THE STATISTICAL INTERPRETATION OF DEGREES OF FREEDOM

WILLIAM J. MOONAN
University of Minnesota
Minneapolis, Minnesota

1. Introduction

THE CONCEPT of "degrees of freedom" has a very simple nature, but this simplicity is not generally exemplified in statistical textbooks. It is the purpose of this paper to discuss and define the statistical aspects of degrees of freedom and thereby clarify the meaning of the term. This shall be accomplished by considering a very elementary statistical problem of estimation and progressing onward through more difficult but common problems until finally a multivariate problem is used. The available literature which is devoted to degrees of freedom is very limited. Some of these references are given in the bibliography and they contain algebraic, geometrical, physical and rational interpretations. The main emphasis in this article will be found to be on discovering the degrees of freedom associated with certain standard errors of common and useful significance tests, and that for some models, parameters are estimated directly or indirectly, by certain degrees of freedom. The procedures given here may be put forth completely in the system of estimation which utilizes the principle of least squares. The application given here are special cases of this system.

2.

In most statistical problems it is assumed that n random variables are available for some analysis. With these variables, it is possible to construct certain functions called statistics with which estimations and tests of hypotheses are made. Associated with these statistics are numbers of degrees of freedom. To elaborate and explain what this means, let us start out with a very simple situation. Suppose we have two random variables, y_1 and y_2. If we pursue an objective of statistics, which is called the reduction of data, we might construct the linear function, $Y_1 = \frac{1}{2} y_1 + \frac{1}{2} y_2$.

This function estimates the mean of the population from which the random variables were drawn. For that matter so does any other linear function of the form, $Y_1 = a_{11} y_1 + a_{12} y_2$ where the a's are real equal numbers. When the coefficients of the random variables are equal to the reciprocal of the number of them, the statistic defined is the sample mean. This statistic may be chosen here for logical reasons, but its specification really comes from the theory of estimation mentioned before. We also could construct another linear function of the random variables, $Y_2 = \frac{1}{2} y_1 - \frac{1}{2} y_2$.

This contrast statistic is a measure of how well our observations agree since it yields a measure of the average difference of the variables. These statistics, Y_1 and Y_2, have the valuable property that they contain all the available information relevant to discerning characteristics of the population from which the y's were drawn. This is true because it is possible to reconstruct the original random variables from them. Clearly, $Y_1 = Y_2 = y_1$ and $Y_1 - Y_2 = y_2$. We discern that we have constructed a pair of statistics which are reduceable to the original variables, but they state the information contained in the variables in a more useful form. There are certain other characteristics worth noticing. The sum of the coefficients of the random variables of Y_2 equals zero and the sum of the products of the corresponding coefficients of the random variables of Y_1 and Y_2 equals zero. That is, $(\frac{1}{2})(\frac{1}{2}) + (\frac{1}{2})(-\frac{1}{2}) = 0$. This latter property is known as the quasi-orthogonality of Y_1 and Y_2. This property is analogous to the property of independence which is associated with the random variables.

In changing our random variables to the statistics we have performed a quasi-orthogonal transformation. Quasi-orthogonal transformations are of special interest because the statistics to which they lead have valuable properties. In particular, if our data are composed of random variables from a normal population, these statistics are independent in the probability sense, (i. e., stochastically independent) or in other words, they are uncorrelated. That remark has a rational interpretation which says that the statistics used are not overlapping in the information they reveal about the data. As long as we preserve the property of orthogonality we will be able to reproduce the original random variables at will. This reproductive property is guaranteed when the coefficients of the random variables of the statistics are mutually orthogonal (i. e., every statistic is orthogonal to every other one), since the determinant of such coefficients does not vanish when this is true, our equations (statistics) have a solution which is the explicit designation of the original random variables. The determinant for this problem is

$$(1) \quad \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} = (\tfrac{1}{2})(-\tfrac{1}{2}) - (\tfrac{1}{2})(\tfrac{1}{2}) = -\tfrac{1}{2} = 0$$

There is another valuable property of quasi-orthogonal transformations which we shall come to a little later.

3.

If we have three observations, we can construct three mutually quasi-orthogonal statistics. Again we might let Y_1 be the mean of the random variables with Y_2 and Y_3 as contrast statistics. Specifically, let $Y_1 = \frac{1}{3} y_1 = \frac{1}{3} y_2 = \frac{1}{3} y_3$. There exist two other mutually quasi-orthogonal linear statistics which might be chosen, and it can be said that we enjoy the freedom of two choices in the statistics we actually use to summarize the data. We could let

$$(2) \quad Y_2 = \tfrac{1}{2} y_1 - \tfrac{1}{2} y_2 + \tfrac{0}{2} y_3; \; Y_3 = \tfrac{1}{3} y_1 + \tfrac{1}{3} y_2 - \tfrac{2}{3} y_3.$$

or,

$$(3) \quad Y_2 = \tfrac{0}{2} y_1 + \tfrac{1}{2} y_2 - \tfrac{1}{2} y_3; \; Y_3 = \tfrac{2}{3} y_1 - \tfrac{1}{3} y_2 - \tfrac{1}{3} y_3.$$

(It can be shown that there exists an infinity of possible choices!)

Either pair of the statistics which we have chosen together with Y_1 can be shown to reproduce the random variables y_1, and y_2 and y_3. As a consequence, they possess all the information that the original variables do. In general, if we have n random variables, we might construct a statistic representing the sample mean (which estimates θ) and have n - 1 choices or degrees of freedom for other mutually quasi-orthogonal linear statistics to summarize the data. Each degree of freedom then corresponds to a mutually quasi-orthogonal linear function of the random variables. In general, the term degree of freedom does not necessarily refer to a linear function which is orthogonal to all the others which are or may be constructed; however, in common usage it usually does refer to quasi-orthogonal linear functions.

When the observational model we are working with contains only parameter which is estimated by a linear function, there is little purpose in specifying the remaining degrees of freedom in the form of contrasts. For instance, if our model is $y_i = \theta + e_i$ is normally distributed with zero mean and variance σ^2, i. e., $N(o, \sigma^2)$, and $i = 1, \ldots, n$, we would also like to estimate σ^2. Unfortunately, this parameter is not estimated directly by linear functions other than Y_1.

Before proceeding, the other property of quasi-orthogonal transformations will be discussed. One might inquire about the relationship of the number called the sum of the squares to the y_i's to the sum of squares of the Y_j's. If we require this number to be invariant, then

$$(4) \quad \sum_{j=1}^{n} Y_j^2 = \sum_{i=1}^{n} y_i^2.$$

For two statistics, we can write in matrix notation,

$$(5) \quad \begin{Bmatrix} Y_1 \\ Y_2 \end{Bmatrix} \begin{Bmatrix} a_{11} & a_{11} \\ a_{21} & a_{22} \end{Bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = Y = Ay.$$

Then, $\sum_{j=1}^{n} y_j^2 = Y' \, Y = (a \, y) \, '(Ay) = y' \, A' \, Ay.$

Now if $\Sigma \; Y_j^2 = Y'Y$ is to equal $\sum_{i=1}^{n} y_i^2 = y'y$,

then A'A is a two row-two column matrix with ones in the main diagonal, i. e., $A'A = \begin{pmatrix} 1 & o \\ o & 1 \end{pmatrix}$

A matrix, A', which when multiplied by its transpose, A, equals a unit matrix, then A' is called an orthogonal matrix and the y_i's which are transformed to the Y_j's by this matrix are said to be orthogonally transformed. You will notice that the matrix of the coefficients of Y_1 and Y_2 of section 2 is not an orthogonal matrix since

$$A'A = \begin{Bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{Bmatrix}$$

If the coefficients of the Y's had been $1/\sqrt{2}$'s instead of $\frac{1}{2}$'s then A' would be an orthogonal matrix. Because the matrix of our transformations does not fulfill the accepted mathematical definition of orthogonal transformations, but one very much like them, they are termed, for the purposes of this paper, quasi-orthogonal transformations. However, it seems unnatural to beginning students to define Y_1 as $y_1 = \frac{1}{\sqrt{2}} y_1 + \frac{1}{\sqrt{2}} y_2$. Actually, for Y_1 any linear function with positive and equal coefficients would serve as well as Y_1 itself for they would be logically equivalent and mathematically reducible to the usual definition of the sample mean. If we are to use the common-sense statistics, obviously something must be done in order to preserve the property (4). One thing that can be done is to change our definition of what the sum of squares of the jth linear function, Y_j^2, would be. Let us define the sum of squares associated with the linear function Y_j to be

$$(6) \quad SS(Y_j) = \frac{(a_{1j} \, y_1 + a_{2j} \, y_2 + \ldots + a_{nj} \, y_n)^2}{a_{1j}^2 + a_{2j}^2 + \ldots + a_{nj}^2}$$

Using this definition instead of just the numerator of it, property (4) will be preserved. As an illustration of this formula let j = 1 and

$Y_1 = \frac{1}{3} y_1 + \frac{1}{3} y_2 + \frac{1}{3} y_3$, then

$$(7)\ SS(Y_1) = \frac{(\frac{1}{3} y_1 + \frac{1}{3} y_2 + \frac{1}{3} y_3)^2}{(\frac{1}{3})2 + (\frac{1}{3})2 + (\frac{1}{3})2} \quad \frac{(\sum_{i=1}^{3} y_i)^2}{3}$$

or for n random variables, $SS(Y_1) = (\sum_{i=1}^{n} y_i)^2/n$.

Further, if $y_1 = 24$, $y_2 = 18$ and $y_3 = 36$, then $SS(Y_1) = 2028$, and if we use (2), then $SS(Y_2) = 18$ and $SS(Y_3) = 150$. Note that $SS(Y_1) + SS(Y_1) + SS(Y_2) + SS(Y_3) = 2196$ and that

$$\sum_{i=1}^{3} y_i^2 = 24^2 + 18^2 + 36^2 = 2196$$

Thus the sum of squares of the linear function equals the sum of squares of the random variables. These results can, of course, be generalized to the n-variable case. Clearly, the sum of squares of the two linear functions Y_2 and Y_3 equals the total sum of squares of the random variables minus the sum of squares associated Y_1, so:

$$(8)\ SS(Y_2) + SS(Y_3) = \sum_{i=1}^{3} y_i^2 - SS(Y_1) = \sum_{i=1}^{3} y_i^2 - \frac{(\sum_{i=1}^{3} y_i)^2}{3}$$

or, in general,

$$(9)\ SS(Y_2) + \dots + SS(Y_n) = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

Now define the sample variance of a set of linear functions as the average of the sums of squares associated with the contrast linear functions. We see that for the special case where n = 3, our division for this average will be 2 because three are two sums of squares to be averaged in (8). This argument accounts for the degrees of freedom divisor which has been traditionally difficult to explain to beginning students in the formula

$$(10)\ S^2 = \frac{\sum_{i=1}^{n} y_i}{n-1} - \frac{(\sum_{i=1}^{n} y_i)^2}{n(n-1)} = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}$$

The statistic Y_1, accounts for one degree of freedom in the numerator of the formula for Student's t and the denominator is a function of (10) and is associated with n - 1 degrees of freedom. Note that it is not necessary to construct the contrast degrees of freedom to obtain the sums of squares associated with them.

4.

The problem just presented is a simple analysis of variance (anova) type and leads to the test of the hypothesis, $\theta = \theta_0$. The next logical elaboration would be to consider Fisher's t test of the hypothesis $\theta_1 = \theta_2$. The observation model is $y_{ik} = \theta_k + e_{ik}$, where $i = 1, \dots, n_k$; $k = 1, 2$ and e_{ik} are $N(0, \sigma^2 = \sigma_1^2 = \sigma_2^2)$. The orthogonal linear functions which estimate the parameters θ_1 and θ_2, are respectively,

$$Y_1 = \frac{1}{n_1} y_{11} + \dots + \frac{1}{n_1} y_{n_1 1} + \frac{0}{n_2} y_{12} + \dots + \frac{0}{n_2} y_{n_2 2}$$

$$\text{and } Y_2 = \frac{0}{n_1} y_{11} + \dots + \frac{0}{n_1} y_{n_1 1} + \frac{1}{n_2} y_{12} + \dots + \frac{1}{n_2} y_{n_2 2}.$$

Then,

$$(11)\ SS(Y_3) + \dots + SS(Y_{n_1+n_2}) = \sum_{i=1}^{n_k} \sum_{j=1}^{2} y_{ik} - SS(Y_1) -$$

$$SS(Y_2) = \sum_{i=1}^{n_k} \sum_{j=1}^{2} Y_{ij}^2 - \frac{(\sum_{i=1}^{n_1} y_{i1})^2}{n_2} - \frac{(\sum_{i=1}^{n_2} y_{i2})^2}{n_2} =$$

$$\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2$$

and if we average these sum of squares, the appropriate denominator will be $n_1 + n_2^{-2}$. The numerator of Fisher's t is $Y_1 - Y_2$ under the null hypothesis $\theta_1 = \theta_2$ and the denominator is a function of (11) and is associated with $n_1 + n_2 - 2$ degrees of freedom.

5.

As another example, we might consider the regression model, $y_i = \theta + \beta(x_i - \bar{x}) + e_i$, where $i = 1, \dots, n$ and e_i are $N(0, \sigma^2_{y.x})$. The linear functions of interest are $Y_1 = \frac{1}{n} y_n$ and $Y_2 = \frac{(X_1 - \bar{X})}{n} y_1 + \dots + \frac{(X_n - \bar{X})}{n} y_n$. For these functions, Y_1 is used to estimate the mean, θ and Y_2, being an an average product of the deviation x's and concomitant y's, leads to an estimate of the unknown constant of proportionality, β. This is rationally and algebraically true, since if y_i and $(x_i - \bar{x})$ tend to proportionately increase and decrease simultaneously or inversely, Y_2 will tend to increase absolutely. However, if y_i and $(x_i - \bar{x})$ do not proportionately rise and fall simultaneously or inversely, Y_2 will tend to be zero. This can be shown by the following table. In this table, several sets of x's designated by x_{ik}, $k = 1, \dots, 3$, each of which have the same mean, 4, are substituted in Y_2 together with their corresponding y_i's. The values of the Y_{2k} are given in the bottom line of Table I.

TABLE I

EVALUATION OF Y_2 FOR CHANGING VALUES OF x_i IN THE SIMPLE REGRESSION MODEL

y_i \ x_{ik}	x_{i1}	x_{i3}	x_{i5}
1	2	6	5
2	4	4	2
3	6	2	5
Y_{2k}	$\frac{4}{3}$	$-\frac{4}{3}$	0

Using (6) we find

$$(12)\ SS(Y_1) = \frac{(\sum_{i=1}^{n} y_i)^2}{n} \quad \text{and } SS(Y_2) = \frac{[\sum_{i=i}^{n}(x_i - \bar{x})y_i]^2}{\sum_{i=i}^{n}(x_i - \bar{x})^2}$$

Consequently, to find the sample estimate of $\sigma^2_{y \cdot x}$,

$$(13)\ SS(Y_3)+\ldots+SS(Y_n) = \sum_{i=1}^{n} y_i^2 - SS(Y_1) - SS(Y_2)=i=1$$

$$\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} - \frac{[\sum_{i=1}^{n}(x_i - \bar{x})y_i]^2}{\sum_{i=1}^{n}(x_i - x)^2} =$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 - b \sum_{i=1}^{n} (x_i - \bar{x})y_i = \sum_{i=1}^{n} (y_i - \tilde{y}_1)^2$$

where b is the usual regression coefficient for predicting y from a knowledge of x and $\tilde{y}$ is the predicted value of y_i. Again to find the variance associated with these sums of squares we divide their sums of squares by the number of degrees of freedom from which these sums of square were derived. This number is n - 2. Under the null hypothesis, $\beta = 0$, the denominator of the t test, $t = b/S.E_b$, has (n - 2) degrees of freedom and the numerator is associated with one degree of freedom.

6.

It is fairly laborious to calculate the $SS(Y_j)$ and because of this it is desirable to have a method whereby the sum of squares associated with several linear functions may be conveniently found. The proof of the method is fairly long and will not be reproduced here. Its exposition will have to suffice.

Let a_j be the coefficient vector of the random variables of the jth degree of freedom and let y be the observation vector, $(y_1\ y_2, \ldots, y_n)$. With these values construct the following system of equations:

$$p_1(a_1 \cdot a_1) + p_2(a_1 \cdot a_2) + \ldots + p_m(a_1 \cdot a_m) = (a_1 \cdot y)$$

(14)

$$p_1(a_2 \cdot a_1)+p_2(a_2 \cdot a_2)+\ldots+p_m(a_2 \cdot a_m) = (a_2 \cdot y)$$

$$\vdots \qquad \vdots \qquad \ddots \qquad \vdots \qquad \vdots$$

$$p_1(a_m \cdot a_1)+p_2(a_m \cdot a_2)+\ldots+p_m(a_m \cdot a_m) = (a_m \cdot y)$$

When these equations are solved, by whatever method is convenient, the sum of squares for the m degrees of freedom, $Y_1, Y_2, \ldots, Y_m (m \leq n)$ is given by

$$(15)\ p_1(a_1 \cdot y) + p_2(a_2 \cdot y)+\ldots+p_m(a_m \cdot y)$$

The method reveals the correct sum of squares whether or not the degrees of freedom are mutually orthogonal, but we shall illustrate it for the orthogonal case. Consider again (2) and then let

$$a_2 = (\tfrac{1}{2}, -\tfrac{1}{2}, \overline{2}),\ a_3(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3})$$

and $y = (y_1, y_2, y_3) = (24, 18, 36)$. Corresponding to (14) we have

$$(16)\ p_2(\tfrac{1}{2}) + p_3\ (0) = 3 \qquad p_2(0) + p_3(\tfrac{2}{3}) = 10.$$

Therefore $p_2 = 6$ and $p_3 = 15$, then $SS(Y_2)+SS(Y_3)= 6(3)+(15)(10) = 168$. In some previous work in section 3, we found $SS(Y_2) = 18$ and $SS(Y_3) = 150$, so this result checks. In this problem, Y_1 was neglected in order to show that (4) is quite general for any $m<n$.

7.

All of these principles may be easily generalized to the multivariate case. What is needed is to use matrix variables instead of the single ones we have been using. Using the Least Squares Principle, the ideas presented here (and many others) have been applied to multivariate analysis of variance in reference number 4. The following and last example is taken from this source. Suppose,

$$y_1^1 = 11,\ y_2^1 = 5,\ y_3^1 = 8;\ y_1^2 = 2,\ y_2^2 = 6,\ y_3^3 = 13.$$

Here the superscripts indicate which variate is being considered (these numbers are not to be confused with powers), and the subscripts designate the variables. Also let

$$Y_1^a = \frac{y_1^a}{3} + \frac{y_2^a}{3} + \frac{y_3^a}{3},\ Y_2^a = \frac{y_1^a}{2} - \frac{y_2^a}{2} \text{ and } Y_3^a =$$

$$\frac{y_1^a}{3} + \frac{y_2^a}{3} - \frac{2y_3^a}{3},$$

where a = 1, 2. We have, corresponding to (14)
$p_1^1 (\frac{1}{3}) + p_2^1 (0) + p_3^1 (0) = 8$

(17) $p_1^1 (0) + p_2^1 (\frac{1}{2}) + p_3^1 (\frac{2}{3}) = 0$

$p_1^1 (0) + p_2^1 (0) + p_3^1 (\frac{2}{3}) = 0$

Therefore, $p_1^1 = 24$, $p_2^1 = 6$, $p_3^1 = 0$ and using (15) we find (24)(8) + (4)(2) + (0)(0) = 210 which is equal to $11^2 + 5^2 + 8^2$. For the second variate $p_1^2 (\frac{1}{3}) + p_2^2 (0) + p_3^2 (0) = 7$.

(18) $p_1^2 (0) + p_2^2 (\frac{1}{2}) + p_3^2 (0) = -2$

$p_1^2 (0) + p_2^2 (0) + p_3^2 (\frac{2}{3}) = -6$

Solving, we get $p_1^2 = 21$, $p_2^2 = -4$, $p_3^2 = -9$ and corresponding to (15), (21)(7) + (-4)(-2) + (-9)(-6)= 209 which is equal to $2^2 + 6^2 = 13^2$. The sum of cross-products of these three vector degrees of freedom for the two variates may be found in one of two ways; either (24)(7) + 6(-2) + 0(-6) = 156 or (21)(8) + (-4)(3) + (-9)(10) = 156. Both results are equal to (11)(2) + (5)(6) + (8)(13). The matrix

210	156
156	209

corresponds to the total sum of squares and cross products for the bivariate sample observations which have been transformed by the vector degrees of freedom Y_j^a, j=1, 2, 2. We note that the sums of squares and cross-products of the variables for each variate is preserved by the orthogonal vector set of degrees of freedom. This simple problem serves to illustrate this invariance property for a multivariate case.

8. Summary

We have seen that certain statistical problems are formulated in terms of linear functions of the random variables. These linear functions, called degrees of freedom, served the purpose of presenting the data in a more usable form because the functions led directly or indirectly to estimates of the parameters of the observation model and the estimate of variance of the observations. Moreover, these estimates may be used to test hypotheses about the population parameters by the standard statistical tests.

Modern statistical usage of the concept of degrees of freedom had its inception in Student's classic work, reference 7, which is often considered the paper which was necessary to the development of modern statistics. Fisher, beginning with his frequency distribution study, reference 2, has generalizations to work in their many contributions to the general theory of regression analysis.

This paper has resulted from an attempt to bring clarification to the statistical interpretation of degrees of freedom. The author feels that his attempt will not be altogether successful for there remain many questions which students may or should ask that have not been answered here. A satisfactory exposition could be given by a complete presentation of the theory of least squares which is slanted towards the problems of modern regression theory of the analysis of variance type. This discussion would appropriately take book form, however.

REFERENCES

1. Cramer, Harald, Mathematical Methods of Statistics (Princeton, N. J.: Princeton University Press, 1946).
2. Fisher, Ronald A., "Frequency Distribution of the Values of the Correlation In Samples from an Independently Large Population," Biometrika, X (1915), pp. 507-521.
3. Johnson, Palmer O., Statistical Methods in Research (New York: Prentice Hall, Inc., 1948).
4. Moonan, William J., The Generalization of the Principles of Some Modern Experimental Designs for Educational and Psychologic Research. Unpublished thesis, University of Minnesota, Minneapolis, Minnesota, 1952.
5. Rulon, Phillip J., "Matrix Representation of Models for the Analysis of Variance and Covariance," Psychometrika, XIV (1949), pp. 259-278.
6. Snedecor, George W., Statistical Methods (Ames, Iowa: Collegiate Press, 1946).
7. Student, "The Probable Error of the Mean," Biometrika, VI (1908), pp. 1-25.

8. Tukey, John W., "Standard Methods of Analyzing Data," Proceedings: Computation Seminar (New York: International Business Machines Corporation, 1949), pp. 95-112.
9. Walker, John W., "Degrees of Freedom," Journal of Educational Psychology, XXI (1940), pp. 253-269.
10. Walker, Helen M., Mathematics Essential for Elementary Statistics (New York: Henry Holt and Co., 1951).
11. Yates, Frank, The Design and Analysis of Factorial Experiments, Imperial Bureau of Soil Science, Technical Communication No. 35, Harpenden, England: 1937.