Establishing χ^2 Properties of Sums of Squares Using Induction

SHAYLE R. SEARLE and FRIEDRICH PUKELSHEIM*

The between-classes sum of squares in a between- and within-classes analysis of variance has, under normality, a χ^2 distribution. Although "substantial mathematical machinery" (Stigler 1984) is often used in classroom derivation of this distribution, it can be avoided by using induction and independence properties of standard normal variables. This is the derivation given here for unequal subclass numbers data. Independence of the between- and within-classes sums of squares is also shown.

KEY WORDS: Helmert transformation; Independent sums of squares; One-way classification; Unbalanced data.

1. INTRODUCTION

Stigler (1984) rightly points out that "In introductory courses in mathematical statistics, the proof that the sample mean $\bar{X}$ and sample variance s^2 are independent when one is sampling from normal populations is commonly deferred until substantial mathematical machinery has been developed" (p. 134). Contrasting this, Stigler then gives a nice proof for the one-sample case that requires understanding nothing more than normality and independence, together with the definition of a χ^2 variable as the sum of squares of independent and identically distributed (iid) standard normal variables. The sum of independent χ^2 variables being distributed as χ^2 is also used. His method of proof, which relies on induction on sample size, is extended here to sums of squares in a one-way classification with unbalanced data (unequal subclass numbers data). It is in this situation of the analysis of variance of unbalanced data that Stigler's "substantial mathematical machinery" is, generally speaking, nowhere more evident; for teaching purposes there are great advantages in being able to avoid such complexities, as is done here.

For observations y_{ij} for $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$, assume that observations having the same value of i (those in the ith class) are identically distributed with a normal density having mean μ_i and variance σ^2. Write this as

$$y_{ij} \sim \text{iid } N(\mu_i, \sigma^2) \quad \text{for } j = 1, \ldots, n_i, \tag{1}$$

and let this be true for each $i = 1, \ldots, a$. Assume also that observations in each class are independent of those in every other class. Thus, using $v(y_{ij})$ for the variance of y_{ij} and $\text{cov}(y_{ij}, y_{hk})$ for the covariance between y_{ij} and y_{hk},

$$v(y_{ij}) = \sigma^2 \quad \text{and} \quad \text{cov}(y_{ij}, y_{hk}) = 0 \tag{2}$$

for all i, j and for all h, k except $i = h$ with $j = k$.

The sample mean of the observations in class i will be denoted by

$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i. \tag{3}$$

We define (for subsequent convenience) partial sums of the n_i values:

$$s_i = \sum_{t=1}^{i} n_t \quad \text{and particularly} \quad s_a = \sum_{t=1}^{a} n_t = n. \tag{4}$$

In addition, the mean of all observations in all a classes will be denoted by

$$m_a = \bar{y}. = \sum_{i=1}^{a} n_i \bar{y}_i \Big/ \sum_{i=1}^{a} n_i$$
$$= \sum_{i=1}^{a} n_i \bar{y}_i / s_a = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}/s_a. \tag{5}$$

Then from (2), (3), and (5),

$$v(\bar{y}_i) = \sigma^2/n_i \quad \text{and} \quad v(m_a) = \sigma^2/s_a; \tag{6}$$

$$\text{cov}(y_{ij}, \bar{y}_i) = \sigma^2/n_i \quad \text{and} \quad \text{cov}(\bar{y}_i, \bar{y}_{i'}) = 0, \tag{7}$$

for $i \neq i'$; and for $i = 1, \ldots, a$

$$\text{cov}(\bar{y}_i, m_a) = n_i\sigma^2/(n_i s_a) = \sigma^2/s_a. \tag{8}$$

Four distributional results are taken as known: (a) that normal variables having zero covariance are independent; (b) that linear combinations of normal variables are normally distributed; (c) that a χ^2_k variable [having k degrees of freedom (df)] is definable as the sum of squares of k iid standard normal variables; and (d) that sums of independent χ^2 variables are χ^2 variables. These results are referred to frequently in what follows.

We deal with between- and within-class sums of squares defined for a classes as

$$B_a = \sum_{i=1}^{a} n_i(\bar{y}_i - m_a)^2 = \sum_{i=1}^{a} n_i \bar{y}_i^2 - s_a m_a^2 \tag{9}$$

and

$$W_a = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \tag{10}$$

with

$$W_a = \sum_{i=1}^{a} W_i \quad \text{for} \quad W_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2. \tag{11}$$

2. INDEPENDENCE

The independence of B_a and W_a stems directly from (a). Consider one of the terms in B_a of (9) that is squared, say $\bar{y}_i - m_a$, and a similar term from W_a of (10), say $y_{hj} - \bar{y}_k$. The covariance of these terms for $i = h$ is, from (7) and (8),

$$\text{cov}(\bar{y}_i - m_a, y_{ij} - \bar{y}_i)$$
$$= \text{cov}(\bar{y}_i, y_{ij}) - v(\bar{y}_i) - \text{cov}(m_a, y_{ij}) + \text{cov}(m_a, \bar{y}_i)$$

*Shayle R. Searle is Professor, Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, NY 14853. Friedrich Pukelsheim is Professor, Lehrstuhl für Stochastik, Institut für Mathematik, Universität Augsburg, Memminger Strasse 6, D-8900, Augsburg, Federal Republic of Germany. The work of the first author was done while on leave from Cornell University and supported by a Senior U.S. Scientist Award from the Alexander von Humboldt-Stiftung. This is Paper BU-470 in the Biometrics Unit, Cornell University.

$$= \sigma^2(1/n_i - 1/n_i - 1/s_a + 1/s_a)$$
$$= 0;$$

and similarly for $i \neq h$,

$$\mathrm{cov}(\bar{y}_i - m_a, y_{hj} - \bar{y}_h) = \sigma^2 0\left(0 - 0 - \frac{1}{s_a} + \frac{1}{s_a}\right) = 0.$$

Thus $\bar{y}_i - m_a$ and $y_{hj} - \bar{y}_h$ have zero covariance, and so, since by (b) each of them is normally distributed, they are by (a) independent. Since this is so for all i, h, and j, it is true for *all* pairs of terms that are squared, one in B_a and one in W_a. Therefore B_a and W_a are independent.

3. INDUCTION

By induction on a we show that $B_a/\sigma^2 \sim \chi^2_{a-1}$. The starting point is the case of $a = 2$. From (9),

$$B_2 = n_1\bar{y}_1^2 + n_2\bar{y}_2^2 - (n_1\bar{y}_1 + n_2\bar{y}_2)^2/(n_1 + n_2)$$
$$= n_1 n_2(\bar{y}_1 - \bar{y}_2)^2/(n_1 + n_2).$$

From (6) and (7),

$$v(\bar{y}_1 - \bar{y}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{(n_1 + n_2)\sigma^2}{n_1 n_2}; \quad (12)$$

under the hypothesis $H : \mu_1 = \mu_2$, (12) and (b) yield

$$\bar{y}_1 - \bar{y}_2 \sim N[0,(n_1 + n_2)\sigma^2/n_1 n_2].$$

Therefore on defining $g = [n_1n_2/(n_1 + n_2)]^{1/2}(\bar{y}_1 - \bar{y}_2)$, we have $g \sim N(0,1)$; so since $B_2/\sigma^2 = g^2$, (c) gives $B_2/\sigma^2 \sim \chi^2_1 = \chi^2_{2-1}$. Thus $B_a/\sigma^2 \sim \chi^2_{a-1}$ is certainly true for $a = 2$. With this as a base we now show that assuming $B_a/\sigma^2 \sim \chi^2_{a-1}$ yields $B_{a+1}/\sigma^2 \sim \chi^2_a$; that is, induction on a establishes that $B_a/\sigma^2 \sim \chi^2_a$ is true generally.

Relationships between m_a and m_{a+1}, and between B_a and B_{a+1}, are needed that are extensions of well-known recurrence formulae for sample means and variances given in Searle (1983) and Stigler (1984). First, from (4) and (5)

$$m_{a+1} = \sum_{i=1}^{a+1} n_i\bar{y}_i/s_{a+1}$$
$$= (s_a m_a + n_{a+1}\bar{y}_{a+1})/(s_a + n_{a+1})$$
$$= m_a + n_{a+1}(\bar{y}_{a+1} - m_a)/s_{a+1}. \quad (13)$$

Second, from (9)

$$B_{a+1} = \sum_{i=1}^{a+1} n_i\bar{y}_i^2 - s_{a+1}m_{a+1}^2,$$

and on using (13), this is

$$\begin{aligned}
B_{a+1} &= \sum_{i=1}^{a} n_i\bar{y}_i^2 + n_{a+1}\bar{y}_{a+1}^2 \\
&\quad - s_{a+1}[m_a + n_{a+1}(\bar{y}_{a+1} - m_a)/s_{a+1}]^2 \\
&= \sum_{i=1}^{a} n_i\bar{y}_i^2 + n_{a+1}\bar{y}_{a+1}^2 - (s_a + n_{a+1})m_a^2 \\
&\quad - 2m_a n_{a+1}(\bar{y}_{a+1} - m_a) \\
&\quad - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2/s_{a+1} \\
&= \sum_{i=1}^{a} n_i\bar{y}_i^2 - s_a m_a^2 + n_{a+1}(\bar{y}_{a+1} - m_a)^2 \\
&\quad - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2/s_{a+1} \\
&= B_a + n_{a+1}(1 - n_{a+1}/s_{a+1})(\bar{y}_{a+1} - m_a)^2;
\end{aligned}$$

that is,

$$B_{a+1} = B_a + \delta \quad \text{for } \delta = \left(\frac{n_{a+1}s_a}{s_{a+1}}\right)(\bar{y}_{a+1} - m_a)^2. \quad (14)$$

Now $\bar{y}_{a+1}$ and m_a are independent, and so

$$v(\bar{y}_{a+1} - m_a) = \sigma^2(1/n_{a+1} + 1/s_a) = s_{a+1}\sigma^2/n_{a+1}s_a.$$

Hence, under the hypothesis

$$H : \mu_1 = \mu_2 = \cdots = \mu_{a+1}, \quad (15)$$

$\bar{y}_{a+1} - m_a \sim N[0, (s_{a+1}/n_{a+1}s_a)\sigma^2]$. Thus, just as in deriving $B_2/\sigma^2 \sim \chi^2_1$, we have $\delta/\sigma^2 \sim \chi^2_1$. Furthermore, $B_a = \Sigma_{i=1}^a n_i(\bar{y}_i - m_a)^2$ and for $i = 1, \ldots, a$,

$$\mathrm{cov}(\bar{y}_i - m_a, \bar{y}_{a+1} - m_a) = \sigma^2\left(0 - \frac{1}{s_a} - 0 + \frac{1}{s_a}\right) = 0.$$

Therefore, since $\bar{y}_i - m_a$ and $\bar{y}_{a+1} - m_a$ are by (b) normally distributed, they are by (a) independent. Therefore in (14) B_a and δ are independent; so with $B_a/\sigma^2 \sim \chi^2_{a-1}$ and $\delta/\sigma^2 \sim \chi^2_1$, this independence means from (d) that $B_{a+1}/\sigma^2 = B_a/\sigma^2 + \delta/\sigma^2 \sim \chi^2_{a-1+1} = \chi^2_a$. Thus is the χ^2 property of B_a proven, without recourse to any "substantial mathematical machinery."

4. THE WITHIN-CLASS SUM OF SQUARES

The χ^2 property of W_a can now be derived from that of B_a. Suppose that $n_i = 1$ for $i = 1, \ldots, a$. Then B_a becomes $\Sigma_{i=1}^a(y_i - \bar{y})^2$ for $\bar{y} = \Sigma_{i=1}^a y_i/a$, and so by the immediately preceding result, $\Sigma_{i=1}^a (y_i - \bar{y})^2/\sigma^2 \sim \chi^2_{a-1}$. A special case of this is W_i of (11):

$$W_i/\sigma^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2/\sigma^2 \sim \chi^2_{n_i-1}.$$

Hence, since the W_is are distributed independently, $W/\sigma^2 = \Sigma_{i=1}^a W_i/\sigma^2$ has by (d) a χ^2 distribution on $\Sigma_{i=1}^a(n_i - 1) = n. - a$ df; that is, $W/\sigma^2 \sim \chi^2_{n\cdot-a}$.

5. APPLICATION

The ultimate application of these results is, of course, that under the hypothesis (15), which gives $B_a/\sigma^2 \sim \chi^2_{a-1}$ independently of $W_a/\sigma^2 \sim \chi^2_{n\cdot-a}$, the ratio

$$F = \frac{(B_a/\sigma^2)/(a - 1)}{(W_a/\sigma^2)/(n. - a)} = \frac{B_a/(a - 1)}{W_a/(n. - a)}$$

has the F distribution on $(a - 1)$ and $(n. - a)$ df and can be used as a test statistic for the hypothesis (15).

6. EXTENDING HELMERT'S TRANSFORMATION

In the simple case of x_i for $i = 1, \ldots, n$ with $x_i \sim$ iid $N(0,\sigma^2)$, the χ^2_{n-1} distribution of $S^2/\sigma^2 = \Sigma_{i=1}^n(x_i - \bar{x})^2/\sigma^2$ can be derived by using what is known (e.g., Lancaster 1972) as Helmert's transformation:

$$u_i = \sum_{j=1}^{i} \lambda_{ij}x_j \quad \text{for } i = 2, \ldots, n, \quad (16)$$

with $\lambda_{ij} = 1/[i(i - 1)]^{1/2}$ for $j = 1, 2, \ldots, i - 1$ and $\lambda_{ii} = -[(i - 1)/i]^{1/2}$. It is then easily shown that the u_i of (16) are iid $N(0,\sigma^2)$ and that $S^2 = \Sigma_{i=2}^n u_i^2$; then (d) gives $S^2/\sigma^2 \sim \chi^2_{n-1}$.

An extension of (16) provides a second proof that $B_a/\sigma^2 \sim \chi^2_{a-1}$. It uses

$$z_i = \sum_{j=1}^{i} t_{ij}\bar{y}_i \quad \text{for } i = 2, \ldots, a, \tag{17}$$

with $t_{ij} = n_j(n_i/s_{i-1}s_i)^{1/2}$ for $j = 1, \ldots, i-1$ and $t_{ii} = -(n_is_{i-1}/s_i)^{1/2}$. It can then be shown that $v(z_i) = \sigma^2$ and $\text{cov}(z_i, z_h) = 0$ for all $i \neq h = 2, \ldots, a$ and that $\Sigma_{i=2}^{a} z_i^2 = B_a$. Hence the z_i are iid $N(0, \sigma^2)$, and so $B_a/\sigma^2 \sim \chi^2_{a-1}$.

In passing, observe that $n_i = 1$ for all i reduces t_{ij} and t_{ii} of (17) to λ_{ij} and λ_{ii}, respectively, of (16)—as one would expect.

A matrix comment is not out of order: on defining $t_{ij} = 0$ for $j = i + 1, \ldots, a$ and $i = 2, \ldots, a$, the resulting $(a-1) \times a$ matrix $\mathbf{T} = \{t_{ij}\}$ for $i = 2, \ldots, a$ and $j = 1, \ldots, a$ is related to a more general Helmert-style matrix of Irwin (1942), quoted as $\mathbf{H}$ in (4) of Lancaster (1972). The relationship is

$$\mathbf{H} = \begin{bmatrix} \mathbf{n}'/\sqrt{s_a} \\ \mathbf{T} \end{bmatrix} \mathbf{D},$$

where $\mathbf{n}'$ is the row vector $[n \ldots n_a]$ and $\mathbf{D}$ is the diagonal matrix of diagonal elements $1/(n_1)^{1/2}, \ldots, 1/(n_a)^{1/2}$.

[*Received February 1985. Revised May 1985.*]

REFERENCES

Irwin, J. O. (1942), "On the Distribution of a Weighted Estimate of Variance and on Analysis of Variance in Certain Cases of Unequal Weighting," *Journal of the Royal Statistical Society*, Ser. A, 105, 115–118.

Lancaster, H. O. (1972), "The Helmert Matrices," *The American Mathematical Monthly*, 65, 5–12.

Searle, S. R. (1983), "The Recurrence Formulae for Means and Variances," *Teaching Statistics*, 5, 7–10.

Stigler, S. M. (1984), "Kruskal's Proof of the Joint Distribution of $\bar{X}$ and s^2," *The American Statistician*, 38, 134–135.

On Minimum Variance Unbiased Estimators

K. X. KARAKOSTAS*

The problem of finding minimum variance unbiased estimators of various parameters for parametric distributions is an important one in statistics. This article gives analytical formulas for the minimum variance unbiased estimators of parametric functions, which are usually used in a classroom, for two types of densities. The first type is the one-parameter regular exponential family, and the second is a two-parameter family of a continuous random variable whose range depends on the unknown parameters.

KEY WORDS: Completeness; One-parameter regular exponential family; Parametric functions; Range depending on unknown parameter; Sufficient statistic.

1. INTRODUCTION

Suppose that we have a random sample $X_1, X_2, \ldots, X_n$ from a distribution $f(x;\theta)$, $\theta \in \Theta$. Then an interesting theoretical problem is to characterize those parametric functions $h(\theta)$ for which a minimum variance unbiased estimator exists and to give its analytical form.

In Section 2 the one-parameter regular exponential family

$$f(x;\theta) = c(\theta)\exp[Q(\theta)T(x)]v(x), \qquad \theta \in \Theta, \tag{1.1}$$

is considered. For that family, this article characterizes a class of parametric functions $h(\theta)$ for which a minimum variance unbiased estimator exists and gives an analytical formula for finding that estimator. Density functions of the form $f(x;\theta_1,\theta_2) = Q(\theta_1,\theta_2)M(x)$, $\theta_1 < x < \theta_2$, $\theta_1 < \theta_2$, are considered in Section 3. For such families, ready-to-use analytical formulas are given for a minimum variance unbiased estimator for either $h(\theta_1,\theta_2)$, $h(\theta_1)$, or $h(\theta_2)$.

2. REGULAR EXPONENTIAL FAMILY

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution whose form belongs to (1.1). Then it is well known that $T = \Sigma_{i=1}^{n} T(X_i)$ is complete and sufficient for θ, with

$$g(t;\theta) = [c(\theta)]^n\exp[Q(\theta)t]v^*(t;n). \tag{2.1}$$

Following Guenther (1978), we see that if the distribution $g(t;\theta)$ of a complete and sufficient statistic T of θ can be written as

$$g(t;\theta) = w(t)g^*(t;\theta^*)h(\theta), \tag{2.2}$$

where $w(t)$ and $h(\theta)$ are, respectively, functions of t and θ only and $g^*(t;\theta^*)$ is another density of the same form as $g(t;\theta)$ with possibly a different parameter θ^*, then $u(t) = 1/w(t)$ is the minimum variance unbiased estimate for $h(\theta)$.

Now let $h(\theta) = [c(\theta)]^k\exp[Q(\theta)r]$, where $k \leq n$ is a nonnegative integer, be the function of θ for which a minimum variance unbiased estimate is requested. Then from (2.1) we have

$$\begin{aligned} g(t;\theta) &= [c(\theta)]^k\exp[Q(\theta)r][c(\theta)]^{n-k} \\ &\quad \times \exp[Q(\theta)(t-r)]v^*(t;n) \\ &= [c(\theta)]^k\exp[Q(\theta)r]g^*(t-r;\theta^*) \\ &\quad \times [v^*(t-r;n-k)]^{-1}v^*(t;n), \end{aligned} \tag{2.3}$$

*K. X. Karakostas is Lecturer, Statistics Section, Department of Mathematics, University of Ioannina, Ioannina 45332, Greece. The author is grateful to both referees for their helpful comments and suggestions.