Establishing $\chi^2$ Properties of
Sums of Squares Using Induction
(and no Matrices)

Shayle R. Searle
Friedrich Pukelsheim

Preprint Nr. 55
Februar 1985
BU-869-M
Establishing $\chi^2$ Properties of Sums of Squares Using Induction (and no Matrices)

Shayle R. Searle * and Friedrich Pukelsheim
Institut für Mathematik
Universität Augsburg, Federal Republic of Germany

ABSTRACT
The between-classes sum of squares in a between-and-within classes analysis of variance has, under normality, a $\chi^2$ 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.

*On leave from Cornell University, Ithaca, N.Y., supported by a U. S. Senior Scientist Award from the Alexander von Humboldt-Stiftung.
1. Introduction

Stigler (1984) points out, quite rightly, 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." Contrasting this, Stigler then gives a very nice proof for the one-sample case that requires understanding nothing more than normality and independence, together with the definition of a $\chi^2$ variable as the sum of squares of independent and identically distributed standard normal variables. The sum of independent $\chi^2$ variables being distributed as $\chi^2$ 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, so that 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 $i$'th class) are identically distributed with a normal density having mean $\mu_i$ and variance $\sigma^2$. Write this as

$$y_{ij} \sim \text{i.i.d. } N(\mu_i, \sigma^2) \quad \text{for } j = 1, \ldots, n_i;$$

and let this be true for each $i = 1, \ldots, a$. Also assume 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 $y_{hk}$,

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

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

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

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

$$s_i = \sum_{i=1}^{a} n_i, \text{ and particularly } s_a = \sum_{i=1}^{a} n_i = n.$$  

And the mean of all observations in all $a$ classes will be denoted by

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

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

$$v(\bar{y}) = \sigma^2 / n_i \quad \text{and } v(m_a) = \sigma^2 / s_a.$$
cov(y_{ij}, \bar{y}_i) = \sigma^2 \quad \text{and, for } i \neq i', \quad \text{cov}(\bar{y}_i, \bar{y}_{i'}) = 0; \quad (7)

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. \quad (8)

Four distributional results are taken as known: (i) that normal variables having zero covariance are independent; (ii) that linear combinations of normal variables are normally distributed; (iii) that a \( \chi^2_k \) variable (having \( k \) degrees of freedom) is definable as the sum of squares of \( k \) i.i.d. normal standard variables; and (iv) that sums of independent \( \chi^2 \) variables are \( \chi^2 \) variables. These results are referred to frequently in what follows.

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

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

and

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

with

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

2. Independence

The independence of \( B_a \) and \( W_a \) stems directly from (i). 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 \( y_{hj} - \bar{y}_h \). The covariance of these terms for \( i = h \) is, from (7) and (8)

\[
\text{cov}(\bar{y}_i - m_a, y_{ij} - \bar{y}_i) = \text{cov}(\bar{y}_i, y_{ij}) - \nu(\bar{y}_i) - \text{cov}(m_a y_{ij}) + \text{cov}(m_a, \bar{y}_i) = \sigma^2 (1/n_i - 1/n_i - 1/s_a + 1/s_a) = 0;
\]

and, similarly for \( i \neq h \),

\[
\text{cov}(\bar{y}_i - m_a, y_{hj} - \bar{y}_h) = \sigma^2 (0 - 0 - 1/s_a + 1/s_a) = 0.
\]

Thus \( \bar{y}_i - m_a \) and \( y_{hj} - \bar{y}_h \) have zero covariance and so, since by (ii) each of them is normally distributed, they are by (i) independent. Since this is so for all \( i, h \) and \( j \) it is true for every pair 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)
\[ \nu(\bar{y}_1 - \bar{y}_2) = \sigma^2(1/n_1 + 1/n_2) = (n_1 + n_2)\sigma^2/n_1n_2, \tag{12} \]
and so, under the hypothesis \( H : \mu_1 = \mu_2 \), (12) and (ii) yield
\[ \bar{y}_1 - \bar{y}_2 \sim N[0, (n_1 + n_2)\sigma^2/n_1n_2]. \]

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

Relationships between \( m_a \) and \( m_{a+1} \), and \( B_a \) 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)

\[ m_{a+1} = \sum_{i=1}^{a+1} n_i \bar{y}_i / s_{a+1} \tag{13} \]

and 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

\[ B_{a+1} = \sum_{i=1}^{a} n_i \bar{y}_i^2 + n_{a+1} \bar{y}_{a+1}^2 - s_{a+1} \left[ m_a + n_{a+1}(\bar{y}_{a+1} - m_a)/s_{a+1} \right]^2 \]

\[ = \sum_{i=1}^{a} n_i \bar{y}_i^2 + n_{a+1} \bar{y}_{a+1}^2 - (s_a + n_{a+1})m_a^2 \]

\[ - 2m_a n_{a+1}(\bar{y}_{a+1} - m_a) - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2 / s_{a+1} \]

\[ = \sum_{i=1}^{a} n_i \bar{y}_i^2 - s_a m_a^2 + n_{a+1}(\bar{y}_{a+1} - m_a)^2 - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2 / s_{a+1} \]

\[ = B_a + n_{a+1}(1 - n_{a+1}/s_{a+1})(\bar{y}_{a+1} - m_a)^2, \]

i.e.,
\[ B_{a+1} = B_a + \delta \quad \text{for} \quad \delta = (n_{a+1}s_a/s_{a+1})(\bar{y}_{a+1} - m_a)^2. \tag{14} \]

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

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

Hence, under the hypothesis

\[ H : \mu_1 = \mu_2 = \ldots = \mu_{a+1} \tag{15} \]
\[ \bar{y}_{a+1} - m_a \sim N(0, \frac{s_{a+1}}{n_{a+1} s_a}). \]

Thus, just as in deriving \( B_a/\sigma^2 \sim \chi^2_1 \), we have \( \delta/\sigma^2 \sim \chi^2_1 \). Furthermore, \( B_a = \sum_{i=1}^a n_i (\bar{y}_i - m_a)^2 \) and for \( i = 1, \ldots, a \),\( \text{cov}(\bar{y}_i - m_a, \bar{y}_{a+1} - m_a) = \sigma^2 (0 - 1/s_a - 0 + 1/s_a) = 0 \). Therefore, since \( \bar{y}_i - m_a \) and \( \bar{y}_{a+1} - m_a \) are by (i) independent. Therefore in (14) \( B_a \) and \( \delta \) are independent; and so with \( B_a/\sigma^2 \sim \chi^2_{a-1} \) and \( \delta/\sigma^2 \sim \chi^2_1 \), this independence means from (iv) 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 \( \chi^2 \) property of \( B_a \) proven, without recourse to any "substantial mathematical machinery".

4. The within-class sum of squares

The \( \chi^2 \) 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 \( \sum_{i=1}^a (y_i - \bar{y})^2 \) for \( \bar{y} = \sum_{i=1}^a y_i/a \), and so by the immediately preceding result, \( \sum_{i=1}^a (y_i - \bar{y})^2/\sigma^2 \sim \chi^2_{a-1} \). A special case of this is \( W_i \) of (11): \( W_i/\sigma^2 = \sum_{j=1}^a (y_{ij} - \bar{y}_i)^2/\sigma^2 \sim \chi^2_{a-1} \). Hence, since the \( W_i \)s are distributed independently, \( W/\sigma^2 = \sum_{i=1}^a W_i/\sigma^2 \) has, by (iv), a \( \chi^2 \) distribution on \( \sum_{i=1}^a (n_i - 1) = n. - a \) degrees of freedom, i.e., \( W/\sigma^2 \sim \chi^2_{n.-a} \).

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 \( W_a/\sigma^2 \sim \chi^2_{n.-a} \), 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 \( (n. - a) \) degrees of freedom, 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 \( x_i \sim i.i.d. N(0, \sigma^2) \), the \( \chi^2_{a-1} \) distribution of \( S^2/\sigma^2 = \sum_{i=1}^n (x_i - \bar{x})^2/\sigma^2 \) can be derived 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

\[
\lambda_{ij} = 1/\sqrt{i(i - 1)} \quad \text{for } j = 1, 2, \ldots, i - 1,
\]

and

\[
\lambda_{ii} = -\sqrt{(i-1)/i}.
\]

It is then easily shown that the \( u_i \) of (16) are i.i.d. \( N(0, \sigma^2) \) and that \( S^2 = \sum_{i=2}^n u_i^2 \); then (iv) 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

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

with

$$t_{ij} = n_j \sqrt{n_i/s_{i-1} s_i} \quad \text{for } j = 1, \ldots, i - 1$$

and

$$t_{ii} = -\sqrt{n_i s_{i-1}/s_i}.$$

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 $\sum_{i=2}^{a} z_i^2 = B_a$. Hence the $z_i$ are i.i.d. $N(0, \sigma^2)$ 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 $\lambda_{ij}$ and $\lambda_{ii}$ respectively, of (16) — as one would expect.

A final, 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 $T = \{t_{ij}\}$ for $i = 2, \ldots, a$ and $j = 1, \ldots, a$ is related to a more general Helmert-style matrix of Irwin (1942), quoted as $H$ in (4) of Lancaster (1972). The relationship is

$$H = \left[ \frac{n'/\sqrt{s_a}}{T} \right] D$$

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

References


Bestellungen nimmt entgegen:
Institut für Mathematik
der Universität Augsburg
Memminger Str. 6
8900 Augsburg

Verzeichnis der erschienenen Preprints


<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Authors</th>
<th>Year</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Characteristics (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Plasma in the Unit Circle (1983),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Math. Meth. in Appl. Sci. 6 (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>their Relation to Independence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>System Polyhedra (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Regularization (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>with Thermal Memory (1983), Mat.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>257 a 264, 1984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Numer. Math. 68, Eirkhäuser-Verlag</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>by Asymptotic Regularization (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Problemen durch Thermostatkontrollen am festen Rand (1984)</td>
</tr>
</tbody>
</table>


44. Pawlow, I.: Boundary Control of Degenerate Two-Phase Stefan Problems (1984)
51. Appell, J.: Deux Méthodes Topologiques pour la Résolution des Équations Elliptiques Non linéaires sans Compacité
52. Appell, J.; de Pascale Espedito: Théorèmes de Bornage pour l'Opérateur de Nemyckii dans les Espaces Idéaux
53. Appell, J.: Misure di Non Compattezza in Spazi Ideali