DOCUMENT RESUME

ED 397 088 TM 025 122

AUTHOR Hester, Yvette

TITLE Understanding that ANOVA Effects Are Perfectly Uncorrelated.

PUB DATE 25 Jan 96

NOTE 20p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, January 1996).

PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Analysis of Variance; *Computation; *Correlation; Heuristics; *Research Design

IDENTIFIERS *Balanced Designs; *Sum of Squares

ABSTRACT

Analysis of variance (ANOVA) was invented in the 1920s to partition variance of a single dependent variable into uncorrelated parts. Having uncorrelated parts makes the computations involved in ANOVA incredibly easier. This was important before computers were invented, when calculations were all done by hand, and also were done repeatedly to check for calculation errors. This paper demonstrates that ANOVA effects in a balanced design are perfectly uncorrelated. A mathematical proof that the four sums-of-squares (SOS) partitions (two main effect, one two-way interaction, and error) for a factorial two-way design are all uncorrelated, i.e., sum exactly to the SOS of the dependent variable is presented, and a small heuristic data set is included in an appendix to illustrate the proof. (Contains 71 references.) (Author/SLD)

Understanding That ANOVA Effects are Perfectly Uncorrelated

Yvette Hester
Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA, January 25, 1996

Fisher invented ANOVA in the 1920's to partition variance of a single dependent variable into uncorrelated parts. Having uncorrelated parts makes the computations involved in ANOVA incredibly easier. This was important before computers were invented, when calculations were all done by hand, and also were done repeatedly, to check for calculation errors.

The present paper demonstrates that ANOVA effects in a balanced design are perfectly uncorrelated. A mathematical proof that the 4 sums-of-squares partitions (2 main effect, 1 two-way interaction, and error) for a factorial two-way design are all uncorrelated, i.e., sum exactly to the SOS of the dependent variable is presented and a small heuristic data set is included to illustrate the proof.

Fisher invented ANOVA in the 1920's to partition variance of a single dependent variable into uncorrelated parts. Having uncorrelated parts makes the computations involved in ANOVA incredibly easier. This was important before computers were invented, when calculations were all done by hand, and also were done repeatedly, to check for calculation errors.

The present paper demonstrates that ANOVA effects in a balanced design are perfectly uncorrelated. A mathematical proof that the 4 sums-of-squares partitions (2 main effect, 1 two-way interaction, and error) for a factorial two-way design are all uncorrelated, i.e., sum exactly to the SOS of the dependent variable is presented and a small heuristic data set is included to illustrate the proof.

Let A be the independent variable with levels 1,..,j,...a and subjects 1,...,i,...,n. For the one factor case, we can describe the influences responsible for the performance of the i^{th} subject in the j^{th} treatment group by writing the ij^{th} response in terms of the sum of (the overall mean performance of all subjects) and (the difference between the j^{th} treatment mean and the overall mean) and (the unexplained component of the i^{th} subject's score).

The statistical model for the one factor completely randomized design with fixed effects is given by

$$x_{ij} = \mu + \left(\mu_j - \mu\right) + \left(x_{ij} - \mu_j\right),$$

which completely accounts for the ij^{th} response (Kennedy & Bush, 1985).

For ease of notation, let $\alpha_j = (\mu_j - \mu)$ and $\varepsilon_{ij} = \left(x_{ij} - \mu_j\right)$, allowing us to rewrite the model as

$$(1) \qquad x_{ij} = \mu + \alpha_j + \varepsilon_{ij}.$$

Let $\bar{x}_{..} = \hat{\mu} \rightarrow \mu$ and $\bar{x}_{.j} = \hat{\mu}_j \rightarrow \mu_j$ be the least-squares estimators of the population parameters in the above model. (Note that $\cdot$ indicates that the subscript varies over all cases whereas the explicit subscript remains fixed. For example, $\bar{x}_{.j}$ is the mean of the jth treatment group over all subjects 1,...,n.) Thus, the <u>working</u> model is given as

$$(2) \qquad x_{ij} = \bar{x}_{..} + \left(\bar{x}_{.j} - \bar{x}_{..}\right) + \left(x_{ij} - \bar{x}_{.j}\right)$$

where $\alpha_j = \left(\bar{x}_{.j} - \bar{x}_{..}\right)$ and $\varepsilon_{ij} = (x_{ij} - \bar{x}_{.j})$. Note that α_j denotes the effect for the j^{th} level of the independent variable A and that $\sum_j^a \alpha_j = 0$. (The assumption of fixed effects is important for this result). Also, recall that $\varepsilon_{ij} \cap N, ID, \left(0, \sigma^2\right)$.

To generalize to the two-factor case, let A and B be two independent variables with levels 1,...,j,...,a for A, levels 1,...,k,...b for B, and subjects 1,...,i,...,n. Note that the set of all values

$$x_{ijk},\ for\ all\ i = 1,\ldots,n,\ j = 1,\ldots,a\ and\ k = 1,\ldots,b$$

can be thought of as a vector with $n \cdot a \cdot b$ entries. For example, suppose n = 2, a = 2, and b = 3. Then this vector has $2 \cdot 2 \cdot 3 = 12$ entries. We write $X = \left(x_{ijk}\right)$ to stand for the vector having $n \cdot a \cdot b$ entries. This is an $n \times a \times b$ vector, commonly called a tensor, (a tensor can be conceptualized as a 3-dimensional matrix), whose mean is given by

$$\bar{X} = \frac{1}{nab} \sum_{i=1}^{n} \sum_{j=1}^{a} \sum_{k=1}^{b} x_{ijk}.$$

We can now describe and completely account for the ijk^{th} response in a similar manner to the one-factor case ij^{th} response by generalizing the statistical model in equation (1) to

$$(1)^* \qquad x_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}.$$

For ease of notation in writing down the generalized model, we will use the following shorthand:

$$\sum_i = \sum_{i=1}^{n},\ \sum_j = \sum_{j=1}^{a},\ \sum_k = \sum_{k=1}^{b}.$$

It follows that

$$\sum_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{a},\ \sum_{ik} = \sum_{i=1}^{n}\sum_{k=1}^{b},\ \sum_{jk} = \sum_{j=1}^{a}\sum_{k=1}^{b},\ and \sum_{ijk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b}.$$

The least-squares type estimators are now given as

(a) $\bar{x}_{...} = \dfrac{1}{abn} \sum\limits_{ijk} x_{ijk} \rightarrow \mu,$ (b) $\bar{x}_{.j.} = \dfrac{1}{bn} \sum\limits_{ik} x_{ijk} \rightarrow \mu_j,$

(c) $\bar{x}_{..k} = \dfrac{1}{an} \sum\limits_{ij} x_{ijk} \rightarrow \mu_k,$ (d) $\bar{x}_{.jk} = \dfrac{1}{n} \sum\limits_{i} x_{ijk} \rightarrow \mu_{jk}.$

Using these estimators, we can define the components of (1)* as

$$\alpha_j = \left(\bar{x}_{.j.} - \bar{x}_{...}\right), \quad \beta_k = \left(\bar{x}_{..k} - \bar{x}_{...}\right), \quad \alpha\beta_{jk} = \left(\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...}\right),$$

and $\varepsilon_{ijk} = \left(x_{ijk} - \bar{x}_{.jk}\right).$

Using the estimator $\bar{x}_{...}$ for μ and subtracting it from both sides of equation (1)*, we have

$$(*) \qquad x_{ijk} - \bar{x}_{...} = \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}.$$

We will use equation (*) and the least-squares type estimators (a) - (d) to prove the following claim.

Claim: SS(Total) = SS(A) + SS(B) + SS(AB) + SS(Error) where all the SS terms are uncorrelated.

Proof: If we square both sides of (*) and sum over all ijk, we have

$$(**) \qquad \sum_{ijk} (x_{ijk} - \bar{x}_{...})^2 = \sum_{ijk} (\alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk})^2$$

$$= \sum_{ijk} \alpha_j^2 + \sum_{ijk} \beta_k^2 + \sum_{ijk} (\alpha\beta_{jk})^2 + \sum_{ijk} \varepsilon_{ijk}^2$$

(mixed terms) $$+ \quad 2\left[\begin{array}{c} \sum_{ijk}\alpha_j\beta_k + \sum_{ijk}\alpha_j\alpha\beta_{jk} + \sum_{ijk}\alpha_j\varepsilon_{ijk} + \sum_{ijk}\beta_k\alpha\beta_{jk} \\ + \sum_{ijk}\beta_k\varepsilon_{ijk} + \sum_{ijk}\alpha\beta_{jk}\varepsilon_{ijk} \end{array}\right].$$

It is important to note at this point that if it can be shown that each of $\overline{\alpha}, \overline{\beta}, \overline{\alpha\beta}, and\ \overline{\varepsilon}$ is zero, then each of the mixed terms in (**) represents the covariance of two tensors. In general, the covariance of two tensors U and V is given by

$$\text{cov}(U,V) = \sum_{ijk}(u_{ijk} - \overline{U})(v_{ijk} - \overline{V}).$$

Thus, it will suffice to show that each of $\overline{\alpha}, \overline{\beta}, \overline{\alpha\beta}, and\ \overline{\varepsilon}$ is zero and that the mixed terms above is each equal to zero, since U and V are uncorrelated (perpendicular or orthogonal) if and only if

$$\text{cov}(U,V) = 0.$$

<u>Subclaim 1:</u> Each of $\overline{\alpha}, \overline{\beta}, \overline{\alpha\beta}, and\ \overline{\varepsilon}$ is zero.

I. Consider $\overline{\alpha}$.

$$\overline{\alpha} \quad = \quad \frac{1}{abn}\sum_{ijk}\alpha_j \quad = \quad \frac{1}{abn}\sum_{ijk}\left(\overline{x}_{.j.} - \overline{x}_{...}\right)$$

$$= \quad \frac{1}{abn}\sum_{ijk}\overline{x}_{.j.} - \frac{1}{abn}\sum_{ijk}\overline{x}_{...}$$

$$= \frac{1}{abn} nb \sum_j \bar{x}_{.j.} - \frac{1}{abn} abn\, \bar{x}_{...}$$

$$= \frac{1}{a} \sum_j \bar{x}_{.j.} - \bar{x}_{...} \qquad = \qquad \bar{x}_{...} - \bar{x}_{...}$$

$$= \quad 0.$$

II. Consider $\bar{\beta}$.

$$\bar{\beta} = \frac{1}{abn} \sum_{ijk} \beta_k = \frac{1}{abn} \sum_{ijk} \left(\bar{x}_{..k} - \bar{x}_{...} \right)$$

$$= \frac{1}{abn} \sum_{ijk} \bar{x}_{..k} - \frac{1}{abn} \sum_{ijk} \bar{x}_{...}$$

$$= \frac{1}{abn} na \sum_k \bar{x}_{..k} - \frac{1}{abn} abn\, \bar{x}_{...}$$

$$= \frac{1}{b} \sum_k \bar{x}_{..k} - \bar{x}_{...} \qquad = \qquad \bar{x}_{...} - \bar{x}_{...}$$

$$= \quad 0.$$

III. Consider $\bar{\varepsilon}$.

$$\bar{\varepsilon} = \frac{1}{abn} \sum_{ijk} \varepsilon_{ijk} = \frac{1}{abn} \sum_{ijk} \left(\bar{x}_{ijk} - \bar{x}_{.jk} \right)$$

$$= \frac{1}{abn}\sum_{ijk}\bar{x}_{ijk} - \frac{1}{abn}\sum_{ijk}\bar{x}_{.jk}$$

$$= \bar{x}_{...} - \frac{1}{ab}\sum_{jk}\bar{x}_{.jk} \quad = \quad \bar{x}_{...} - \bar{x}_{...}$$

$$= 0.$$

IV. Consider $\overline{\alpha\beta}$.

$$\overline{\alpha\beta} = \frac{1}{abn}\sum_{ijk}\alpha\beta_{jk} = \frac{1}{abn}\sum_{ijk}\left(\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...}\right)$$

$$= \frac{1}{abn}\sum_{ijk}\bar{x}_{.jk} - \frac{1}{abn}\sum_{ijk}\bar{x}_{.j.} - \frac{1}{abn}\sum_{ijk}\bar{x}_{..k} + \frac{1}{abn}\sum_{ijk}\bar{x}_{...}$$

$$= \frac{1}{ab}\sum_{jk}\bar{x}_{.jk} - \frac{1}{a}\sum_{j}\bar{x}_{.j.} - \frac{1}{b}\sum_{k}\bar{x}_{..k} + \bar{x}_{...}$$

$$= \bar{x}_{...} - \bar{x}_{...} - \bar{x}_{...} + \bar{x}_{...}$$

$$= 0.$$

Thus, each of $\bar{\alpha}$, $\bar{\beta}$, $\overline{\alpha\beta}$, *and* $\bar{\varepsilon}$ is zero and subclaim 1 is proved.

Note that all of the possible mixed or combination terms of the four components

$$\alpha_j = \left(\bar{x}_{.j.} - \bar{x}_{...}\right), \quad \beta_k = \left(\bar{x}_{..k} - \bar{x}_{...}\right), \quad \alpha\beta_{jk} = \left(\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...}\right),$$

and $\varepsilon_{ijk} = \left(x_{ijk} - \bar{x}_{.jk}\right)$ from equation (*) are represented in equation (**).

Thus, since each of $\bar{\alpha}$, $\bar{\beta}$, $\overline{\alpha\beta}$, *and* $\bar{\varepsilon}$ is zero, now showing each mixed term equal to zero will consequently show that

$$\alpha \perp \beta,\ \alpha \perp \alpha\beta,\ \alpha \perp \varepsilon,\ \beta \perp \alpha\beta,\ \beta \perp \varepsilon,\ \textit{and}\ \ \alpha\beta \perp \varepsilon.$$

<u>Subclaim 2:</u> Each of the 6 mixed terms in equation (**) is equal to 0.

I. Consider the first mixed term of (**), $\sum_{ijk} \alpha_j \beta_k$.

$$\sum_{ijk} \alpha_j \beta_k \quad = \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \sum_{k=1}^{b} \alpha_j \beta_k \quad = \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} \beta_k \right)$$

$$= \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} (\bar{x}_{..k} - \bar{x}_{...}) \right) \qquad \text{by definition of (*)}$$

$$= \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} \left(\frac{1}{an} \sum_{i=1}^{n} \sum_{j=1}^{a} x_{ijk} - \bar{x}_{...} \right) \right) \qquad \text{by (c)}$$

$$= \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \alpha_j \left(\frac{1}{an} \sum_{k=1}^{b} \sum_{i=1}^{n} \sum_{j=1}^{a} x_{ijk} - b\bar{x}_{...} \right)$$

$$= \quad \sum_{i=1}^{n} \sum_{j=1}^{a} \alpha_j \left(b\bar{x}_{...} - b\bar{x}_{...} \right) \qquad \text{by (a)}$$

$$= \quad 0.$$

II. Consider the second mixed term of (**), $\sum_{ijk} \alpha_j \alpha\beta_{jk}$.

$$\sum_{ijk} \alpha_j \alpha\beta_{jk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b} \alpha_j \alpha\beta_{jk} = \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} \alpha\beta_{jk} \right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} (\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...}) \right) \text{ by definition of } (*)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(\sum_{k=1}^{b} \bar{x}_{.jk} - \sum_{k=1}^{b} \bar{x}_{.j.} - \sum_{k=1}^{b} \bar{x}_{..k} + \sum_{k=1}^{b} \bar{x}_{...} \right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(\frac{1}{n} \sum_{k=1}^{b}\sum_{i=1}^{n} x_{ijk} - \frac{1}{bn} \sum_{k=1}^{b} \sum_{i=1}^{n}\sum_{k=1}^{b} x_{ijk} - \frac{1}{an} \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b} x_{ijk} + b\bar{x}_{...} \right) \text{ by (d),(b),(c)}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(b\bar{x}_{.j.} - \frac{1}{bn} \sum_{k=1}^{b} bn\bar{x}_{.j.} - b\bar{x}_{...} + b\bar{x}_{...} \right) \quad \text{by (b) and (a)}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(b\bar{x}_{.j.} - \sum_{k=1}^{b} \bar{x}_{.j.} \right) = \sum_{i=1}^{n}\sum_{j=1}^{a} \alpha_j \left(b\bar{x}_{.j.} - b\bar{x}_{.j.} \right)$$

$$= 0.$$

III. Consider the third mixed term of (**), $\sum_{ijk} \alpha_j \varepsilon_{ijk}$.

$$\sum_{ijk} \alpha_j \varepsilon_{ijk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b} \alpha_j \varepsilon_{ijk} = \sum_{j=1}^{a}\sum_{k=1}^{b} \alpha_j \left(\sum_{i=1}^{n} \varepsilon_{ijk} \right)$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha_j\left(\sum_{i=1}^{n}(x_{ijk} - \bar{x}_{.jk})\right) \quad \text{by definition of } (*)$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha_j\left(\sum_{i=1}^{n}x_{ijk} - \sum_{i=1}^{n}\bar{x}_{.jk}\right)$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha_j\left(n\bar{x}_{.jk} - n\bar{x}_{.jk}\right) \quad \text{by (d)}$$

$$= 0.$$

IV. Consider the fourth mixed term of (**), $\sum_{ijk}\beta_k\alpha\beta_{jk}$.

$$\sum_{ijk}\beta_k\alpha\beta_{jk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\alpha\beta_{jk} = \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(\sum_{j=1}^{a}\alpha\beta_{jk}\right)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(\sum_{j=1}^{a}(\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k} + \bar{x}_{...})\right) \quad \text{by definition of } (*)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(\sum_{j=1}^{a}\bar{x}_{.jk} - \sum_{j=1}^{a}\bar{x}_{.j.} - \sum_{j=1}^{a}\bar{x}_{..k} + \sum_{j=1}^{a}\bar{x}_{...}\right)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(\frac{1}{n}\sum_{j=1}^{a}\sum_{i=1}^{n}x_{ijk} - \frac{1}{bn}\sum_{j=1}^{a}\sum_{i=1}^{n}\sum_{k=1}^{b}x_{ijk} - \frac{1}{an}\sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{j=1}^{a}x_{ijk} + a\bar{x}_{...}\right) \quad \text{by (d),(b),(c)}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(a\bar{x}_{..k} - a\bar{x}_{...} - \frac{1}{an}\sum_{j=1}^{a}an\bar{x}_{..k} + a\bar{x}_{...}\right) \quad \text{by (c),(a), (c)}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(a\bar{x}_{..k} - \sum_{j=1}^{a}\bar{x}_{..k}\right) = \sum_{i=1}^{n}\sum_{k=1}^{b}\beta_k\left(a\bar{x}_{..k} - a\bar{x}_{..k}\right)$$

$$= 0.$$

V. Consider the fifth mixed term of (**), $\sum_{ijk}\beta_k\varepsilon_{ijk}$.

$$\sum_{ijk}\beta_k\varepsilon_{ijk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\varepsilon_{ijk} = \sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\left(\sum_{i=1}^{n}\varepsilon_{ijk}\right)$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\left(\sum_{i=1}^{n}(x_{ijk} - \bar{x}_{.jk})\right) \quad \text{by definition of (*)}$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\left(\sum_{i=1}^{n}x_{ijk} - \sum_{i=1}^{n}\bar{x}_{.jk}\right) = \sum_{j=1}^{a}\sum_{k=1}^{b}\beta_k\left(n\bar{x}_{.jk} - n\bar{x}_{.jk}\right) \quad \text{by (d)}$$

$$= 0.$$

VI. Consider the sixth mixed term of (**), $\sum_{ijk}\alpha\beta_{jk}\varepsilon_{ijk}$.

$$\sum_{ijk}\alpha\beta_{jk}\varepsilon_{ijk} = \sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b}\alpha\beta_{jk}\varepsilon_{ijk} = \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha\beta_{jk}\left(\sum_{i=1}^{n}\varepsilon_{ijk}\right)$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha\beta_{jk}\left(\sum_{i=1}^{n}(x_{ijk} - \bar{x}_{.jk})\right) \quad \text{by definition of (*)}$$

$$= \sum_{j=1}^{a}\sum_{k=1}^{b}\alpha\beta_{jk}\left(\sum_{i=1}^{n}x_{ijk} - \sum_{i=1}^{n}\bar{x}_{.jk}\right)$$

$$= \sum_{j=1}^{a} \sum_{k=1}^{b} \alpha\beta_{jk}\left(n\bar{x}_{.jk} - n\bar{x}_{.jk}\right) \qquad \text{by (d)}$$

$$= \quad 0.$$

Thus, since all mixed terms of (**) equal 0,

$$\sum_{ijk}\left(x_{ijk} - \bar{x}_{...}\right)^2 = \sum_{ijk}\alpha_j^2 + \sum_{ijk}\beta_k^2 + \sum_{ijk}\alpha\beta_{jk}^2 + \sum_{ijk}\varepsilon_{ijk}^2$$

which is the mathematical equivalent of

SS(Total) = SS(A) + SS(B) + SS(AB) + SS(Error), since each of

$\overline{\alpha}$, $\overline{\beta}$, $\overline{\alpha\beta}$, and $\overline{\varepsilon}$ is zero.

Consequently, since all possible covariance combinations equal 0,

$$\alpha \perp \beta,\ \alpha \perp \alpha\beta,\ \alpha \perp \varepsilon,\ \beta \perp \alpha\beta,\ \beta \perp \varepsilon,\ and\ \alpha\beta \perp \varepsilon .$$

Thus, the 4 sums-of-squares partitions (2 main effects, 1 two-way interaction, and error) for a completely randomized factorial two-way design with fixed effects are all uncorrelated.

Appendix A

This appendix consists of an example using a small heuristic data set and calculations illustrating how to work through the proof.

Example: 18 students, 9 male and 9 female, are distributed randomly among 3 training conditions: cooperative learning, lecture and control. Let A be the independent variable representing gender and B be the independent variable representing training condition. This example represents a two-way (2 x 3) balanced design where A has 2 levels and B has 3 levels.

Let Y be the dependent variable representing grade/performance on a 10 point test over the chosen topic. The following table represents test scores as a function of training condition and gender.

	Training Condition	Male ($j = 1$)	Female ($j = 2$)
($k = 1$)	Cooperative Learning	5 6 7	8 8 9
($k = 2$)	Lecture	7 9 9	4 5 6
($k = 3$)	Control	2 3 4	2 3 6

number of subjects per group by gender ($n = 3$) $i = 1, \ldots, 3$

A - gender ($a = 2$) $j = 1, 2$

B - training condition ($b = 3$) $k = 1, \ldots, 3$

$a \cdot b \cdot n = 18$

$$\bar{x}_{...} = \frac{1}{abn}\sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b} x_{ijk} = \frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3} x_{ijk}$$

$$= \frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{2}\left(x_{ij1} + x_{ij2} + x_{ij3}\right) = \frac{1}{18}\sum_{i=1}^{3}\left[\left(x_{i11} + x_{i21}\right) + \left(x_{i12} + x_{i22}\right) + \left(x_{i13} + x_{i23}\right)\right]$$

$$= \frac{1}{18}\left[\begin{array}{c}\left(x_{111} + x_{121} + x_{112} + x_{122} + x_{113} + x_{123}\right) + \left(x_{211} + x_{221} + x_{212} + x_{222} + x_{213} + x_{223}\right) \\ + \left(x_{311} + x_{321} + x_{312} + x_{322} + x_{313} + x_{323}\right)\end{array}\right]$$

$$= \frac{1}{18}[(5 + 8 + 7 + 4 + 2 + 2) + (6 + 8 + 9 + 5 + 3 + 3) + (7 + 9 + 9 + 6 + 4 + 6)]$$

$$= \frac{103}{18} = 5.7222222$$

$$\bar{x}_{.j.} = \frac{1}{9}\sum_{i=1}^{3}\sum_{k=1}^{3} x_{ijk} \Rightarrow \begin{cases} \bar{x}_{.1.} = \frac{1}{9}(52) = 5.7777778 \\ \bar{x}_{.2.} = \frac{1}{9}(51) = 5.6666667 \end{cases}$$

$$\bar{\alpha} = \frac{1}{18}\sum_{i=1}^{3}\sum_{k=1}^{3}\sum_{j=1}^{2}(\bar{x}_{.j.} - \bar{x}_{...}) = \frac{1}{18}\sum_{i=1}^{3}\sum_{k=1}^{3}[(\bar{x}_{.1.} - \bar{x}_{...}) + (\bar{x}_{.2.} - \bar{x}_{...})]$$

$$= \frac{1}{18}\sum_{i=1}^{3}\sum_{k=1}^{3}((5.7777778 - 5.7222222) + (5.6666667 - 5.722222))$$

$$= \frac{1}{18}\sum_{i=1}^{3}\sum_{k=1}^{3}(.0555556 - .0555556) = 0$$

$$\bar{x}_{..k} = \frac{1}{6}\sum_{i=1}^{3}\sum_{j=1}^{2} x_{ijk} \Rightarrow \begin{cases} \bar{x}_{..1} = \frac{1}{6}(5+6+7+8+8+9) = \frac{1}{6}(43) = 7.166666667 \\ \bar{x}_{..2} = \frac{1}{6}(7+9+9+4+5+6) = \frac{1}{6}(40) = 6.666666667 \\ \bar{x}_{..3} = \frac{1}{6}(2+3+4+2+3+6) = \frac{1}{6}(20) = 3.333333333 \end{cases}$$

$$\sum_{i=1}^{3}\sum_{j=1}^{2}\alpha_j\left(\sum_{k=1}^{3}(\bar{x}_{..k}-\bar{x}_{...})\right) = \sum_{i=1}^{3}\sum_{j=1}^{2}\alpha_j\left[\begin{array}{r}(7.16666667-5.72222222)\\ +(6.66666667-5.7222222)\\ +(3.33333333-5.72222222)\end{array}\right]$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{2}\alpha_j(1.44444447+.94444447-2.38888887)$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{2}\alpha_j(0) = 0.$$

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Benton, R.L. (1991). Statistical power considerations in ANOVA. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 119-132). Greenwich, CT: JAI Press.

Carter, D.S. (1979). Comparison of different shrinkage formulas in estimating population multiple correlation coefficients. Educational and Psychological Measurement, 39, 261-266.

Clayton, K. N. (1984). An introduction to statistics. Columbus, OH: Merrill.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Crask, M.R., & Perreault, W.D., Jr. (1977). Validation of discriminant analysis in marketing research. Journal of Marketing Research, 14, 60-68.

Cronbach, L.J. (1957). The two disciplines of psychology. American Psychologist, 12, 671-684.

Cronbach, L.J. (1975). Beyond the two disciplines of psychology. American Psychologist, 30, 116-127.

Daniel, L.G. (1989, January). Use of the jackknife statistic to establish the external validity of discriminant analysis results. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 305 382)

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

Eason, S.H., & Daniel, L.G. (1989, January). Trends and methodological practices in several cohorts of dissertations. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston. (ERIC Document Reproduction Service No. ED 306 299)

Edgington, E.S. (1974). A new tabulation of statistical procedures in APA journals. American Psychologist, 29, 25-26.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Fan, Xitao. (1992, April). Canonical correlation analysis as a general data-analytic model. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 348 383)

Fornell, C. (1978). Three approaches to canonical analysis. Journal of the Market Research Society, 20, 166-181.

Gage, N.L. (1963). Paradigms for research on teaching. In N.L. Gage (Ed.), Handbook of research on teaching (pp. 94-141). Chicago: Rand McNally.

Glasnapp, D. R., & Poggio, J. P. (1985). Essentials of statistical analysis for the behavioral sciences. Columbus, OH: Merrill.

Goodwin, L.D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Gravetter, F. J., & Wallnau, L. B. (1985). Statistics for the behavioral sciences. St. Paul, MN: West.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Huberty, C.J. (1987). On statistical testing. Educational Researcher, 16(8), 4-9.

Huberty, C.J, & Morris, J.D. (1988). A single contrast test procedure. Educational and Psychological Measurement, 48, 567-578.

Hudson, W.D. (1969). The is/ought question. London: MacMillan.

Humphreys, L.G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 70, 873-876.

Humphreys, L.G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. Journal of Educational Psychology, 66, 464-472.

Kaiser, H.F. (1976). Review of Factor analysis as a statistical method.

Educational and Psychological Measurement], 36, 586-589.
Kennedy, J., & Bush, A. (1985). An introduction to the design and analysis of experiments in behavioral research. Washington, DC: University Press of America.
Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.
Keppel, G., & Zedeck, S. (1989). Data analysis for research designs. New York: W. H. Freeman.
Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.
Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole. (pp. 69-98)
Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.
Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. American Psychologist, 43, 635-642.
Lincoln, Y.S., & Guba, E.G. (1985). Naturalistic inquiry. Beverly Hills: SAGE.
Lunneborg, C.E. (1987). Bootstrap applications for the behavioral sciences. Seattle: University of Washington.
Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.
Minium, E. W., & Clarke, R. B. (1982). Elements of statistical reasoning. New York: John Wiley and Sons.
Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
Patton, M.Q. (1975). Alternative evaluation research paradigm. Grand Forks: University of North Dakota Press.
Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.
Rosnow, R.L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.
Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). Human experimental psychology. New York: Oxford University Press.
Snyder, P., & Lawson, S. (in press). Evaluating statistical significance using corrected and uncorrected magnitude of effect size estimates. In B. Thompson (Guest Editor) of the special issue on statistical significance testing, Journal of Experimental Education.
Sowell, E. J., & Casey, R. J. (1982). Research methods in education. Belmont, CA: Wadsworth.
Spence, J. T., Cotton, J. W., Underwood, B. J., & Duncan, C. P. (1983). Elementary statistics (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
Strike, K.A. (1979). An epistemology of practical research. Educational Researcher, 8(1), 10-16.
Swaminathan, H. (1989). Interpreting the results of multivariate analysis of variance. In B. Thompson (Ed.), Advances in social science methodology (pp. 205-232, Vol. 1). Greenwich, CT: JAI Press.
Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Beverly Hills: SAGE.
Thompson, B. (1987a). Review of Foundations of behavioral research (3rd ed.) by F. Kerlinger. Educational Research and Measurement, 47, 1175-1181.
Thompson, B. (1987b, April). The use (and misuse) of statistical significance testing: Some recommendations for improved editorial policy and practice. Paper presented at the annual meeting of the American Education Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 287 868)
Thompson, B. (1988a, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)
Thompson, B. (1988b, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the

Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)
Thompson, B. (1988c). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.
Thompson, B. (1988d). Misuse of chi-square contingency table test statistics. Educational and Psychological Research, 8, 39-49.
Thompson, B. (1988e). Program FACSTRAP: A program that computes bootstrap estimates of factor structure. Educational and Psychological Measurement, 48, 681-686.
Thompson, B. (1988f). Review of Analyzing multivariate data by N. Cliff. Educational and Psychological Measurement, 48, 1129-1135.
Thompson, B.(1989a). Asking "what if" questions about significance tests. Measurement and Evaluation in Counseling and Development, 22, 66-68.
Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-6.
Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationship: A Monte Carlo study. Educational and Psychological Measurement, 50, 15-31.
Thompson, B. (1991a). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.
Thompson, B. (1991b). Review of Data analysis for research designs by G. Keppel & S. Zedeck. Educational and Psychological Measurement, 51, 500-510.
Thompson, B. (1992a). DISCSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.
Thompson, B. (1992b). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.
Thompson, B. (in press). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality.
Tucker, M.L. (1991). A compendium of textbook views on planned versus post hoc tests. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 107-118). Greenwich, CT: JAI Press.
Thompson, B. (1993, April). The General Linear Model (as opposed to the classical ordinary sums of squares) approach to analysis of variance should be taught in introductory statistical methods classes. Paper presented at the annual meeting of the American Educational Research Association, Atlanta. (ERIC Document Reproduction Service No. ED 358 134)
Tuthill, D., & Ashton, P. (1983). Improving educational research through the development of educational paradigms. Educational Researcher, 12(10), 6-14.
Haase, T., & Thompson, B. (1992, January). The homogeneity of variance assumption in ANOVA: What it is and why it is required. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston.
Willson, V. L. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 222 522)