Which Sums of Squares Are Best In Unbalanced Analysis of Variance?

Donald B. Macnaughton*

Note added on May 10, 1998: A few months after I originally published this paper I discovered a particular (infrequently occurring) situation in which the recommended HTO approach to unbalanced analysis of variance is invalid. (The HTOS approach, which I recommend in appendix D as an extension of the HTO approach, is valid in this situation.) I shall discuss this matter in a forthcoming paper.

Three fundamental concepts of science and statistics are entities, variables (which are formal representations of properties of entities), and relationships between variables. These concepts help to distinguish between two uses of the statistical tests in analysis of variance (ANOVA), namely

- to test for relationships between the response variable and the predictor variables in an experiment
- to test for relationships among the parameters of the model equation in an experiment.

Two methods of computing ANOVA sums of squares are:

- Higher-level Terms are Omitted from the generating model equations (HTO = SPSS ANOVA EXPERIMENTAL ≈ SAS Type II ≈ BMDP4V with Weights are Sizes)
- Higher-level Terms are Included in the generating model equations (HTI = SPSS ANOVA UNIQUE = SPSS MANOVA UNIQUE = SAS Type III = BMDP4V with Weights are Equal = BMDP2V = MINITAB GLM = SYSTAT MGLH = Data Desk Type 3).

This paper evaluates the HTO and HTI methods of computing ANOVA sums for squares for fulfilling the two uses of the ANOVA statistical tests. Evaluation is in terms of the hypotheses being tested and relative power. It is concluded that (contrary to current practice) the HTO method is generally preferable when a researcher wishes to test the results of an experiment for evidence of relationships between variables.

KEY WORDS: Relationships between variables; Relationships among parameters; Philosophy of ANOVA; Power of ANOVA.

1. INTRODUCTION

Methods of computing analysis of variance (ANOVA) sums of squares for unbalanced experiments were introduced by Yates in 1934. Recently there has been controversy over which method is “best”. This expository paper addresses some areas of the controversy.

Section 2 gives references to earlier work. Sections 3 - 6 discuss preparatory topics. Sections 7 - 11 describe two uses of the ANOVA statistical tests. Sections 12 and 13 describe two methods of computing ANOVA sums of squares. Sections 14 - 18 evaluate the two methods for fulfilling the two uses. Five appendices extend the ideas.

I hope that knowledgeable readers will indulge my initial concentration (in sections 3 - 6) on some fundamental concepts of human thought. These concepts may at first seem trivial or obvious. However, these concepts deserve careful study because they are foundations for many other concepts in science and statistics, including the conclusions of this paper.

Most of the discussion that follows is non-mathematical. However, readers who enjoy a mathematical crescendo may get some satisfaction from the elegant simplicity of works of three statisticians excerpted in appendix C.

2. HISTORY

Readers wishing to trace the development of ideas about unbalanced ANOVA will find the works by the following authors of interest (given here in chronological order): Yates (1934), Kempthorne (1952), Scheffé (1959:112-119), Elston and Bush (1964), Gosslee and Lucas (1965), Bancroft (1968), Overall and Spiegel (1969), Speed (1969), Francis (1973), Urquhart, Weeks, and Henderson (1973), Appelbaum and Cramer (1974), Burdick, Herr, O’Fallon, and O’Neill (1974), Carlson and Timm (1974), Kutner (1974), Kutner (1975), Hocking and Speed (1975), Overall, Spiegel, and Cohen (1975), Golhar and Skillings (1976), Hocking and Speed (1976), Keren and Lewis (1976), O’Brien (1976), Speed and Hocking (1976), Heiberger and Laster (1977), Nelder (1977), Aitkin (1978), Herr and Gaebelein (1978), Hocking, Hackney, and Speed (1978), Speed, Hocking, and Hackney (1978), Urquhart and Weeks (1978), Burdick (1979), Frane (1979), Bryce, Scott, and Carter (1980), Burdick and Herr (1980), Cramer and Appelbaum (1980), Goodnight (1980), Hocking, Speed, and Coleman (1980), Searle (1980), Steinhorst and Everson (1980), Overall, Lee, and Hornick (1981), Rubin, Stroud, and Thayer (1981), Searle (1981), Searle, Speed, and Henderson

*Donald B. Macnaughton is president of MatStat Research Consulting Inc. 246 Cortleigh Blvd., Toronto Ontario, Canada M5N 1P7. Email: donmac@matstat.com This article is a revised version of a paper presented at the Joint Statistical Meetings in Boston, August 1992. The author thanks Donald S. Burdick for the proof of theorem 2, and Donald S. Burdick, Donald F. Burrill, Ramon C. Littell, and Jacqueline Oler for helpful comments.

(1981), Spector, Voissem, and Cone (1981), Calinski (1982), Howell and McConaughy (1982), Nelder (1982), Steinhorst (1982), Aitkin (1983), Littell and Lynch (1983), Schmoyer (1984), Johnson and Herr (1984), Hocking (1985), Elliott and Woodward (1986), Pendleton, Von Tress, and Bremer (1986), Finney (1987), Knoke (1987), Milligan, Wong, and Thompson (1987), Searle (1987), Helms (1988), Singh and Singh (1989), Turner (1990), and Macdonald (1991). An excellent overview and further early references are given by Herr (1986).

3. ENTITIES

If you observe your train of thought, you will probably agree that you think about various "things". For example, during the next few moments you might think about, among other things, a friend, an appointment, today's weather, and an idea. Each of these things is an example of an *entity*.

The concept of *entity* is perhaps the broadest of all human concepts, because literally everything (whether it exists or not) is an instance of an entity. Table 1 illustrates the broadness by listing a variety of entity types.

TABLE 1

A List of Some Common Entity Types
(some categories overlap)

- physical objects (examples: trees, automobiles, protons)
- processes (examples: a leaf blowing in the wind, a machine building another machine, a chemical reaction)
- events
- organisms
- minds
- symbols
- forces (examples: force needed to lift a physical object, magnetic forces)
- mathematical entities (examples: sets, functions, numbers, spaces, vectors)
- relationships between entities
- properties of entities

Most people view most entities as existing in two different places: in the external world and in our minds. We use the entities in our minds mainly to stand for the entities in the external world. This helps us to understand the external world. In language we represent entities with nouns.

The concept of *entity* does not often appear in scientific or statistical discussions because it is not often necessary to discuss things at such a general level. Instead, discussions usually concern one or more *types* of entities, which are best referred to by their type names. (For example, medical scientists often study a of type of entity called *human beings*.) Or discussions may refer to one or more individual entities, which are best referred to by their individual names. However, it is useful to be aware of the central role that the concept of *entity* plays in human thought.

4. PROPERTIES OF ENTITIES

Associated with every entity are attributes or *properties.* Table 2 lists some entity types and some of the properties associated with entities of each type.

TABLE 2

Entity Types with Examples of
Some of Their Properties

Entity Type	Properties of Entities of this Type
physical objects	weight chemical composition age
persons	height intelligence quotient blood type political affiliation whether presently alive
forces	magnitude direction locus of application
national economies	gross national product cost of living rate of inflation
populations	size proportion of the population having a specified level of a property
events	probability of occurrence whether occurred duration
works of art	beauty

Properties are an important aspect of entities because we can only know or experience an entity by knowing or experiencing its properties.

Kendall, Stuart, and Ord (1987:1.1-1.3) discuss the role of properties in statistics.

5. VALUES OF PROPERTIES OF ENTITIES

For any particular entity, each of its properties has a *value.* People usually report the value of a property of an entity by one or more words or by a number. For example, table 3 lists some of the properties and the associated values for the entity known as the United Nations Building in New York City.

TABLE 3

Properties of the United Nations Building
and Their Associated Values

Property	Value of the Property
height	tall (i.e., the word *tall*)
height in meters	165.8
primary building materials	concrete, glass, steel

In language we often use adjectives and adverbs to report the values of properties. For example, we might use the adjective *tall* to report (the value of) the height (property) of a building, or the adverb *quickly* to report (the value of) the speed (property) of (the process of) someone running in a race.

Adjectives and adverbs are useful for reporting the values of properties because they are compact—within a single word we can both identify the property of interest and indicate a particular value of it. However, adjectives and adverbs are also imprecise. If we need higher precision in the report of the value of a property, we can use numbers because numbers can represent any degree of precision we wish.

If we wish to determine the value of a property of an entity, we can apply an appropriate measuring instrument to the entity. If the instrument is measuring correctly, it will return a value to us that is an estimate of the value of the property in the entity at the time of the measurement.

References to values of properties of entities are such a fundamental part of people's thinking that we usually make these references automatically, without being specifically aware that we are using the general concept of a value of a property of an entity. Therefore, the importance of values of properties of entities in models of the external world may be sometimes underestimated.

Computer models of the external world are playing increasingly important roles in science, business, and government. Therefore, it is interesting to note the central roles that entities, properties, and values play in such models:

- Each table in a computer database is associated with a different type of entity about which the user of the database wishes to keep information. For any given table, the rows of the table represent different instances of entities of the type associated with the table. The columns represent different properties of entities of this type. And the intersection of a row and a column contains the value of the property associated with the column for the particular entity that is associated with the row.
- The concepts of entities and properties appear directly in knowledge representation in artificial intelligence (expert) systems where they are sometimes organized into "semantic networks" or "frames".
- Entities and properties appear in object-oriented programming languages in the sense that the objects and attributes (variables) of such languages are simply entities (or sometimes classes of entities) and properties respectively.

6. RELATIONSHIPS BETWEEN PROPERTIES (RELATIONSHIPS BETWEEN VARIABLES)

6.1 Science as a Study of Relationships Between Properties

In view of the broad generality of the concepts of entities and properties, it is helpful to consider scientific research in terms of those concepts. In those terms, much of science can be seen as a study of *relationships between properties of entities.*

One can characterize a relationship between properties as follows:

> There is a *relationship* in entities between a property y and one or more other properties x_1, x_2, ..., x_p if any of the following (equivalent) conditions are satisfied:
>
> - the measured value of y in the entities "depends" on the measured values of the x's in the entities or
> - the measured value of y in the entities varies wholly or partially "in step" with the measured values of the x's in the entities or
> - y is some *function* of the x's in the entities—that is
>
> $$y = f(x_1, x_2, \ldots, x_p) + \varepsilon \qquad (1)$$
>
> (where I discuss the term ε below).

For example, medical scientists have discovered that there is a relationship in humans between the property "concentration of insulin in the bloodstream" and the property "rate of carbohydrate metabolism". Specifically, as insulin in the bloodstream increases (within a certain range), carbohydrate metabolism also increases.

Scientists often summarize their findings of a relationship between properties with a graph, such as figure 1.

We can see the generality of the concept of a relationship between properties of entities if we examine the so-called laws of science, and if we note that many of these laws are statements of relationships between properties of entities. For example, the ideal gas law, $PV = nRT$, that relates pressure *(P),* volume *(V),* amount *(n),* and temperature *(T)* of an ideal gas is a statement of a relationship between certain properties of a mass of gas. (The R is the constant of proportionality.)

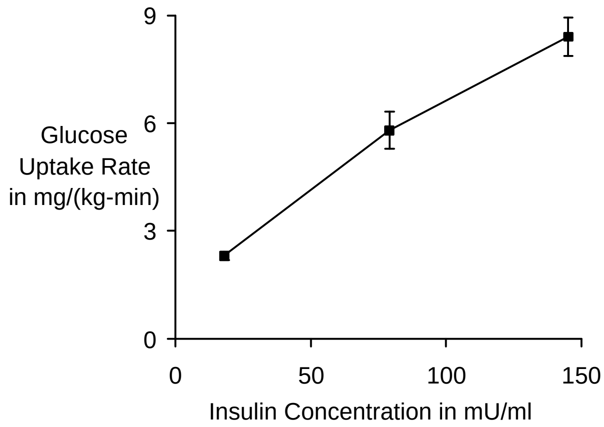

Figure 1. A graph showing the relationship between carbohydrate metabolism (specifically glucose uptake) and insulin concentration for fifteen normal young adults. Data are from an experiment by Gottesman et al (1982). The height of each black square indicates the mean glucose uptake when the subjects were maintained at the insulin concentration shown on the horizontal axis directly beneath the square. (Plasma glucose was maintained at approximately 92 mg/dl throughout.) The horizontal bars show plus and minus the standard error of each glucose uptake mean.

TABLE 4

Classification of "Laws" Defined in *Dictionary of Scientific Terms*

Type of Statement	%	Count*
relationship between properties	75	184
non-relationship between properties (including 10 conservation laws)	11	27
law of mathematics (axiom or theorem)	6	14
relationship between entities	4	9
value of a property	2	5
distribution of the values of a property	2	4
existence of a property	1	2
existence of an entity	<1	1
other	0	0

*The sum of the counts is greater than 213 because some laws contained two or more independent statements, and each such statement was classified separately. For some laws, instances of the one or more statements that constitute the law could sometimes, by taking different points of view, be classified in more than one of the first eight ways. In those cases, the count reflects the way of classifying the statement that was judged easiest to understand.

Similarly, Einstein's equation, $E = mc^2$ is a statement of a relationship between two properties of matter: contained energy *(E)* and mass *(m)*. (The c^2 is the constant of proportionality, which Einstein has shown to be equal to the square of the speed of light.)

To explore the generality of the concept of a relationship between properties, two assistants each scanned each page of the 2088-page *McGraw-Hill Dictionary of Scientific and Technical Terms* (Parker 1989) for entries that contain the word *law* in the definiendum. They found 213 entries that define different "laws" of science. For each entry I then tried to express the definition in terms of the concepts of entities, properties, and relationships between properties. This yielded the classification shown in table 4.

The most common type of statement in science—a statement of a relationship between properties—is also the most important because knowledge of relationships between properties gives us the ability to predict (and sometimes control) the values of properties in new similar entities, and such ability is often of great value. For example, knowledge of the relationship between "concentration of insulin in the bloodstream" and "rate of carbohydrate metabolism" in humans has helped doctors to control diabetes, a disease that is characterized by poor carbohydrate metabolism.

Summarizing: Study of relationships between properties of entities is a central activity of science because knowledge of such relationships gives us the ability to predict and control (values of) properties, and such ability is often of great value.

6.2 Properties as Variables

Bypassing some details (see Macnaughton 1997), we can roughly say that scientists and statisticians usually refer to properties as *variables*. Much of statistics is aimed at providing techniques to help scientists study relationships between variables, whether through the *t*-test, ANOVA, regression, exploratory data analysis, nonparametric analysis, categorical analysis, time series analysis, survey analysis, factor analysis, correspondence analysis, or various other techniques.

In the rest of the paper I use standard terminology and discuss properties and relationships between properties mainly in terms of variables and relationships between variables. However, readers new to the ideas may find it helpful to keep in mind that the variables that are discussed in science are simply representations of properties of entities. And the *value* of any variable represents the value of the property in the associated entity (usually at a particular time).

(In the physical sciences, properties, variables, or values are sometimes called *physical quantities.*)

Barnett (1988) gives a general discussion of the concept of a relationship between variables.

6.3 Response Variables and Predictor Variables

In studying relationships between variables, scientists often classify the variables in a research project into response variables and predictor variables. The *response* variables are the variables that the scientist would like to discover how to control (or at least predict). The *predictor* variables are the variables that the scientist will control (or just measure) in an attempt to discover how to control (predict) the values of the response variables.

For example, when medical scientists study the relationship between "concentration of insulin in the bloodstream" and "rate of carbohydrate metabolism", they view "rate of carbohydrate metabolism" as the response variable because they wish to learn how to control carbohydrate metabolism by controlling insulin concentration (and not the other way around).

Most scientists find it efficient to concentrate on learning how to control or predict a single variable at a time in the entities they are studying. Therefore, without loss of relevant generality, this paper discusses research projects (or standalone units of analysis in research projects) that have a single response variable (which may be measured just once or repeatedly in entities) and one or more predictor variables.

6.4 A Definition of a Relationship Between Properties (Variables)

The characterization of a relationship between properties given in section 6.1 is statistically weak so, since we need a statistical definition below, let us convert the characterization into such a definition. I begin with two preliminary definitions:

Definition: The *expected value* of a variable in a population is the average value of the variable across all the entities in the population under a given set of conditions.

Definition: The *expected value of a variable* in a population *conditioned on the values of one or more other variables* is the average value of the variable across all the entities in the population under a given set of conditions including the condition that the other variables have particular stated values.

I shall represent the expected value of some variable y as $E(y)$, and I shall represent the expected value of y conditioned on the values of variables $x_1,\ldots,x_p$ as $E(y|x_1,\ldots,x_p)$.

Using the concept of expected value, let us now consider a formal definition of a relationship between properties (variables):

Definition: If y is a variable that reflects a measured property of entities in some population, and if $x_1,\ldots,x_p$ are a set of one or more other variables that reflect distinct other measured properties of the entities (or of the entities' environment), then a *relationship* exists in the entities between y and the $x_1,\ldots,x_p$ if, for each integer i, where $1 \le i \le p$

$$E(y|x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_p)$$
$$\ne E(y|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p).$$

If $p = 1$, the inequality simplifies to

$$E(y|x_1) \ne E(y).$$

Each of the p inequalities is deemed to be satisfied if there is at least one set of specific values of the *x*s that satisfies the inequality. (Notes: (1) A different set of values of the *x*s may be used for each inequality; (2) if y is not a numeric variable, then for a relationship to exist, the p inequalities must each be satisfied for at least one recoding of the values of y into numeric values. A different recoding may be used for each inequality.)

Note that this definition of a relationship between properties is operationally equivalent to the characterizations of a relationship given in section 6.1 because if a state of affairs satisfies the definition, it will also satisfy any of the characterizations, and vice versa.

Other mathematical definitions of a relationship between properties are given (in terms of causal relationships) by Granger (1980), Chowdhury (1987), and Poirier (1988).

6.5 The Null and Alternative Hypotheses

For any type of entity and any response variable and any set of one or more predictor variables, exactly one of the following two hypotheses is true:

- *Null Hypothesis:* there is *no* relationship between the response variable and any of the predictor variables in the population of entities of this type
- *Alternative Hypothesis:* there *is* a relationship between the response variable and one or more of the predictor variables in the population.

Scientists usually begin study of a relationship between variables with the *formal* assumption that null hypothesis is true. (Informally we usually suspect and hope that the alternative hypothesis is true or there would be no point in seeking evidence of relationships in the particular set of variables we are studying.) The practice of beginning with the (impossible-to-prove) assumption that the null hypothesis is true is entailed by the principle of parsimony, which tells us to keep things as simple as possible. The simplest situation is that of no relationship, so we begin with that assumption.

After making the assumption that the null hypothesis is true, scientists who wish to study a relationship then perform an appropriate research project in an attempt to invalidate the assumption. If the results of the research project show reasonable evidence that a relationship exists, then the scientific community (through informal consensus) rejects the null hypothesis and concludes that a relationship between variables similar to that suggested by the results probably actually exists.

Tukey (1989:176) suggests that there may be a relationship (albeit sometimes very weak) in entities between *all* measurable pairs of variables, regardless of the variables' identities. Given that point, we may ask why it is necessary to begin with the assumption that the null hypothesis is true. Researchers begin initial study of a relationship between variables with the (formal) assumption that there is no relationship in order to avoid the problem of thinking that they know more about a relationship than they actually do. And only if all of the following conditions are satisfied do careful researchers accept the existence of a particular relationship between variables:

- someone has performed an appropriate empirical research project
- the research project has found reasonable (see below) evidence of a relationship between the variables
- the research project has been carefully scrutinized for errors (and perhaps replicated) by the scientific community (and anyone else who is interested)
- nobody has been able to come up with a reasonable alternative explanation of the results of the research project (see Mosteller 1990, Lipsey 1990, Macnaughton 1997).

6.6 Experiments and Causation

When scientists can control ("manipulate") the values of variables (as opposed to being able only to observe them), they usually study relationships between variables by performing "experiments".

> *Definition:* An *experiment* consists of our manipulating one or more predictor variables in one or more entities while we observe a response variable in the entities.

Of course, we manipulate a variable in entities by somehow causing the variable to have certain values of our choosing in the entities. For example, in a medical experiment we can manipulate the concentration of insulin in the bloodstreams of patients by administering differing amounts of insulin to them.

When feasible, experiments are preferred to observational (i.e., non-manipulative) research projects because relationships between variables found in experiments usually allow us to infer causation while relationships between variables found in observational research projects usually do not allow us to (confidently) infer causation—they only allow us to infer association.

This paper concentrates on analyzing the results of experiments. And although some of the points below also apply to non-experimental (i.e., observational) research projects, interpretation of such research projects is more difficult, and beyond the present scope.

Landmark books about the design and analysis of experiments are by Fisher (1935 [1990]), Kempthorne (1952), Cochran and Cox (1957 [1992]), Cox (1958 [1992]), Finney (1960), Winer (1971), and Box, Hunter, and Hunter (1978).

6.7 ANOVA

If certain often-satisfiable assumptions (discussed below) are adequately satisfied, ANOVA is well suited to analyze the results of an experiment to help determine if there is evidence of a relationship between the response variable and one or more of the predictor variables.

ANOVA works by taking as input the results of a properly-carried-out experiment. That is, the input is the (ordered) set of values of the response and predictor variables that were measured in the entities that participated in the experiment. Through a mathematical procedure, ANOVA provides as its output a set of numbers called *p*-values. A *p*-value is:

the probability of obtaining the evidence
(or stronger evidence)
available from the results of the experiment
that the particular type of relationship
(between the response variable
and the predictor variables)
that is suggested by the results
actually exists
if in fact there is no such relationship.

Thus the lower a *p*-value, the more improbable it is that the obtained result would be obtained if there is no relationship. Thus if a *p*-value for a relationship is low enough, we can tentatively reject the null hypothesis and conclude that there is a relationship between the response variable and the relevant predictor variable(s).

In evaluating the work of others, scientists often use the reasonable convention that a *p*-value must be less than a critical value of .05 (or sometimes .01) before they will reject the null hypothesis and tentatively conclude that the relationship that is associated with a *p*-value actually exists.

The practice of computing a *p*-value and examining it to determine whether it is less than a critical value is called a *statistical test* of the hypothesis that the associated relationship exists.

Once we have concluded (from a statistical test or otherwise) that a particular relationship exists, we can then use our knowledge of the relationship to help us make predictions or exercise control.

6.8 Does ANOVA Detect Relationships Between Variables?

Some readers may question whether we can view the use of the statistical tests in ANOVA (or in its simplest incarnation, the *t*-test) as a means for detecting relationships between variables. For example, consider a research project that uses a *t*-test to check if there evidence of a difference between women and men in their response to some standard form of medical treatment (as measured by some medically acceptable measure of the response). Most readers will agree that a response variable is clearly present in this research project, namely the measured value of the "response" to the treatment for each person who participates in the research project. But some readers may question whether there is a predictor variable, or whether it is useful to view this research project in terms of seeking evidence of a relationship between variables.

We can answer these questions by first noting that the predictor variable in the research project is the variable "gender", which reflects an important property of the patients. ("Gender" is a variable in the sense that for any patient in the research project this variable has a particular value, namely, "female" or "male", and the values *vary* somewhat from patient to patient.) Thus we can view the research project as an attempt to see if there is a relationship in humans between the variables "gender" and "response".

Similarly, in an *n*-way ANOVA there are *n* different predictor variables, each of which represents a different property of the entities that are under study, or of the entities' environment. (Of course, in the case of a treatment that is *applied* to the entities, the associated predictor variable reflects a property of the entities *after* they have received the treatment. That is, the predictor variable reflects the amount of the treatment—possibly zero—that an entity received.) Thus, as we shall see in more detail below, we can use ANOVA to help us determine whether there is evidence of a relationship between the response variable and one or more of the predictor variables in an experiment.

But even if we agree that there are response and predictor variables present in every ANOVA, the question still remains whether it is *useful* to view ANOVA as a technique for detecting relationships between variables. To answer that question note that:

- It is precisely instances of the concept of relationships between variables in entities that most scientists are interested in detecting and describing when they use ANOVA in research projects. That is, most scientists are precisely interested in determining whether the expected value of some variable *y* depends on the values of certain other *x*-variables. They are interested in making this determination because if they find such a relationship, then we (as society) can confidently use the knowledge of the relationship to make predictions or to exercise control.
- As noted above, virtually all the statistical techniques as they are used in empirical research can be viewed as techniques for studying (i.e., detecting and describing) relationships between variables. Viewing ANOVA and the other techniques as techniques for studying relationships between variables helps to unify the techniques, and this unification facilitates understanding. The unification is especially helpful to newcomers to the field of statistics because it helps them to view the field in terms of an easy-to-understand and practical concept. (The statistical techniques for studying standalone distributions are instances of the study of relationships between variables in the sense that such techniques study relationships between a response variable and a set of predictor variables when the set of predictor variables is empty. The techniques for studying relationships with non-empty sets of predictor variables collapse neatly into this degenerate case.)
- Viewing ANOVA as a technique for studying relationships between variables helps to unravel certain important statistical problems, as we shall see later in this paper.

6.9 Why We Need Statistical Tests

Of course, we need not use a statistical test if we have discovered a very strong relationship between variables because in that case the results of the research project will usually leave no doubt as to the existence of the relationship. However, nowadays new strong relationships between variables are not often discovered, perhaps because most of the strong relationships have already been discovered. Thus most relationships that are currently studied are weak enough that statistical tests are necessary.

Recall Tukey's (1989:176) suggestion that there may be a relationship (albeit sometimes very weak) in entities between *all* measurable pairs of variables, regardless of the variables' identities. Given that suggestion (and despite the arguments made in section 6.5 for beginning by assuming that the null hypothesis is true), we can still ask why we need to use statistical tests to determine whether there is a evidence of a relationship between a particular pair of variables when, in fact, there almost surely is.

We need statistical tests because, as Tukey notes, in addition to wishing to know with confidence whether there is a relationship between the variables, we usually need information about the direction or *profile* of the relationship. (The profile tells us whether, when the value of a given predictor variable increases in entities, the value of the response variable can be expected to *increase* or to *decrease* or perhaps to sometimes increase and sometimes decrease depending on the values of other variables.) Only when we perform statistical tests can we be confident that the modest relationships between variables that are typically suggested by the results of modern research accurately reflect profiles that we can expect to find in similar entities in similar situations.

6.10 Interactions and Simple Relationships Between Variables

It is useful to classify relationships between variables into interactions and simple relationships:

Definition: If:

- there is a relationship in entities between a response variable y and p predictor variables x_1, x_2, ..., x_p where $p \geq 2$ and
- the conditioning of y on the x's cannot be represented as a mathematical sum of separate simpler conditionings of y on one or more proper subsets of the same x's and
- there is no higher-level interaction between the x_1, x_2, ..., x_p and other predictor variables with respect to their joint relationship to y

we say the relationship is a *p-way interaction* between the x_1, x_2, ..., x_p with respect to their joint relationship to y.

(The minor difficulty with the word *interaction* appearing in the body of its own definition can be resolved with a more elaborate recursive definition, which I omit here for simplicity.)

Note that I have defined the concept of an interaction in terms of the more fundamental concept of a relationship between variables. Interactions were invented by Fisher (1935, ch. VI) mainly as a means for detecting any form of relationship that might exist between the response variable and the predictor variables in a research project.

Definition: If:

- there is a relationship in entities between a response variable y and a single predictor variable x_1 and
- there is no evidence that there is an interaction between x_1 and any other predictor variable(s) with respect to its relationship to y

we say the relationship is a *simple* relationship (sometimes called a *main effect* relationship).

The last bulleted paragraph in each of the preceding two definitions is usually not included in definitions of *interaction* and *simple relationship*. I discuss the usefulness of these paragraphs in section 2 of appendix B.

6.11 Summary And Preview

In sections 3 - 6 I discussed the concept of a relationship between variables and I noted that one use of the ANOVA statistical tests is to test for evidence of relationships between the response variable and the predictor variable(s) in the population of entities that is studied in an experiment. In section 10 (after some preparatory work in sections 7 - 9) I discusses a contrasting second use of the ANOVA statistical tests.

7. GROUP TREATMENT TABLES

We can summarize the layout of most experiments with a group treatment table, as used extensively by Winer (1971). For example, table 5 summarizes a 2 × 3 experiment in which two predictor variables, *A* and *B*, are manipulated while we observe a response variable, *y*.

TABLE 5
Group Treatment Table for a 2 × 3 Experiment

These Variables are Manipulated Between Groups of Entities		Group
A	*B*	
a0	*b0*	*g1*
a0	*b1*	*g2*
a0	*b2*	*g3*
a1	*b0*	*g4*
a1	*b1*	*g5*
a1	*b2*	*g6*

Each of the *g*'s in the right side of the table represents a different "treatment group" of the entities that participate in the experiment. For example, in a medical experiment the treatment groups might be groups of patients.

The values shown in the two columns in the left side of the table show which combination of the values of the two predictor variables (i.e., which "treatment combination") we apply to the entities in each of the six groups. For example, the fourth row of the table indicates that we apply level 1 of *A* and level 0 of *B* to the entities in group *g4*. In a medical experiment treatment *A* might be insulin and treatment *B* might be another drug that we suspect will enhance the effectiveness of insulin.

Of course, after the entities have received their treatment combinations for an appropriate length of time, we measure the value of the response variable in each entity. In a medical experiment the response variable might be some measure of the rate of carbohydrate metabolism in patients.

After we have carried out the experiment, if the underlying assumptions of ANOVA (see below) are adequately satisfied, we can then use ANOVA to help us analyze the results of the experiment to determine if there evidence of a relationship between the response variable (*y*) and one or both of the predictor variables (*A* and *B*).

The experiment summarized in table 5 is called a 2×3 experiment because the first predictor variable (*A*) has two different values in the experiment and the second predictor variable (*B*) has three different values, and all of the $2 \times 3 = 6$ combinations of values of the predictor variables are present in the experiment.

In experiments that use repeated measurements, we can use the column dimension in the right side of a group treatment table to show both the repeated measurement of the response variable in the entities and the different values of the within-entities predictor variables in the groups of entities in the experiment. For example, table 6 summarizes a $2 \times 3 \times 2$ repeated measurements experiment in which we manipulate the first two variables (*A* and *B*) within the experimental entities, and we manipulate (or possibly simply observe) *C* between the groups of entities.

TABLE 6

Group Treatment Table for a $2 \times 3 \times 2$ Repeated Measurements Experiment

This Variable Varies Between Groups of Entities	These Variables are Manipulated Within Entities						
	a0			*a1*			← *A*
C	*b0*	*b1*	*b2*	*b0*	*b1*	*b2*	← *B*
c1	*g1*	*g1*	*g1*	*g1*	*g1*	*g1*	
c2	*g2*	*g2*	*g2*	*g2*	*g2*	*g2*	

As before, the *g1*'s and *g2*'s represent the two treatment groups of entities in the experiment, and the *a0, a1,* and *b0, b1, b2,* and *c1, c2* denote the different values of predictor variables *A*, *B*, and *C* respectively. The table implies that we measure the value of the response variable six successive times in each entity in each group (one time for each column in the right side of the table, and each time after setting the within-entities predictor variables to the appropriate values), hence the name "repeated measurements".

We can also use group treatment tables to describe experiments that use blocking designs, or fractional factorial designs, or incomplete block designs, or nested designs, or other more unusual experimental designs. A good test of whether you understand the design of an experiment is whether you can draw a group treatment table for it.

8. FULLY CROSSED EXPERIMENTS, UNBALANCED EXPERIMENTS, AND EMPTY CELLS

Definition: An experiment is *fully crossed* if all the possible combinations of the values chosen for the predictor variables are present in the experiment. (Fully crossed experiments are also sometimes called *factorial* experiments.)

The experiments summarized in tables 5 and 6 are both fully crossed. Scientists often use fully crossed experiments because such experiments allow study of all the possible interactions between the predictor variables with respect to their joint relationship to the response variable, and because the results of such experiments are relatively easy to analyze.

Customarily, scientists design experiments so that the same number of experimental entities is assigned to each cell in the group treatment table. For example, if the experiment summarized in table 5 is a medical experiment, we might design it so that each cell in the table has a group of twenty patients assigned to it, implying that there is a total of $6 \times 20 = 120$ patients in the experiment.

If a fully crossed experiment has the same number of experimental entities assigned to each cell in the group treatment table, the experiment is a *balanced* experiment; otherwise the experiment is *unbalanced.* Searle (1988) gives a more complete definition of balance, which covers experiments that are not fully crossed.

Scientists usually design experiments to be balanced experiments because an unbalanced experiment usually provides no advantage over the associated balanced experiment, and because unbalanced experiments are harder to analyze. However, after an experiment is performed, some of the data for one or more of the experimental entities will often be, for some reason, unavailable, and therefore the experiment will have become unbalanced. For example, a patient may withdraw from a medical experiment partway through, and thus at least one value of the response variable for the patient will be unavailable, and therefore the experiment will have become unbalanced.

This paper addresses the common situation in experimental research in which:

- an experiment whose results are being analyzed is unbalanced
- there is no evidence that the imbalance is related to the values of the response or predictor variables
- each cell in the group treatment table has at least one value of the response variable associated with it—that is, there are no "empty" cells in the table.

(Experiments with empty cells are uncommon because most scientists are aware that research projects with empty cells are more difficult to analyze, so they usually design their experiments in ways that minimize the chance that empty cells will occur.)

9. MODEL EQUATIONS

9.1 The Cell-Means Model Equation

The cells in the right side of a group treatment table (i.e., the locations of the *g*'s) are the "cells" of the cell-means model equation that scientists sometimes use to model the behavior of the response variable in experiments. For the experiment summarized in table 5, the cell-means model equation is:

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \qquad (2)$$

where:

- y_{ijk} = the value of the response variable for the *k*th entity in the treatment group (cell) that received level *i* of *A* and level *j* of *B* (*i* = 1, 2; *j* = 1, 2, 3)
- μ_{ij} = the hypothetical expected value of the response variable for the measurements of the response variable in all the entities in the population if they were given the treatment combination associated with the *ij* cell in the group treatment table under conditions identical to those of the experiment
- ε_{ijk} = an "error" term reflecting the difference between the μ_{ij} and the y_{ijk}.

Irwin (1931), Elston and Bush (1964), Speed (1969), and Urquhart, Weeks, and Henderson (1973) were early contributors to the development of the cell-means model equation.

9.2 The Error Term and the Underlying Assumptions

The error term in an ANOVA model equation is important because the ANOVA *p*-values that are computed from the results of an experiment may be incorrect unless the following assumptions about the error term are adequately (but not necessarily fully) satisfied:

- For all the (combinations of) values of the predictor variable(s) used in the experiment, there must be no relationship between the error term and the response variable, or between the error term and any of the predictor variables either in the population of entities under study or in the sample of entities used in the experiment.
- For each cell in the group treatment table, the values of the error term must be distributed with a normal distribution in both the population and the sample.
- The distribution of the values of the error term must have the same expected variance in all the cells in the group treatment table in both the population and the sample.

In this paper I refer to these assumptions as the "underlying assumptions" of ANOVA. It is, of course, an important step in analyzing the results of any experiment to check how well these assumptions are satisfied. Fortunately, they are adequately satisfied in many experiments, especially if the entities in the sample are randomly selected from the population (not always feasible), and if the entities in the sample are randomly assigned to the various treatment groups.

9.3 The Overparameterized Model Equation

We can also model the behavior of the response variable in the experiment summarized in table 5 as:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \phi_{ij} + \varepsilon_{ijk} \qquad (3)$$

where:

- y_{ijk} = the value of the response variable for the *k*th entity in the treatment group (cell) that received level *i* of *A* and level *j* of *B* (the same as in the cell-means model equation)
- μ = the hypothetical grand mean of the values of the response variable that would be obtained if a balanced version of the experiment was performed on the entire population of entities that are under study (other definitions consistent with this paper are possible)
- α_i = the hypothetical simple effect on the mean of the values of the response variable (i.e., on the mean of the values of the y_{ijk}) of giving an entity level *i* of variable *A*
- β_j = the hypothetical simple effect on the mean of the values of the response variable of giving an entity level *j* of variable *B*
- ϕ_{ij} = the hypothetical interaction effect on the mean of the values of the response variable (independent of either of the above two simple effects) of concurrently giving an entity level *i* of variable *A* and level *j* of variable *B* [sometimes also written $(\alpha\beta)_{ij}$]
- ε_{ijk} = an error term reflecting the difference between the sum of the preceding four terms and the value of y_{ijk}; this is the same random variable with the same assumptions as the ε_{ijk} in (2).

The model equation in (3) is called an *overparameterized* model equation. The α's, β's, and ϕ's (but not the ε's) in the overparameterized model equation are called the *parameters* of the equation.

Fisher and Mackenzie (1923) hinted at the idea of the overparameterized model equation, and the idea was developed by Fisher's colleagues, especially Allan and Wishart (1930) and Yates (1933, 1934).

For the experiment summarized in table 5, the link between the cell-means model equation and the overparameterized model equation is

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \phi_{ij}.$$

9.4 Using Model Equations to Make Predictions

An obvious use of model equations is to make predictions of the value of the response variable in new entities that are similar to the entities that participated in the research project. To make these predictions we first obtain numerical estimates of the values of the parameters in the appropriate model equation.

For example, suppose we have performed the experiment summarized in table 5. We can analyze the results of the experiment to obtain numerical estimates for the μ, the 2 α's, the 3 β's and the 2 $\times$ 3 = 6 ϕ's shown in (3). Then, given an entity for which we wish to make a prediction, we determine the values of the predictor variables for that entity, and then we substitute the estimated numerical values of the parameters that correspond to the values of the predictor variables into the right-hand side of the model equation, and then (ignoring the error term because we don't know its value) we add the substituted estimated numerical values of the parameters together to get the predicted value of the response variable for the entity.

We usually determine the estimates of the values of the parameters in a model equation by requiring that the predictions we obtain when we use the estimates to help us make predictions be *as accurate as possible*. This leads to the least-squares method under which we substitute the values of the response variable and the predictor variables from the results of a research project into certain differential equations. Then we (or a computer) solve the equations to obtain estimated values of the parameters such that the sum of the squared errors in the predictions is minimized if we use the model equation together with the estimated values of the parameters to make predictions for all the values of the response variable obtained in the research project.

9.5 The Estimates of the Values of the Parameters in an Overparameterized Model Equation Are Not Unique

Overparameterized model equations are so named because there are more parameters in such an equation than there are non-empty cells in the group treatment table that describes the associated research project. For example, in the experiment summarized in table 5 and (3), there are 2 $\times$ 3 = 6 cells in the group treatment table. But if we count all the parameters in (3), we can see that there are 1 μ + 2 α's + 3 β's + (2 $\times$ 3) ϕ's = 12 parameters in the equation. Since there are more parameters than cells, it follows from linear algebra and calculus that it is impossible, without further information, to write a set of least-squares differential equations (employing data from a research project that is consistent with the model equation) that we can then solve to obtain unique estimates of the values of the parameters.

(If an unsaturated model equation is used—see below—then there are more parameters in the model equation than there are cells in the associated *collapsed* group treatment table.)

Although we cannot write least-squares equations and solve them for *unique* estimates of the values of the parameters in an overparameterized model equation, linear algebra and calculus still allow us to write least-squares equations and solve them for *non-unique* estimates of the values of the parameters. Of course, these non-unique estimates are not completely non-unique—that is, jointly free to assume *any* values—or the estimates would be meaningless. Instead, these non-unique estimates are always constrained to be estimates of the values of the parameters that minimize the sum of the squared errors in prediction of the value of the response variable across all the values that were obtained in the research project.

9.6 Sigma Restrictions

It is generally easier to obtain *the* solution to a set of equations that has a unique solution than to obtain *a* solution to a set of equations that has a non-unique solution. Therefore, without loss of relevant generality, we can facilitate solving for estimates of the values of the parameters in an overparameterized model equation by forcing a unique solution on the estimates. Scientists usually do this by pre-defining certain restrictions (constraints) on the estimates. These restrictions are specified in terms of equations that state reasonable relationships among the estimates. When the restriction equations are taken together with the original least-squares differential equations, the full set of equations has a unique solution that we can easily obtain through matrix algebra with a computer or hand calculator. This gives us unique estimates of the values of the parameters.

In view of the way they are written, the additional restrictions on the estimates of the values of the parameters are called *sigma restrictions* (or sometimes called *side conditions*). The following four equations show the sigma restrictions that scientists sometimes use for the overparameterized model equation shown in (3):

$$\begin{aligned} &\textstyle\sum_i \alpha_i = 0 \\ &\textstyle\sum_j \beta_j = 0 \\ &\textstyle\sum_i \phi_{ij} = 0 \qquad \forall\, j \\ &\textstyle\sum_j \phi_{ij} = 0 \qquad \forall\, i. \end{aligned} \tag{4}$$

Although the sigma restrictions give us "unique" estimates of the values of the parameters in the model equation, these estimates are unique only relative to the particular set of sigma restrictions that we have chosen. And, in general, if we choose another set of sigma restrictions (there are infinitely many choices), we will obtain another "unique" set of estimates of the values of the parameters.

9.7 The Use of Overparameterized Model Equations

Because the estimates of the values of the parameters in an overparameterized model equation are not unique, scientists who are analyzing the data of an experiment usually do not, as a practical matter, bother to obtain estimates of the values of the parameters in the associated overparameterized model equation.

(Scientists may, however, be interested in having the computer supply the predicted value of the response variable for each cell in the group treatment table, and these predicted values, which are not always simple cell means, can be readily computed from the estimated values of the parameters. Perhaps surprisingly, but of course ultimately necessary for reasonableness, for a given model equation, the predicted values of the response variable are *independent* of the particular set of [linear] constraints [i.e., sigma restrictions] that we have chosen to facilitate solving for the estimates of the values of the parameters.)

Because overparameterized equations have "too many" parameters, and because the estimated values of the parameters are not unique, some scientists have completely abandoned overparameterized model equations. However, overparameterized model equations are useful in discussing experiments and ANOVA because:

- overparameterized model equations provide an easy-to-grasp overview of the relationship between the response variable and the predictor variables in an experiment, including illustrating the roles that simple effects and interactions play in the relationship
- the parameters in overparameterized model equations can be used to describe what is being tested in statistical tests, and are especially helpful in providing understandable descriptions of tests of interactions
- the parameters in overparameterized model equations can help to characterize different possible forms of the relationship between the response and predictor variables, as an aid to power calculations
- overparameterized model equations can help to characterize the computation of sums of squares in ANOVA, as discussed in sections 12 - 17.

As suggested by the last item in the preceding list, some of the following discussion is in terms of overparameterized model equations with sigma restrictions. I use these equations because they facilitate understanding. However, the conclusions I draw are independent of the form of the model equation. That is, in order to draw the conclusions, we need not use sigma restrictions to force a unique solution for the estimates of the values of the parameters. (However, sigma restrictions are sometimes necessary for another purpose, which I discuss in a technical note at the end of section 13.2.) And we can draw the same conclusions in this paper using an overparameterized model equation without "forcing" sigma restrictions, or using a cell-means model equation.

10. RELATIONSHIPS AMONG PARAMETERS

10.1 Review of Relationships Between Variables

In section 6 I noted that one use of the statistical tests in ANOVA is to help us analyze the results of an experiment to see whether there is significant evidence of a relationship between the response variable and the predictor variables in the entities in the population under study. For example, for the experiment summarized in table 5 and in (2) and (3), we can use ANOVA to help us determine whether there is evidence of a relationship between the response variable *y* and predictor variable *A*. That is, we can use ANOVA to test whether the expected value of *y* in entities *depends* on the value of *A*. In yet other words, using the formal definition of a relationship given in section 6.4, we can test whether

$$H_0: \quad E(y) = E(y|A). \tag{5}$$

And if our test provides sufficient evidence that H_0 is *not* satisfied, we can then conclude that there is a relationship between *y* and *A*. Let us call this use of a statistical test in ANOVA *testing for relationships between variables.*

10.2 Relationships Among Parameters

A second use of the statistical tests in ANOVA is to test whether subsets of the (population) *parameters in the model equation* bear particular numerical relationships to one another. For example, for the experiment summarized in table 5 and in (2) we may wish to use ANOVA to test the hypothesis that the following relationship exists among the parameters of (2):

$$H_0': \quad \sum_j \mu_{ij} / b \qquad \text{equal } \forall\ i \tag{6}$$

where:

b = the number of different values of predictor variable *B* appearing in the experiment.

To understand H_0' it is helpful to arrange the population means for all the cells in the group treatment table in a two-dimensional array, with the values of predictor variable *A* indexing the rows and the values of predictor variable *B* indexing the columns, as in array 1.

(The *a* and *b* in the array are the number of different values appearing in the experiment of predictor variables *A* and *B* respectively.)

The rightmost column in the array contains the mean of the cell means for each row. That is

$$\bar{\mu}_{i\bullet} = \sum_j \mu_{ij} / b.$$

Thus in terms of array 1, the hypothesis H_0' states that if we compute the mean of the population cell means for each row in the array, then these row means will all be equal to one another.

ARRAY 1

Layout of the Population Cells Means
For the Experiment Summarized in Table 5
To Facilitate Understanding H_0'

		Level of B			Row Mean of Cell Means
		1	...	b	
Level of A	1	μ_{11}	...	μ_{1b}	$\bar{\mu}_{1\bullet}$
	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
	a	μ_{a1}	...	μ_{ab}	$\bar{\mu}_{a\bullet}$

Because H_0' is stated in terms of the parameters of the model equation, let us call the use of ANOVA exemplified in the test of H_0' *testing for relationships among parameters.*

10.3 Are Relationships Among Parameters the Same as Relationships Between Variables?

An obvious question is whether H_0 and H_0' are the same hypothesis. It turns out that although they are related, they are not the same because there are states of affairs in which H_0 is rejected (i.e., there is a relationship between y and A—specifically a particular type of interaction relationship) but H_0' is satisfied. I describe such a state of affairs in the candy sales example in section 6 of appendix B. (On the other hand, whenever there is no relationship between y and A—i.e., H_0 is true—H_0' is also always true.)

10.4 Why Scientists Test for Relationships Among Parameters

There are two reasons why a scientist may wish to perform a test for a relationship among the parameters of a model equation:

1. the scientist may be interested in testing for the existence of a relationship among the parameters *in its own right*
2. the scientist may be using the test of the relationship among parameters as a test of whether there is a relationship between the response variable and one or more of the predictor variables in the research project. That is, the scientist may use a test of a relationship among parameters as a test of a relationship between variables.

Testing for relationships among parameters in their own right is a valid research interest, which I briefly discuss further in sections 14 and 15. However, although the procedure of testing for relationships among parameters has been discussed often by statisticians, interest in this procedure among actual empirical scientists—e.g., medical scientists—is less common because the concept of a relationship among the parameters of a model equation is somewhat abstract, and is therefore (except in the simplest cases) difficult for some scientists to understand.

On the other hand, for the reasons discussed in section 6.8, most scientists usually *are* interested in testing for relationships between variables. In that situation, if a scientist uses a test of a hypothesis of a relationship among parameters (such as H_0') to test a hypothesis of a relationship between variables (such as H_0), this approach, although valid, is less efficient because it leads to two difficulties:

The first difficulty is that if we use a test of a relationship among parameters to test for evidence of a relationship between variables, then in order to understand what is being tested, we must understand the concept of a relationship among the parameters of the model equation. However, as noted above, the concept of a relationship among the parameters of the model equation (whether cell-means or overparameterized) is somewhat difficult to understand, as any statistician who has tried to explain this concept to a client will perhaps agree.

Furthermore, if we use a test of a relationship among parameters to test for evidence of a relationship between variables, we must understand how the concept of a relationship among parameters relates to the concept of a relationship between variables, another somewhat difficult task.

On the other hand, if we simply concentrate on tests of relationships between variables, we (and our clients) need only understand the concept of a relationship between variables, which is relatively easy to understand, especially if we view the concept in terms of the value of the response variable *depending* on the values of the relevant predictor variable(s).

The second difficulty with using a test of a hypothesis of a relationship among parameters (e.g., H_0') to test a hypothesis of a relationship between variables (e.g., H_0) is that this approach may confuse us into thinking that only a subset of the valid statistical tests are valid. And selecting statistical tests from this subset will lead us to use sub-optimal statistical tests, as I discuss in sections 14 through 18.

(In sections 6 and 10 I discussed two uses of the ANOVA statistical tests. Although ANOVA is often employed in these uses, there are other uses as well. Hoaglin, Mosteller, and Tukey [1991 ch. 2] discuss another classification of the uses of ANOVA.)

11. SUMMARY AND PREVIEW

In sections 1 through 10 of this paper I discussed two uses of the ANOVA statistical tests, namely:

- to test for relationships between variables (section 6) and
- to test for relationships among the parameters of the model equation (section 10).

I noted that scientists are often interested in testing for relationships between variables. In sections 12 through 18 I will discuss two methods of computing ANOVA sums of squares and will evaluate the two methods of computing sums of squares for fulfilling the two uses of the ANOVA statistical tests.

12. RESIDUAL SUMS OF SQUARES

In sections 9.3 - 9.6 of this paper I noted that for any overparameterized model equation, and for the results of any research project that is consistent with the model equation and that has no empty cells in the group treatment table, we can (using the method of least squares and possibly using the sigma restrictions) "fit" the equation to the results of the research project and thereby obtain estimates of the values of the parameters in the equation.

For example, suppose we have performed the 2 × 3 experiment summarized in table 5. In section 9.3 I noted that this experiment has the following overparameterized model equation:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \phi_{ij} + \varepsilon_{ijk}. \qquad (3)$$

In section 9.4 I noted that we can use the method of least squares to fit (3) to the results of the experiment and thereby obtain estimates of the values of the μ, α's, β's, and ϕ's in the equation for the population of entities we are studying.

I also noted that once we have obtained estimates of the values of the parameters, we can then substitute the appropriate estimates into the model equation to predict the value of the response variable for each entity in each cell in the group treatment table in the research project. (Of course, for any given cell in the group treatment table, the predicted value for all the entities in the cell is the same.)

Using the foregoing ideas, let us consider a definition of the concept of a residual sum of squares:

> *Definition:* For any research project in which the response variable has numeric values, the *residual sum of squares* of a model equation in the research project is the sum across all the values of the response variable in the research project of the squared deviation of the value predicted by the model equation for an entity from the actual measured value of the response variable in the entity. That is, for any model equation in a research project, the residual sum of squares is
>
> $$\mathrm{SS}_r = \sum (measured - predicted)^2.$$

It is, of course, the residual sum of squares that the least-squares procedure minimizes in order to determine the estimates of the values of the parameters. I use the concept of the residual sum of squares of a model equation shortly.

13. TWO METHODS OF COMPUTING ANOVA SUMS OF SQUARES

A critical step in computing the *p*-values in an ANOVA is to compute different "sums of squares" from the results of the experiment. (Please distinguish ANOVA sums of squares from the closely related residual sums of squares discussed in the preceding section.) Various methods of computing ANOVA sums of squares are available, two of which I discuss in this section.

We can characterize both methods of computing ANOVA sums of squares in terms of the difference between the residual sums of squares of two model equations (Yates 1934:63, Scheffé 1959, Searle 1971). I shall call these equations the two *generating equations* for an ANOVA sum of squares.

13.1 The HTO Method of Computing Sums of Squares

One method of computing ANOVA sums of squares is to have Higher-level Terms Omitted from the two generating model equations (HTO). For example, if we have performed the experiment summarized in table 5, then under the HTO method we can compute the ANOVA sum of squares for the *A* simple relationship (main effect) by first fitting (separately) the following two new model equations to the data:

$$y_{ijk} = \mu \qquad + \beta_j + \varepsilon_{ijk} \qquad (7)$$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad (8)$$

where in each equation (if the associated term is present):

$$\sum_i \alpha_i = 0$$

$$\sum_j \beta_j = 0.$$

Then under the HTO method, the sum of squares for the *A* simple relationship is the residual sum of squares for (7) minus the residual sum of squares for (8). [This difference will always be non-negative because (8) has an additional term, and thus will always provide at least as close a fit to the data as (7).] Thus (7) and (8) are the generating model equations for the HTO sum of squares for the *A* simple relationship for the experiment.

By examining (7) and (8) we can see that under the HTO method of computing sums of squares we are testing the effect on *y* of a change in the value of variable *A* by studying the *reduction* in the residual sum of squares if we include a term for variable *A* in the model equation, using a model equation *with no interaction term.* It is customary to say that the interaction term is a *higher-level* term than the *A* term because the interaction term refers to two of the predictor variables (i.e., *A* and *B*) while the *A* term refers to only one. I use the name HTO to reflect the fact that (although terms at the same level are included) Higher-level Terms are Omitted from the two generating model equations.

Generating equations (7) and (8) are called *unsaturated* model equations because not all the possible terms in the standard overparameterized model equation for the experiment [i.e., equation (3)] are included in the equations. We can always use the least-squares procedure to obtain estimates of the values of the parameters in an unsaturated overparameterized model equation. If we wish, we can use sigma restrictions to facilitate the solution.

The procedure of computing an HTO sum of squares by computing the difference in the residual sums of squares of two generating model equations illustrates what is being computed when we compute the sum of squares. There are, of course, computationally or algebraically more efficient (but conceptually less transparent) procedures for performing the same computation as discussed briefly in the second part of appendix C. For detailed coverage of the methods, see the landmark discussions by Hocking (1985) and Searle (1987).

13.2 The HTI Method of Computing Sums of Squares

A second method of computing ANOVA sums of squares is to have Higher-level Terms Included in the two generating model equations (HTI). For example, if we have performed the experiment summarized in table 5, then under the HTI method we can compute the ANOVA sum of squares for the *A* simple relationship by first fitting (separately) the following two model equations to the data:

$$y_{ijk} = \mu \qquad + \beta_j + \phi_{ij} + \varepsilon_{ijk} \tag{9}$$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \phi_{ij} + \varepsilon_{ijk} \tag{10}$$

where we use the sigma restrictions from (4).

Then under the HTI method, the sum of squares for the *A* simple relationship is the residual sum of squares for (9) minus the residual sum of squares for (10).

By examining (9) and (10) we can see that under the HTI method of computing sums of squares we are testing the effect on *y* of a change in the value of variable *A* by studying the reduction in the residual sum of squares if we include a term for variable *A* in the model equation, but this time we are using model equations *with* an interaction term. I use the name HTI to reflect the fact that (in addition to including terms at the same level) Higher-level Terms are Included in the two generating model equations.

[On a technical matter, in unsaturated generating model equations, in addition to allowing us to obtain unique estimates of the value of the parameters, some of the sigma restrictions play a second role: they prevent the interaction terms in the model equation from wrongly accounting for variation in the values of the response variable that should be accounted for by lower-level terms or, as in (9), that should not be accounted for at all. This approach, which is entailed by the principle that each interaction term in a model equation should be independent of the effects of all of the other terms, solves a problem identified by Searle (1971, 1987:339-340) and Nelder (1977:50) concerning certain ANOVA sums of squares that would undesirably turn out to be zero.]

13.3 General Comments

I have illustrated the HTO and HTI methods of computing ANOVA sums of squares in terms of an experiment with two predictor variables. The distinction between the two methods can be generalized to include experiments with *any* number of predictor variables by noting that we can view each method in terms of fitting two generating model equations to the data: one equation containing the term for the effect being tested, and the other equation lacking the term for the effect being tested. And the desired sum of squares is the difference between the residual sums of squares of the two generating equations. In computing the HTO sums of squares, Higher-level Terms are Omitted from the two generating equations although (with one obvious exception) all the terms at the same level as, and at lower levels (if any) than, the effect being tested are included. On the other hand, in computing the HTI sums of squares, all the Higher-level Terms are Included in the two generating equations along with all the terms at the same and lower levels (with the same one exception).

The HTO and HTI methods of computing ANOVA sums of squares generally yield numerically different values from each other in an unbalanced experiment. However, they always yield identical values in a balanced experiment. Furthermore, in a balanced experiment the two methods yield sums of squares that are identical to the sums of squares obtained through the standard ANOVA techniques for balanced experiments.

In a fully crossed unbalanced experiment with two or more predictor variables (and with no empty cells), it can be shown that the HTO and HTI sums of squares for the highest-level interaction are always identical. Similarly, in an unbalanced experiment with only a single predictor variable, it can be shown that the HTO and HTI sums of squares for the simple relationship are always identical.

Table 7 shows the names that twelve articles and eight statistical computer programs use for the HTO and HTI methods of computing ANOVA sums of squares in experiments with two predictor variables.

TABLE 7
Names Used by Twelve Articles and Eight Computer Programs
for Two Methods of Computing Sums of Squares
in Two-Way Unbalanced ANOVA

Reference	HTO	HTI
ARTICLES		
Yates (1934)	fitting constants	weighted squares of means
Overall and Spiegel (1969)	method 2	method 1
Burdick et al (1974)	method 2	method 1
Kutner (1974)	hypotheses F and G	hypotheses A and B
Hocking and Speed (1975)	hypotheses H_{A**} and H_{B**}	hypotheses H_A and H_B
Speed and Hocking (1976)	hypotheses 5 and 6	hypotheses 1 and 2
Herr and Gaebelein (1978)	EAD (method 3)	STP (method 1)
Speed, Hocking, and Hackney (1978)	hypotheses 3 and 7	hypotheses 1 and 5
Burdick (1979)	hypothesis 3	hypothesis 1
Elliott and Woodward (1986)	hypotheses 3 and 6	hypotheses 1 and 2
Milligan et al (1987)	hypothesis H3 and H7	hypothesis H1 and H5
Singh and Singh (1989)	hypotheses H_5 and H_6	hypotheses H_1 and H_2
COMPUTER PROGRAMS		
SAS GLM (SAS Institute 1990:109)	Type II	Type III
SPSS ANOVA (SPSS 1990:64)	method = experimental	method = unique
SPSS MANOVA (SPSS 1990:377)	-	unique method
BMDP 4V (Dixon 1990a:1155)	weights = sizes	weights = equal
BMDP 2V (Dixon 1990b:489)	-	default method
SYSTAT MGLH (Wilkinson 1990:140)	-	default method
MINITAB GLM (Minitab 1989:8-27)	-	default method
Data Desk Linear Models (Velleman n.d.:39)	-	Partial (Type 3)

Table 7 covers only two-way experiments. For three-way and higher experiments, the SPSS ANOVA EXPERIMENTAL sums of squares are identical to HTO sums of squares. However, for three-way and higher experiments, although some SAS type II sums of squares and some BMDP4V WEIGHTS = SIZES sums of squares are identical to HTO sums of squares, others are not. (Nor are the SAS type II sums of squares and the corresponding BMDP4V WEIGHTS = SIZES sums of squares always identical to each other.) Davidson and Toporek (1979:21) compare sums of squares from the three statistical packages for two- and three-way experiments. In their comparison the SPSS-D (for "default") sums of squares are identical to what are now called SPSS ANOVA EXPERIMENTAL sums of squares and URWAS is the old name for the computer program that is now called BMDP4V.

If you need to compute an HTO (or certain other types of) sum of squares, but if the available statistical package cannot do the computation directly, then (with most packages) you can do the computation indirectly by using the following steps for each sum of squares you wish to compute:

- Specify to the linear regression program in the package the first of the two generating model equations for the sum of squares and then use the regression program to compute the residual sum of squares for that model equation. (Use full-rank dummy-variable specification—which has the same effect as imposing sigma restrictions—in this and the next step in order to bypass the problem discussed in the last paragraph of section 13.2.)

- Specify to the regression program the second of the two generating equations and then use the program to compute the residual sum of squares for *that* equation.
- Subtract the smaller residual sum of squares from the larger to yield the associated ANOVA sum of squares.

Alternatively, you can obtain the same result using the matrix arithmetic procedure given in the second part of appendix C.

14. THE HTO METHOD IS GENERALLY INAPPROPRIATE WHEN TESTING FOR RELATIONSHIPS AMONG PARAMETERS

I have now discussed:

- using ANOVA to test for relationships between variables and using ANOVA to test for relationships among the parameters of the model equation (sections 6 and 10 respectively)
- the HTO and HTI methods of computing ANOVA sums of squares (section 13).

In this and the next three sections I address the questions of which of the HTO and HTI methods is better for testing for relationships between variables, and which is better for testing for relationships among parameters. I begin by considering the HTO method and testing for relationships among parameters.

Several authors have correctly noted that the HTO method of computing sums of squares is generally not appropriate if we wish to test for relationships among parameters (Carlson and Timm 1974, Kutner 1974, Speed, Hocking, and Hackney 1978, Hocking, Speed, and Coleman 1980, Searle 1987). These authors have shown that, in general, under the testing for relationships among parameters approach, the HTO method does not test meaningful, useful, or interesting hypotheses.

For example, consider the two-way ANOVA for the experiment summarized in table 5. Searle (1987:115) notes that the HTO sum of squares for the *A* simple relationship (main effect) in this experiment enables us to test whether the following relationship exists among the parameters of (2), the associated cell-means model equation:

$$\sum_j n_{ij}\mu_{ij} = \sum_j \sum_t \frac{n_{ij}n_{tj}}{n_{\bullet j}}\mu_{tj} \qquad \forall\ i \qquad (11)$$

where:

n_{ij} = number of entities in the *ij* cell in the group treatment table and

$n_{\bullet j} = \sum_i n_{ij}$.

Searle also notes (1987:333) that the same HTO sum of squares for the *A* simple relationship enables us to test whether the following relationship exists among the parameters of (3), the associated overparameterized model equation:

$$\sum_j n_{ij}(\alpha_i + \phi_{ij}) = \sum_j \sum_t \frac{n_{ij}n_{tj}}{n_{\bullet j}}(\alpha_t + \phi_{tj}) \qquad \forall\ i. \quad (12)$$

The important question is whether either of these tests of relationships among the parameters in an unbalanced experiment is of any use. And, as the authors cited above have noted, scientists are virtually never *specifically* interested in testing whether the (equivalent but somewhat arcane) relationships shown in (11) and (12) exist among the parameters of the model equation. This example illustrates why the HTO method of computing sums of squares is generally inappropriate if our goal is to test for relationships among parameters.

15. BUT THE HTI METHOD *IS* APPROPRIATE WHEN TESTING FOR RELATIONSHIPS AMONG PARAMETERS

On the other hand, Searle (1987:335) notes that the HTI test for the *A* simple relationship (main effect) tests whether the following relationship (discussed earlier in section 10.2) exists among parameters of the cell-means model equation:

$$H_0'\text{:} \quad \sum_j \mu_{ij} / b \quad \text{equal } \forall\ i. \qquad (6)$$

Similarly, Searle (1987:335) notes that the same HTI test for the *A* simple relationship tests whether the following relationship exists among the parameters of the overparameterized model equation:

$$\alpha_i + \sum_j \phi_{ij} / b \quad \text{equal } \forall\ i. \qquad (13)$$

If the sigma restrictions are imposed, the relationship in (13) reduces to:

$$\alpha_i = 0 \qquad \forall\ i. \qquad (14)$$

The relationships shown in (6), (13), and (14) are relatively easy to understand and are relationships among parameters that scientists are sometimes interested in testing.

Because statisticians have endorsed the HTI method of computing ANOVA sums of squares (because the associated statistical tests test hypotheses of interest for relationships among parameters), and because many scientists are not familiar with the facts discussed below, many scientists choose the HTI method of computing sums of squares in unbalanced ANOVA *even when they are not interested in testing for relationships among parameters, but are instead actually interested in testing for relationships between variables.* Unfortunately, as I shall discuss below, this choice is sub-optimal.

16. BOTH THE HTO AND HTI METHODS ARE APPROPRIATE WHEN TESTING FOR RELATIONSHIPS BETWEEN THE RESPONSE VARIABLE AND THE PREDICTOR VARIABLES

In the preceding two sections I reviewed the well-known conclusions that the HTO sums of squares are generally not appropriate if our goal is to test for relationships among parameters while the HTI sums of squares are generally appropriate for that goal. That review was included to help readers contrast its conclusions with the main argument of this paper. Let me now return to the main argument and evaluate the HTO and HTI methods of computing sums of squares if our goal is to test for relationships between the response variable and the predictor variables in the population of entities under study in an experiment.

I begin with a definition of a valid statistical test for a relationship between variables:

Definition: In situations in which the associated underlying assumptions are satisfied, an *F*-ratio-based statistical test for a relationship between variables is a *valid* test if:

- in the absence of the indicated relationship the *F*-ratio yielded by the test can be shown to have a central *F*-distribution with known degrees of freedom and
- in the presence of the indicated relationship the expected value of the numerator of the *F*-ratio can be shown to be greater than the expected value of the denominator.

The definition implies that if an ANOVA statistical test for a relationship between variables is valid (and if the underlying assumptions of ANOVA are adequately satisfied), we can then be confident that the *p*-values that are yielded by the test are correct estimates of the probabilities that they purport to estimate. Obviously, it is important that ANOVA statistical tests for relationships between variables be valid.

Let us now consider a theorem that shows that the HTO sums of squares provide valid statistical tests for relationships between variables in certain general situations:

Theorem 1 (proved in appendix A): Consider an n-way fully crossed research project (where $n > 1$) with no empty cells in the group treatment table. Suppose that the underlying assumptions of ANOVA are satisfied. Select any q of the predictor variables where $1 \leq q < n$. Suppose that there are no $(q + 1)$-way or higher-way population interactions involving all of the q selected predictor variables (it being only necessary to consider variables measured in the research project). Then the HTO statistical test that is associated with the q predictor variables is a valid test if we wish to test for evidence of a q-way population relationship between the q selected predictor variables with respect to their joint relationship to the response variable. (That is, if $q > 1$, the test is a valid test for the presence of the q-way interaction, and if $q = 1$, the test is a valid test for the presence of the simple relationship.) Similarly, the n-way HTO test is a valid test if we wish to test for evidence of an n-way population interaction between all of the predictor variables with respect to their joint relationship to the response variable.

Put simply, theorem 1 implies that the HTO sums of squares are (with one qualification, and if the underlying assumptions are adequately satisfied) valid for use in statistical tests if our purpose is to test the results of a research project for evidence of a population relationship between the response variable and the predictor variables.

Theorem 1 can also be proven for statistical tests that are based on the HTI sums of squares.

Thus both the HTO and HTI sums of squares are appropriate for testing for population relationships between the response variable and the predictor variables in an experiment, at least in the absence of associated higher-level interactions.

I discuss why higher-level interactions usually do not create problems when scientists are testing for relationships between variables in appendix B.

17. THE HTO METHOD IS GENERALLY MORE POWERFUL THAN THE HTI METHOD WHEN TESTING FOR RELATIONSHIPS BETWEEN VARIABLES

Definition: If

- we apply an ANOVA statistical test of a relationship between variables to the results of a research project and
- the underlying assumptions of ANOVA are adequately satisfied and
- (most importantly) the *p*-value that is yielded by the test is less than or equal to the critical *p*-value that we have chosen for the test (i.e., usually .05 or .01),

then we say that the statistical test *provides reasonable evidence* that the relationship exists.

Using the concept of *providing reasonable evidence that a relationship exists*, let us now define the power of a statistical test.

Definition: For a given form of a relationship between variables, the *power* of a statistical test of the relationship is the fraction of the times that the test is performed (each time with a fresh random sample from the population of interest) that the test will provide reasonable evidence that the relationship exists, given that the relationship has the specified form.

Within the limits of time and cost constraints, scientists generally prefer to use statistical tests that (when operating in the vicinity of the expected form of the relationship) have the highest possible power because such tests are more likely to discover the sought-after relationship, if it exists. Thus having noted that both the HTO and HTI sums of squares are appropriate for testing for relationships between the response variable and the predictor variables, an important question is: Of the HTO and HTI methods of computing sums of squares, which method yields more powerful statistical tests for detecting relationships between variables?

> *Theorem 2* (proved in appendix C): Consider an n-way fully crossed research project (where $n > 1$) with no empty cells in the group treatment table. Suppose that the underlying assumptions of ANOVA are satisfied. Select any q of the predictor variables where $1 \leq q < n$. Suppose that there are no $(q + 1)$-way or higher-way population interactions among the predictor variables in the research project with respect to their joint relationship to the response variable. Then if a q-way population interaction (or, if $q = 1$, a simple relationship) exists between the selected predictor variables with respect to their joint (its) relationship to the response variable, the statistical test (using within-cell error) for the q-way interaction (or, if $q = 1$, for the simple relationship) based on the HTO sum of squares is at least as powerful for detecting the interaction (simple relationship) as the statistical test (also using within-cell error) based on the HTI sum of squares.

Theorem 2 implies that an HTO statistical test is, under certain general conditions, at least as powerful for detecting relationships between variables as the associated HTI statistical test. However, because the two tests have exactly the same power only when special conditions obtain, usually in unbalanced experiments the HTO test is *more* powerful than the HTI test. Of course, the size of the difference in power between the two tests depends on the pattern of imbalance in the experiment. (Littell and Lynch [1983] examine the simplest case.) Furthermore, often the difference in power will be small. However, given the possible high social cost of a failure to discover a useful existing relationship between variables, the difference in power is large enough to be a consideration when deciding which method of computing sums of squares to use when testing for relationships between variables.

You can easily empirically see the validity of theorem 2 by analyzing several sets of unbalanced experimental data with an unbalanced ANOVA computer program that can compute both types of sums of squares. (Of course, for each experiment there must be no empty cells in the group treatment table, the data must contain evidence of a relationship, and the experiment must have two or more predictor variables.) Except for the highest-level interaction (for which the HTO and HTI sums of squares are identical), whenever an effect is significant, the HTO sum of squares for the statistical test will usually (but not always) be larger than the corresponding HTI sum of squares, and thus the p-value for the HTO sum of squares will usually be smaller than the p-value for the HTI sum of squares, which indirectly implies that statistical tests based on the HTO method of computing sums of squares are generally more powerful.

Theorem 2 implies that statistical tests based on the HTO method of computing sums of squares are preferred to statistical tests based on the HTI method in ANOVA of unbalanced experiments with no empty cells when there are no higher-level interactions if (as is often the case) our goal is to test for relationships between the response variable and the predictor variables. That is the main point of this paper.

I discuss an extension to the HTO approach for three-way and higher experiments in appendix D and I discuss other approaches to ANOVA statistical tests in appendix E.

18. SUMMARY

Entities, properties of entities, and relationships between properties of entities are fundamental concepts of human thought. In science, properties are roughly synonymous with variables. Relationships between variables are of central concern to scientists.

Two uses of the ANOVA statistical tests are:

- to test for population relationships between the response variable and the predictor variables in an experiment
- to test for population relationships among the parameters of the model equation in an experiment.

Two popular approaches for computing ANOVA sums of squares are the HTO approach or the HTI approach.

If we wish to test for relationships between the response variable and the predictor variables in an experiment, both the HTO and HTI approaches are appropriate (valid).

Main point: If we wish to use ANOVA to test for evidence of relationships between variables in an experiment with no empty cells, statistical tests based on the HTO approach are preferred to statistical tests based on the HTI approach because the former tests are generally slightly more powerful.

APPENDIX A: PROOF OF THEOREM 1

The following proof is a generalization of the first part of a theorem proved in the two-way case by Burdick and Herr (1980:239):

As exemplified by Searle (1987), for any given HTO ANOVA statistical test, the hypothesis being tested can be shown in terms of one or more expressions that contain only terms in parameters of the overparameterized model equation. Furthermore, for a given HTO test, these expressions contain only terms in the parameters associated with (*a*) the effect being tested and (*b*) the associated higher-level interactions. For example, see (12).

(*) For any *q*-way effect in an overparameterized model equation (i.e., an effect for a particular interaction or an effect for a particular simple relationship), if:

- there are no associated higher-level interactions in the population between the predictor variable(s) associated with the effect and other predictor variables in the research project with respect to their joint relationship to the response variable and
- there is no associated *q*-way relationship in the population between the predictor variable(s) associated with the effect with respect to their (its) joint relationship to the response variable

then the parameters in the overparameterized model equation that are associated with the effect will be all identical. (This can be derived from the definition of interactive and simple relationships between variables and from the definition of an overparameterized model equation.)

If the parameters associated with a higher-level interaction in an expression of a hypothesis being tested are all identical, we can remove any terms that are associated with that interaction from the expression of the hypothesis in order to obtain a more succinct statement of the hypothesis. (This can be seen by examining the expressions of the hypothesis being tested such as (12) and noting that if the ϕ_{ij}'s are all replaced by a constant *k*, then the corresponding statement of the hypothesis with the *k*'s omitted can be derived from the statement with the *k*'s included through some minor algebra.)

If we remove all the higher-level interaction terms from the expressions of a hypothesis, examination of the resulting expressions reveals that the test is a test (using a weighting to reflect the relative accuracy of the estimated values of the parameters) of whether all the parameters associated with the effect in question are identical. (This can be seen in (12) by noting that if the ϕ's are all removed, and if the α's in the *i* expressions are all set identical to some value *k*, then the expressions in (12) all become tautologically true.)

But from (*) if there is no associated higher-level interaction and no *q*-way relationship, the parameters associated with the effect *will* all be identical.

Thus if there is no associated population higher-level interaction and no population *q*-way relationship, the expression of the hypothesis associated with the HTO sum of squares will be satisfied, and thus (if the underlying assumptions of ANOVA are satisfied) the numerator mean square and the denominator mean square in the *F*-ratio will have the same expected value. Thus if there is no associated population higher-level interaction and no population *q*-way relationship, (and if the underlying assumptions of ANOVA are satisfied), the HTO *F*-ratio will have a central *F*-distribution. And thus the HTO statistical test satisfies the first part of the definition of a valid ANOVA statistical test.

On the other hand, if there *is* a visible (within the context of the experiment) population relationship between the predictor variable(s) associated with the effect of interest with respect to their joint (its) relationship to the response variable (but there is no associated higher-level interaction), then the parameters associated with the effect will *not* be all identical and thus, because the term associated with those non-identical parameters is omitted, the model equation with the omitted term [e.g., (7)] will have an inflated residual sum of squares. Thus the effect mean square based on the difference of the residual sums of squares of the two model equations [e.g., (7) and (8)] will have an expected value that is greater than the error mean square. Thus the expected value of the numerator of the HTO *F*-ratio will be greater than the expected value of the denominator. And thus the HTO statistical test satisfies the second part of the definition of a valid ANOVA statistical test.

And thus the theorem is proved.

APPENDIX B: HIGHER-LEVEL INTERACTIONS

In this appendix I discuss why the qualification about higher-level interactions given at the end of section 16 usually does not create problems when scientists are testing for relationships between variables.

B.1 A Reasonable Approach to ANOVA

Scientists who are testing for relationships between variables usually use ANOVA to help them answer the following two questions:

- As x_1 changes in an entity, can we also expect *y* to change in the entity? That is, is there a relationship between x_1 and *y*?
- Given that there is a relationship between x_1 and *y* in entities, does the profile of the relationship depend on the level of x_2, x_3, ... in the entities? That is, is the relationship an interactive relationship?

A reasonable approach to answering these questions is (after performing an appropriate experiment) to examine the components (rows) in the ANOVA table *beginning with the highest-level interaction and working down to the simple relationships* (main effects). As we work down, if we find a significant component, we can tentatively conclude that there is a relationship between (*a*) the response variable and (*b*) the *set* of one or more predictor variables that are associated with the component. (Of course, if p predictor variables are associated with a significant component, where $p > 1$, we can tentatively conclude that there is a p-way interaction between the p predictor variables with respect to their joint relationship to the response variable.)

Let us use the term *associated lower-level components* to refer to components in an ANOVA table that are associated only with predictor variables that are all involved in a particular higher-level interaction. For example, consider a four-way ANOVA in which the predictor variables are *A*, *B*, *C*, and *D*, and consider the $A \times B \times C$ interaction. Then the associated lower-level components for this interaction are all the components in the ANOVA table that contain one or two of variables *A*, *B*, and *C*. There are six such components, namely *A*, *B*, and *C*, and $A \times B$, $A \times C$, and $B \times C$.

After finding a significant component in an ANOVA table, we can continue (in the order mentioned above) checking components in the table. However, as suggested by Appelbaum and Cramer (1974:340), Aitkin (1978:200), and Cox (1984:16) if we are seeking evidence of relationships between variables, once we have found a significant interaction component we are usually not interested in any of the associated lower-level components. We are usually not interested in these components because the higher-level interaction will have already told us that there is a relationship between the response variable and the associated predictor variables, and the relationship is interactive at the indicated level. And if we are seeking evidence of relationships between variables, checking the associated lower-level components will usually give us no further useful information (beyond giving us more evidence that the relationship exists).

Some writers disagree with the preceding sentence. Sections B.2 - B.5 address arguments that have been made in favor of checking associated lower-level components for statistical significance in the presence of a significant higher-level interaction.

B.2 Lower-Level Relationships Cannot Be Present in the Presence of a Higher-Level Associated Interaction

One argument in favor of checking the associated lower-level components is that such checking enables us to see whether simple relationships or lower-level interactions exist between the response variable and the associated predictor variables. However, if there is an interaction between two or more predictor variables with respect to their joint relationship to the response variable, then (for reasons given below) a simple relationship between the response variable and any of the predictor variables involved in the interaction *cannot exist.* Similarly, if there is an interaction between p predictor variables with respect to their joint relationship to the response variable where $p > 2$, then a lower-level interaction between two or more of the predictor variables involved in the interaction also *cannot exist.* Thus if there is a significant interaction, checking the associated lower-level components of the interaction does not allow us to see whether simple relationships or lower-level interactions exist between the response variable and the associated predictor variables because (for reasons given below) such simple relationships and interactions cannot exist.

The conclusion of the preceding paragraph is derived from the definition of a simple relationship between variables and from the definition of an interaction, which are both given in section 6.10. To help understand the conclusion, let us revisit the definitions. A critical feature of the definitions is that each definition contains a clause that limits each predictor variable to being involved in only one type of relationship (or no relationship) with the response variable at a time. That is, a particular predictor variable can only be involved in a simple relationship or in a particular single interaction at a time, and another relationship between that predictor variable and the response variable *cannot exist*.

The fact that a given predictor variable can only bear a single type of relationship to the response variable at a time is, of course, not an empirical fact, but instead simply results from a definition. Thus we could easily choose some other definition for a simple relationship or an interaction. In particular, we could choose a definition that allows simple relationships and interactions to exist even in the presence of associated higher-level interactions.

However, the requirement of the definitions that a predictor variable can only bear one type of relationship to the response variable at a time is helpful in formulating an easily-grasped concept of a relationship between variables because the requirement simplifies the concept of a relationship without limiting its generality.

In particular, it is important to recognize that the definitions given in section 6.10 of simple and interactive relationships between variables do not prevent us from having *terms* for both an interaction effect and an associated lower-level effect in a model equation at the same time, if we wish. But even when we include terms of various levels in a model equation, it is still perhaps simplest to view the relationship in terms of the highest-level interaction—as a single high-level interaction between all of the associated predictor variables with respect to their joint relationship to the response variable, rather than as a combination of two or more different relationships. From this point of view, once we decide that there is an interaction, there is no need to examine associated lower-level components to see if there is any evidence that the associated lower-level relationships exist because (by definition) they cannot.

B.3 The Final Form of the Model Equation: Practical Considerations

If we are using an overparameterized model equation, another argument in favor of testing the associated lower-level components for significance in the presence of a significant higher-level interaction is that such tests tell us which terms to include in the final form of the model equation. However, knowledge of whether we need to include a simple-effect (or lower-level interaction) term in a model equation is often of little practical use because often in performing ANOVA we are not interested in knowing which terms are required in the final form of the model equation, and we are only interested in knowing which relationships, if any, exist between the response variable and the predictor variables.

We are usually not interested in the final form of the model equation because the model equations used in ANOVA are usually of little use beyond providing a way of understanding what is being tested in a statistical test, and for enabling analysis of the power of statistical tests, and for providing an easily-understood way of describing the computation of sums of squares. That is, ANOVA model equations are not often used in substantive areas to describe relationships between variables or to make predictions. ANOVA model equations are not often used in substantive areas because the discrete-valued terms that such equations contain are cumbersome, and the same descriptions and predictions can be made more easily and with better understanding from the (appropriately collapsed, if necessary) table of predicted means of the values of the response variable for the different cells in the group treatment table, or (perhaps better) from a graph based on the (appropriately collapsed) table of means. (For unsaturated model equations, we may have to obtain estimates of the values of the parameters to enable us to use the model equation to compute the associated collapsed predicted means, but once these means are computed, the parameter estimates and the model equation are of little further use.) Of course, model equations that are *not* ANOVA model equations are often used in substantive areas, especially in the hard sciences, to describe relationships between variables. Such model equations usually contain continuous terms rather than discrete-valued terms.

B.4 The Final Form of the Model Equation: Theoretical Considerations

Another argument in favor of checking the associated lower-level components for significance in the presence of a significant higher-level interaction is that such checks tell us which terms to include in the final form of the model equation, and this knowledge may be useful to scientific theoreticians in understanding the underlying processes in the entities under study.

However, model equations derived in standard ANOVA are not often viewed as reflecting (in any illuminating way) the underlying processes in the entities because the equations are designed to be capable of handling *any* set of means of the values of the response variable for the various cells in the group treatment table—that is, the equations are completely general—and (apart from showing which type of relationship has been found, and if interaction is found, that a model equation with only simple effect terms will not work) such completely general model equations usually cannot provide much insight into the underlying processes in the entities. (On the other hand, the appropriately collapsed cell means can provide useful insight because they can assist the theoretician to test the fit of a specific function to the data to describe the relationship between the response variable and the relevant predictor variables.)

B.5 A Graphical Argument

The lack of usefulness of performing a statistical test of a simple relationship or interaction when there is evidence that the predictor variable(s) associated with the effect or interaction are all involved in a significant higher-level interaction can also be seen in terms of graphs. Performing a statistical test of a lower-level relationship when the predictor variable(s) associated with the relationship is (are all) involved in a significant higher-level interaction is equivalent to looking at a collapsed graph of the predicted cell means of the values of the response variable—collapsed over all the predictor variables except the variables associated with the lower-level relationship. But we should generally never collapse over predictor variables involved in a significant interaction because the resulting graph will be concealing information about the true nature of the relationship between the response variable and the predictor variables. And predictions of the value of the response variable for new individual entities (in the same population under the same conditions) made on the basis of this graph (or on the basis of the associated collapsed cell means) will be less accurate than predictions made on the basis of the graph (or cell means) of the full interaction.

B.6 Exceptions

Despite the points made above, there are rare situations in which it *is* necessary to perform a lower-level statistical test even though an associated predictor variable is also clearly involved in a higher-level interaction. Tukey (1977) gives an example of such a situation, which I have enhanced slightly for realism:

Suppose we work in the marketing department of a candy manufacturer, and suppose a particular brand of candy we sell is available in two flavors, and suppose, for budgetary reasons, it is necessary to discontinue selling one of the flavors and, for obvious reasons, we have decided to discontinue selling the flavor that has the lesser sales. To evaluate the relative sales of the two flavors, consider an observational study in which the entities are a random sample of people in the geographic area where the candy is sold, the response variable is the volume of sales of each flavor of the candy to each person, and the predictor variable is the flavor of the candy. Suppose we have performed this (repeated measurements) study, and suppose we also happen to know the sex of each person who participated in the study, and suppose our application of ANOVA to the data reveals that there is an interaction between flavor and sex with respect to their joint relationship to sales. That is, the difference in sales between the two flavors is different for females from the difference in sales between the two flavors for males.

Despite the fact that we have found an interaction, we still need to know whether one flavor sells significantly better overall than the other, so that we can confidently discontinue selling the lesser-selling flavor. Thus despite the interaction, we still need to perform a statistical test of the simple relationship between flavor and sales to determine if we have good evidence as to which flavor sells better than the other.

[Note that in performing the statistical test of the simple relationship between flavor and sales, we should use the HTO sum of squares because the HTI sum of squares is effectively comparing the collapsed-over-sex *mean of the two cell means* for the first flavor with the collapsed-over-sex mean of the two cell means for the second flavor and, of course, this comparison will fail to detect a difference if the two means of the cell means are identical. However, if the two population sales means (across sex) of the cell means are identical, and if there are more females than males in the population (or more males than females), this does *not* imply that the two flavors have identical sales in the population. Instead, if the population ratio of females to males differs substantially from 1:1, it is possible to build a situation (specifically, a particular type of interaction between flavor and sex) in which there is a substantial difference in overall sales between the two flavors, but the two collapsed-over-sex means of the population cell means are identical. In that situation (and in situations approaching that situation) the HTI test can be expected to fail to detect the difference in overall sales between the two flavors. On the other hand, if the sample is random, it is not possible to build a situation in which the HTO test can be expected to fail to detect an existing difference in overall sales between the two flavors (although, naturally, the power of the HTO test varies with the form of the relationship and the pattern of imbalance).]

The candy sales example shows a practical situation in which we will still be interested in performing a statistical test for a simple relationship (main effect) even though the predictor variable involved in the simple relationship is also involved in a significant interaction. However, it is important to note that the example is unusual because (unlike most other research projects) in the example we are not interested in optimizing the value of the response variable in individual entities, but instead we are interested in optimizing (i.e., maximizing) the *sum* of the values of the response variable across all the entities in the population. (That is, we are interested in maximizing sales across people.) If it were not for the unusual summing feature of this research project, it would generally not make sense to perform the test for the simple relationship in the presence of the higher-level interaction.

That is, if we wish to maximize the sales of the candy to *individuals*, and if there is an interaction between flavor and sex, and if we are only able to offer one flavor to each sex, then we should offer to sell to each individual the flavor that sells best to that individual's sex, with (in view of the interaction) it being possible that one flavor will sell better to one sex, and the other flavor will sell better to the other. Or, in more general terms, if we wish to predict (or possibly control) the values of the response variable in new entities that are similar to those used in the research project (as opposed to wishing to predict or control the *sum* of the values of the response variable *across* such entities), we should view the relationship between the response variable and the predictor variables in its full complexity because only from that point of view can we make the best predictions (or possibly exercise the best control).

Elston and Bush (1964:686) and Frane and Jennrich (1977) give examples that are similar to Tukey's in that they reflect situations in which the scientist is interested in optimizing the *sum* of the values of the response variable across all the entities in the research project as opposed to optimizing the value of the response variable in individual entities.

Thus despite the preceding unusual example, it is still reasonable to conclude from the earlier arguments that enlightened scientists will *usually* not be interested in examining lower-level relationships between the response variable and the predictor variables once they have concluded that an associated higher-level interaction exists.

B.7 Valid Means of Rebuttal

Note that the arguments in sections B.2 to B.5 are not incontrovertible because it is logically impossible to prove that no (further) reasonable use exists for a lower-level test in the presence of a significant associated higher-level interaction. Thus to disprove the arguments given in those sections one need only describe a frequently applicable *reasonable* use for an associated lower-level test in the presence of a significant higher-level interaction.

Here the important word is "reasonable". If we view statistics as a tool of scientific research, any acceptable argument about a reasonable use of lower-level tests in the presence of an associated higher-level interaction must be couchable in terms of an example of a "reasonable" scientific research project. Thus examples with toy data or examples that are designed to teach statistics, but that scientists in a substantive area would not consider reasonable, are themselves not reasonable. And for an argument to be convincing, it must present a situation in which an enlightened scientist (not just a statistician) who is studying relationships between variables would have a clear interest in knowing whether a lower-level test is significant in the presence of a believably significant associated higher-level interaction. This paper claims that such situations, although they do occur, are rare.

Nelder (1977) and Aitkin (1978) and their discussants give further thoughts on the usefulness of performing a statistical test for a simple relationship or lower-level interaction in the presence of a higher-level interaction involving the same predictor variables.

B.8 Breakdown Analyses

I have discussed reasons why if we are seeking evidence of relationships between variables, we are usually not interested in testing associated lower-level ANOVA components in the presence of a believably significant higher-level interaction. However, the procedure of testing associated lower-level components should be clearly distinguished from the procedure of performing a "breakdown analysis" of an interaction. A breakdown analysis consists of analyzing one or more *subsets of the full data.* In the simplest case each subset contains data from only one of the levels of one of the predictor variables that are in the interaction, but includes data for all of the levels of all of the other predictor variables that are in the interaction. (More generally, a breakdown analysis can include data for a subset of the levels of a subset of the predictor variables that are in the interaction, and can also include data for all of the levels of some or all of the other predictor variables that are in the interaction or that are in the experiment.) Although breakdown analyses sacrifice some power and generality, their use is sometimes justified because they can help to simplify complicated relationships.

B.9 Extant But Non-Significant Higher-Level Interactions

I have concluded that a researcher usually need not be interested in performing statistical tests of lower-level components once he or she has decided that the associated predictor variables are all involved in a higher-level interaction if the researcher is seeking evidence of relationships between variables. Therefore, if we are seeking evidence of relationships between variables, the fact (noted at the end of section 16) that some lower-level tests are valid only in the absence of the associated higher-level interactions is generally not a problem in situations in which there is a significant higher-level interaction because, in that case, we are usually not interested in performing the associated lower-level tests.

But suppose an interaction exists in the population, but a statistical test of the interaction fails to find evidence of it. If the effects of the interaction are just below the level of detectability, then the presence of the interaction terms in the HTO test of a lower-level relationship *may* make it more likely that the lower-level relationship test *will* be significant. However, this is desirable since the existence of the interaction implies that there is a relationship between the response variable and the associated predictor variables, and we would like at least one of our statistical tests to find some evidence of it.

On the other hand, if an interaction exists in the population, but our statistical test of the interaction fails to find evidence of it, then the presence of the undetected interaction terms in the HTO test of a lower-level relationship *may* make it *less* likely that the lower-level test will be significant. Further investigation of this possibility is needed. However, the fact [as can be seen by comparing (11) and (12) with (6), (13), and (14)] that the HTO statistical test appropriately weights (using the cell counts) each of the estimates of the values of the parameters according to its relative accuracy while the HTI test uses no such weighting suggests that the HTO test will not often (if ever) be less powerful than the HTI test in this special situation.

B.10 Summary

In section 16 I noted that the HTO sums of squares are appropriate if our goal is to test for evidence of a relationship between variables *in the absence of associated higher-level interactions.* In this appendix I discussed why the qualification about higher-level interactions usually does not give rise to problems.

APPENDIX C: PROOF OF THEOREM 2

Theorem 2 was first stated for the two-way case by Yates (1934:66). As noted by Knoke (1987), theorem 2 can be derived from the fact that the maximum likelihood estimator has maximum power in an ANOVA context, as suggested by Wald (1942). Theorem 2 was directly proved geometrically for the two-way case by Burdick and Herr (1980:239) and algebraically for the two-way case by Hocking (1985:152).

The following geometric proof for the *n*-way case is due to Donald S. Burdick:

Definition: The *response vector* is the vector of all the usable values of the response variable that were obtained in a research project, one element of the vector for each usable value of the response variable that was obtained.

Definition: The *observation space* is the space in which the response vector lies.

Definition: The *population mean vector* lies in the observation space and is identical to the response vector except that each elemental value is replaced by the (generally unknown) population mean value of the response variable for the cell in the group treatment table associated with that element.

The definition of the population mean vector constrains it to lie in a subspace of the observation space called the *estimation space.* The dimension of the estimation space is equal to the number of cells in the group treatment table (although, of course, the number of components in a vector in the estimation space remains the same as the number of components in the response vector).

For any given effect *E*, the associated sum of squares, whether HTO or HTI, is the squared length of the projection of the response vector on a subspace of the observation space. The power of a statistical test of *E* is an increasing function of the noncentrality, which is in turn an increasing function of the length of the same projection of the population mean vector on the subspace.

Using the standard sequential—i.e., orthogonal—partitioning we can decompose the population mean vector into the sum of its projections on two subspaces of the estimation space:

- the subspace of the other effects at the Same or Lower level as *E* (which I call the *SL subspace)*
- the subspace of vectors in the estimation space that are orthogonal to the SL subspace (which I call the *test subspace*).

The test subspace contains the subspace associated with the HTO sum of squares for *E*. (This is because the HTO sum of squares is identical to the next sum of squares in the orthogonal partitioning mentioned in the preceding paragraph.)

The test subspace also contains the subspace associated with the HTI sum of squares for *E*. (This is because the HTI sum of squares is identical to the next sum of squares in an orthogonal partitioning in which the other effects at the same and lower levels and all the effects at the higher levels have already been fitted.)

If there are no higher-level interactions, the component of the population mean vector in the test subspace lies entirely in the subspace associated with the HTO sum of squares. (This is because the HTO subspace is capable of capturing—in the absence of higher-level interactions—all of the variation in the mean vector apart from that captured in the SL subspace.) That in turn implies that the component of the population mean vector in the HTI subspace is a projection of the HTO component with the result that the HTI component is a leg of a right triangle of which the HTO component is the hypotenuse. Since a leg can be no longer than a hypotenuse, the noncentrality for HTI can be no greater than the noncentrality for HTO.

And thus the theorem is proved.

I now summarize two simple matrix algebra procedures that illustrate the ideas of theorem 2. You may find it helpful to program the procedures in a matrix algebra computer language, after which you can try them on different datasets.

You can obtain the matrix of the projection and the resulting sum of squares for the HTO, HTI, HTOS (appendix D), or the sequential statistical tests with the following procedure, which is based on a discussion by Hocking (1985:152-155):

- specify $\boldsymbol{y}$, the response vector, which we assume has n elements
- specify **C,** a cell-means contrast matrix of the hypothesis to be tested; **C** has one row for each degree of freedom in the hypothesis and it has one column for each cell in the group treatment table; each row of **C** specifies a different contrast of the cell means; collectively these contrasts define a hypothesis to be tested; see below for a procedure to generate different versions of **C**
- specify **W,** the counting (incidence) matrix; **W** has n rows and it has one column for each cell in the group treatment table; an entry in a cell of **W** is 1 if the y-value associated with the row is associated with the group-treatment-table cell that is associated with the column, and zero otherwise
- compute **H**, a response-vector contrast matrix of the hypothesis to be tested, by using **W** to scale **C**:
  $$\mathbf{H} = \mathbf{C}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$$
  H has one row for each degree of freedom in the hypothesis and n columns
- compute **P**, the response-vector projection matrix, by squaring and normalizing **H**:
  $$\mathbf{P} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{H} \qquad (15)$$
  P has n rows and n columns
- compute $\boldsymbol{p}$, the projection of the response vector:
  $$\boldsymbol{p} = \mathbf{P}\boldsymbol{y}$$
  $\boldsymbol{p}$ has n elements
- compute SS, the desired ANOVA sum of squares, by computing the squared length of $\boldsymbol{p}$*:*
  $$\text{SS} = \boldsymbol{p}'\boldsymbol{p}.$$

Graybill (1983:69-76, 434-437) discusses the algebra of projections.

If you are using a matrix algebra computer package to do the arithmetic, specifying **C** is the only difficult step in using the foregoing procedure. (Of course, the rows of **C** are simply variations of the contrasts that are used in the balanced case. These variations can be obtained by multiplying the balanced-case coefficients in a row by various functions of the cell counts.) For the two-way case you can derive different versions of **C** from (6) and (11) or from the formulas given by Speed, Hocking, and Hackney (1978, table 1), Herr and Gaebelein (1978, table 2), and Kutner (1974).

For the general case you can use the following procedure to generate an **H** matrix directly, thereby bypassing the need to specify **C** and **W**. Page references below are to Searle's *Linear Models for Unbalanced Data* (1987) and references to IML are to the *SAS IML User's Guide* (SAS Institute, 1988):

- specify the full (but not necessarily saturated) overparameterized model equation [e.g., (8) above] and the reduced overparameterized model equation [e.g., (7) above] whose residual sums of squares you wish to difference to obtain the desired sum of squares
- specify a simple full-rank design sub-matrix for each predictor variable in the ANOVA; a simple full-rank design sub-matrix has *n* rows and it has as many columns as there are degrees of freedom associated with the predictor variable (i.e., one less than the number of values that the associated predictor variable had in the research project); each column reflects a contrast of the values of the response variable for different values of the predictor variable that is associated with the sub-matrix; you can use the SAS IML DESIGNF function to generate a simple full-rank design sub-matrix from the vector of the raw values of a predictor variable
- specify full-rank design sub-matrices for all of the interactions; a full-rank design sub-matrix for an interaction consists of the horizontal direct product of all the simple full-rank design sub-matrices for predictor variables that are associated with the interaction; you can use the SAS IML HDIR function to compute horizontal direct products of matrices
- specify $\mathbf{X}_2$ as the full-rank design sub-matrix for the particular component of the ANOVA for which you wish to compute the sum of squares (that is, $\mathbf{X}_2$ is one of the design sub-matrices you specified in the preceding two steps)
- specify $\mathbf{X}_1$ as the horizontal concatenation of all of the design sub-matrices whose associated terms are in the full model equation specified above including a vector of 1's for the constant term μ but excluding the design sub-matrix for the term being tested; you can use the SAS IML horizontal concatenation operator || to perform horizontal concatenation of matrices
- compute $\mathbf{M}_1$ as defined in Searle's equation (76) on pages 263 and 318:
$$\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1\mathbf{X}_1^+$$
where $\mathbf{X}^+$ denotes the Moore-Penrose generalized inverse of **X**
- compute **A** as a normalized form of $\mathbf{M}_1\mathbf{X}_2$, as defined by Searle's equation (82) on page 264 and equation (90) on pages 272 and 318:
$$\mathbf{A} = \mathbf{M}_1\mathbf{X}_2(\mathbf{M}_1\mathbf{X}_2)^+$$
- if *d* is the number of degrees of freedom in the hypothesis, then the *d* rows of an **H** matrix are equivalent to the *d* characteristic vectors (eigenvectors) of **A** that correspond to the *d* nonzero characteristic values (eigenvalues) as shown by Searle in subsection ***-iv*** on page 235.

After computing an **H** matrix with the foregoing procedure, you can then, if you wish, compute a **C** matrix from **H** as **HW**.

Note that for a given approach to computing a sum of squares, **C** and **H** are generally not unique. That is, any set of linearly independent vectors that generate the subspace of the projection represent a valid statement of the hypothesis being tested (in terms of contrasts of the response vector). Of course, all the different statements of the hypothesis are different but equivalent expressions of the same test of the hypothesis that there is a relationship between the response variable and the particular predictor variable(s) associated with the test.

In particular, $(\mathbf{M}_1\mathbf{X}_2)'$ is another statement of the hypothesis being tested (in terms of contrasts of the response vector) and thus can be used in place of **H** in (15) to compute **P**, thereby bypassing the need to compute **A** and its characteristic vectors.

I noted above that an ANOVA sum of squares is equal to the squared length of a projection **Py**. Another frequently used general theoretical formula for computing ANOVA sums of squares is the quadratic form **y'Ay** where **A** is as defined above. Thus

$$\begin{aligned}\mathbf{y}'\mathbf{A}\mathbf{y} &= (\mathbf{P}\mathbf{y})'\mathbf{P}\mathbf{y} \\ &= \mathbf{y}'(\mathbf{P}'\mathbf{P})\mathbf{y}\end{aligned}$$

thus illustrating that **P** is simply a partitioning ("square root") of **A**.

APPENDIX D: AN EXTENSION TO THE HTO APPROACH FOR THREE-WAY AND HIGHER EXPERIMENTS

D.1 The HTOS Approach to Computing ANOVA Sums of Squares

I have concluded that for the ANOVA statistical test of any particular type of relationship (i.e., a particular simple relationship or interaction) between the response variable and the predictor variables:

- in the absence of higher-level interactions, the HTO method of computing sums of squares is preferred to the HTI method

- in the presence of higher-level interactions that involve *all of the predictor variables associated with the relationship being tested*, we are usually not interested in performing the lower-level statistical test.

Let us now turn to the situation in which there is a higher-level interaction in the population among some of the predictor variables (with respect to their joint relationship to the response variable) but this interaction is *not* associated with *all* of the predictor variables that are associated with the particular relationship between variables that we are currently interested in testing. (In fact, the interaction may not even be associated with *any* of the predictor variables associated with the relationship.) This situation occurs in some experiments with three or more predictor variables. As we shall see, in this situation we are sometimes interested in performing lower-level statistical tests in the presence of higher-level interactions, and thus it is of interest to compare various methods of computing the relevant sums of squares for these tests.

For example, suppose we wish to analyze the results of a fully crossed three-way experiment in which the predictor variables are *A*, *B*, and *C*. We can write the saturated overparameterized model equation for the relationship between the response variable and the predictor variables as:

$$\begin{aligned} y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k \\ + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \\ + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}. \end{aligned}$$

where the $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ terms represent the three two-way interactions and the $(\alpha\beta\gamma)_{ijk}$ term represents the three-way interaction.

Suppose our analysis has found no evidence that predictor variable *A* is involved in any of the interactions, and thus we are interested in testing for the *A* simple relationship (main effect). If we use the HTO approach, we will compute the numerator sum of squares for the *A* statistical test by, in effect, differencing the residual sums of squares of the following two generating model equations:

$$y_{ijkl} = \mu \qquad + \beta_j + \gamma_k + \varepsilon_{ijkl}$$

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkl}.$$

Now suppose that we have evidence (through the *B* × *C* statistical test) that a *B* × *C* interaction is present in the population. Then the question arises whether, as well as including terms for predictor variables *B* and *C* as is done in the preceding two equations, we should also include a term for the *B* × *C* interaction in the two generating model equations. That is, the question arises whether we should compute the sum of squares for the *A* simple relationship by differencing the residual sums of squares of the following two model equations:

$$y_{ijkl} = \mu \qquad + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \varepsilon_{ijkl}$$

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \varepsilon_{ijkl}.$$

We represent this new approach to computing sums of squares with the name HTOS because Higher-level Terms are Omitted *except for any Significant higher-level interactions* that are not associated with all of the predictor variables that are associated with the effect being tested. The HTOS approach was proposed by Heiberger and Laster (1977).

Two questions arise about the HTOS approach:

- Is the HTOS approach valid for testing for relationships between the response variable and the predictor variables?
- Given that the HTOS approach is valid, which approach is more powerful—HTOS, HTO, or HTI?

Although the details are beyond the scope of this paper, it can be shown that the HTOS approach is valid. Furthermore, it can be shown that in the presence of higher-level population interactions between the predictor variables that are *not* associated with all of the predictor variable(s) associated with the test (but in the absence of higher-level interactions that *are* associated with all of the predictor variables associated with the test), the HTOS approach generally provides a more powerful statistical test than either the HTO approach or the HTI approach. (Of course, when there are no such non-associated higher-level interactions, the definitions imply that the HTOS and HTO sums of squares are identical, and thus the associated HTOS and HTO statistical tests will have identical power.)

Thus when we are testing for relationships between the response variable and the predictor variables in unbalanced ANOVA with no empty cells in the group treatment table, the HTOS sums of squares are preferred to both the HTO and the HTI sums of squares because the HTOS sums of squares generally provide slightly more powerful statistical tests.

D.2 The HTOS Cutoff Point

When we use the HTOS approach in three-way and higher experiments we need a procedure for deciding whether to include a particular non-associated higher-level interaction term in the two generating model equations for a given statistical test. A reasonable procedure is to define a critical *p*-value for such interactions, and then to use this critical *p*-value as a cutoff point: If the *p*-value for a non-associated higher-level interaction is greater than the cutoff point, we (or the computer) form the two generating model equations for the effect being tested omitting the term for that interaction from both the model equations. But if the *p*-value for the interaction is less than or equal to the cutoff point, we form the two generating model equations for the effect being tested including the term for that interaction in both the equations.

In choosing the cutoff point it seems a more serious error to omit needed terms from the generating equations (which results in a violation of the assumption of independence of the values of the error term) than to include unneeded terms in the equations (which has no effect on the independence of the values of the error terms), and therefore it seems reasonable to choose a relatively high cutoff point for this procedure such as .1 or .2. A cautious scientist may even decide to use a cutoff point of 1, which is the limiting case. Then terms for all the non-associated higher-level interactions will always be included in the two generating model equations at a cost of a slight decrease in power relative to the appropriate HTOS generating model equations.

D.3 How To Compute HTOS Sums of Squares

Currently, the HTOS sums of squares are not directly available in the ANOVA or general linear model procedures of any of the popular statistical computer packages. However, as noted above, the HTOS sums of squares are identical to the HTO sums of squares (which are available in some packages) except when there are significant non-associated higher-level interactions. If the available package cannot compute HTO sums of squares, or if there is a non-associated higher-level interaction, you can compute the necessary HTOS sums of squares by using the procedures discussed in the last two paragraphs of section 13.3.

APPENDIX E: OTHER APPROACHES TO ANOVA STATISTICAL TESTS

In addition to the HTOS, HTO, and HTI methods of computing and performing ANOVA statistical tests, various other methods have been proposed. (Some of these methods can be defined in terms of differencing the residual sums of squares of two generating model equations.) Therefore, it is useful to ask whether another method might be better than any of the preceding methods if we wish to test for evidence of a population relationship between the response variable and the predictor variables in an experiment.

Although detailed discussion of other methods is beyond the scope of this paper, the following criteria are reasonable for evaluating any proposed method of computing and performing ANOVA statistical tests if we wish to test for evidence of a relationship:

- for any proposed method, the resulting tests should be valid in the sense given in section 16
- of all the valid proposed tests of a given relationship, the test of choice should, in general, have the greatest power
- to eliminate arbitrariness, the *p*-values obtained from a test should not depend on an ordering of the predictor variables (as occurs with tests based on so-called sequential or hierarchical sums of squares) unless there is a clear rationale for using such an ordering
- to lessen the need for a formal procedure to correct for the multiplicity of statistical tests (because such a procedure will often substantially diminish the power of the tests), formally only a single test should be defined for each component in the ANOVA table (although (*a*) as discussed in the preceding section, the definition of any test may be contingent on the outcome of logically prior tests and (*b*) *informally* we may perform as many different tests as we wish to help us to understand the data)
- to lessen the need for a formal procedure to correct for multiplicity, formal tests of pooled components should usually be omitted
- tests can be chosen keeping in mind that they will usually only be used when it is reasonable to assume that associated higher-level interactions are non-existent, as discussed in appendix B.

These criteria rule out some other methods of computing and performing ANOVA statistical tests.

The importance of maximizing the power of the ANOVA statistical tests (while maintaining their validity) implies that it may be reasonable to define criteria for pooling *F*-ratio denominator sums of squares in some ANOVA situations (as discussed by Steinhorst and Everson 1980, and Hocking 1985:150). However, discussion of such denominator pooling should be distinguished from the discussion in this paper, which is about methods of computing *F*-ratio *numerator* sums of squares

REFERENCES

Aitkin, M. 1978. The analysis of unbalanced cross-classifications (with discussion). *Journal of the Royal Statistical Society, Ser. A* 141:195-223.

Aitkin, M. 1983. Unbalanced data (letter). *The American Statistician* 37:93-94.

Allan, F. E., and J. Wishart. 1930. A method of estimating the yield of a missing plot in field experimental work. *Journal of Agricultural Science* 20:399-406.

Appelbaum, M. I., and E. M. Cramer. 1974. Some problems in the nonorthogonal ANOVA. *Psychological Bulletin* 81:335-343.

Bancroft, T. A. 1968. *Topics in Intermediate Statistical Methods, Volume I.* Ames, Iowa: Iowa State University Press.

Barnett, V. 1988. "Relationship". In *Encyclopedia of Statistical Sciences* vol. 8 ed. by S. Kotz and N. L. Johnson. New York: Wiley.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building* New York: Wiley.

Box, G. E. P., and N. R. Draper. 1987. *Empirical Model Building and Response Surfaces.* New York: Wiley.

Bryce, G. R., D. T. Scott, and M. K. Carter. 1980. Estimation and hypothesis testing in linear models - A reparameterization approach to the cell means model. *Communications in Statistics—Theory and Methods* 9:131-150.

Burdick, D. S. 1979. Rows adjusted for columns versus simple row means in two-way ANOVA: When are the two hypotheses equivalent? *Journal of the American Statistical Association* 74:457-458.

Burdick, D. S., and D. G. Herr. 1980. Counterexamples in unbalanced two-way ANOVA. *Communications in Statistics—Theory and Methods* 9:231-241.

Burdick, D. S., D. G. Herr, W. M. O'Fallon, and B. V. O'Neill. 1974. Exact methods in the unbalanced, two-way analysis of variance—A geometric view. *Communications in Statistics* 3:581-595.

Calinski, T. 1982. On some problems in analyzing nonorthogonal designs. In *COMPSTAT, Fifth Symposium on Computational Statistics.* International Association for Statistical Computing.

Carlson, J. E., and N. H. Timm. 1974. Analysis of nonorthogonal fixed effect designs. *Psychological Bulletin* 81:563-570.

Chowdhury, A. R. 1987. Are causal relationships sensitive to causality tests? *Applied Economics* 19:459-465.

Cochran, W. G., and G. M. Cox. 1957. *Experimental Designs* 2nd ed. New York: Wiley.

Cox, D. R. 1984. Interaction (with discussion). *International Statistical Review* 52:1-31.

Cox, D. R. 1958. *Planning of Experiments*. New York: Wiley.

Cramer, E. M., and M. I. Appelbaum. 1980. Nonorthogonal ANOVA—Once again. *Psychological Bulletin* 87:51-57.

Davidson, M. L., and J. D. Toporek. 1979. *ANOVA With General Cell Weights. BMDP Technical Report #59.* Los Angeles, CA: BMDP Statistical Software, Inc.

Dixon, W. J. 1990a. *BMDP Statistical Software Manual* (vol. 2). Berkeley, CA: University of California Press.

Dixon, W. J. 1990b. *BMDP Statistical Software Manual* (vol. 1). Berkeley, CA: University of California Press.

Elliott, A. C., and W. A. Woodward. 1986. Analysis of an unbalanced two-way ANOVA on the microcomputer. *Communications in Statistics—Simulation and Computation* 15:215-225.

Elston, R. C., and N. Bush. 1964. The hypotheses that can be tested when there are interactions in an analysis of variance model. *Biometrics* 20:681-698.

Finney, D. J. 1960. *An Introduction to the Theory of Experimental Design.* Chicago: University of Chicago Press.

Finney, D. J. 1987. Why did he not ...? *The American Statistician* 41:244-245.

Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd. The 8th edition of this seminal work appears in Fisher (1990).

Fisher, R. A. 1990. *Statistical Methods, Experimental Design, and Scientific Inference* edited by J. H. Bennett. Oxford: Oxford University Press.

Fisher, R. A., and W. A. Mackenzie. 1923. Studies in crop variation: II. The manurial response of different potato varieties. *Journal of Agricultural Science* 13:311-320.

Francis, I. 1973. A comparison of several analysis of variance programs. *Journal of the American Statistical Association* 68:860-865.

Frane, J. W. 1979. *Some Computing Methods for Unbalanced Analysis of Variance and Covariance. BMDP Technical Report #52.* Los Angeles, CA: BMDP Statistical Software, Inc. Also published in 1980 under the same title in *Communications in Statistics—Theory and Methods* 9:151-166.

Frane, J., and R. Jennrich. 1977. Discussion of an article by J. A. Nelder. *Journal of the Royal Statistical Society, Ser. A* 140:73.

Golhar, M. B., and J. R. Skillings. 1976. A comparison of several analysis of variance programs with unequal cell size. *Communications in Statistics—Simulation and Computation* 5:43-53.

Goodnight, J. H. 1980. Tests of hypotheses in fixed effects linear models. *Communications in Statistics—Theory and Methods* 9:167-180.

Gosslee, G. D., and H. L. Lucas. 1965. Analysis of variance of disproportionate data when interaction is present. *Biometrics* 21:115-133.

Gottesman, I., L. Mandarino, C. Verdonk, R. Rizza, and J. Gerich. 1982. Insulin increases the maximum velocity for glucose uptake without altering the Michaelis constant in man. *Journal of Clinical Investigation* 70:1310-1314.

Granger, C. W. J. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2:329-352.

Graybill, F. W. 1983. *Matrices with Applications in Statistics* (2nd ed). Belmont, CA: Wadsworth.

Heiberger, R. M., and L. L. Laster. 1977. Maximizing power in the unbalanced higher-way ANOVA. *Proceedings of the American Statistical Association Statistical Computing Section* 234-239.

Helms, R. W. 1988. Comparisons of parameter and hypothesis definitions in a general linear model. *Communications in Statistics—Theory and Methods* 17:2725-2753.

Herr, D. G. 1986. On the history of ANOVA in unbalanced, factorial designs: The first 30 years. *The American Statistician* 40:265-270.

Herr, D. G., and J. Gaebelein. 1978. Nonorthogonal two-way ANOVA. *Psychological Bulletin* 85:207-216.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey. 1991. *Fundamentals of Exploratory Analysis of Variance.* New York: Wiley.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Comparison Procedures.* New York: Wiley.

Hocking, R. R. 1985. *The Analysis of Linear Models*. Monterey, CA: Brooks/Cole.

Hocking, R. R., O. P. Hackney, and F. M. Speed. 1978. “The Analysis of Linear Models with Unbalanced Data.” In *Contributions to Survey Sampling and Statistics, Festschrift for H. O. Hartley* edited by H. A. David. New York: Academic Press.

Hocking, R. R., and F. M. Speed. 1975. A full rank analysis of some linear model problems. *Journal of the American Statistical Association* 70:706-712.

Hocking, R. R., and F. M. Speed. 1976. Letter to the editor. *The American Statistician* 30:206.

Hocking, R. R., F. M. Speed, and A. T. Coleman. 1980. Hypotheses to be tested with unbalanced data. *Communications in Statistics—Theory and Methods* 9:117-129.

Howell, D. C., and S. H. McConaughy. 1982. Nonorthogonal ANOVA: Putting the question before the answer. *Educational and Psychological Measurement* 42:9-24.

Irwin, J. O. 1931. Mathematical theorems involved in the analysis of variance. *Journal of the Royal Statistical Society* 94:284-300.

Johnson, R. E., and D. G. Herr. 1984. The effect of nonorthogonality on the dependence of *F*-ratios sharing a common denominator. *Journal of the American Statistical Association* 79:702-708.

Kempthorne, O. 1952. *The Design and Analysis of Experiments.* New York: Wiley.

Kendall, M., A. Stuart, and J. K. Ord. 1987. *Kendall’s Advanced Theory of Statistics*. Vol 1. London: Charles Griffin.

Keren, G., and C. Lewis. 1976. Nonorthogonal designs—sample versus population. *Psychological Bulletin* 83:817-826.

Knoke, J. D. 1987. Unique analyses for unbalanced, factorial designs. *The American Statistician* 41:245.

Kutner, M. H. 1974. Hypothesis testing in linear models (Eisenhart Model I). *The American Statistician* 28:98-100.

Kutner, M. 1975. Correction. *The American Statistician* 29:70.

Lipsey, M. W. 1990. “Theory as Method: Small Theories of Treatments”. In *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data* ed. by L. Sechrest, E. Perrin, and J. Bunker. Rockville MD: U. S. Department of Health and Human Services.

Littell, R. C., and R. O. Lynch. 1983. “Power Comparison of the Weighted Squares of Means Method (Type III) and the Method of Fitting Constants (Type II) in Unbalanced ANOVA. In *Proceedings of the Eighth Annual SAS Users Group International Conference.* Cary, NC: SAS Institute. 757-764.

Macdonald, R. R. 1991. The importance of being OSCAR or Balance and the analysis of factorial designs. *British Journal of Mathematical and Statistical Psychology* 44:207-220.

Macnaughton, D. B. 1997. *How to Predict the Unknown: An Introduction to Scientific Thinking With Statistics.* Unpublished manuscript.

Milligan, G. W., D. S. Wong, and P. A. Thompson. 1987. Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin* 101:464-470.

Minitab, Inc. 1989. *Minitab Reference Manual Release 7*. State College, PA: author.

Mosteller, F. 1990. “Improving Research Methodology: An Overview”. In *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data* ed. by L. Sechrest, E. Perrin, and J. Bunker. Rockville MD: U. S. Department of Health and Human Services.

Nelder, J. A. 1977. A reformulation of linear models (with discussion). *Journal of the Royal Statistical Society, Ser. A* 140:48-76.

Nelder, J. A. 1982. Linear models and non-orthogonal data. *Utilitas Mathematica* 21B:141-152.

O’Brien, R. G. 1976. Comment on some problems in the nonorthogonal analysis of variance. *Psychological Bulletin* 83:72-74.

Overall, J. E., D. M. Lee, and C. W. Hornick. 1981. Comparison of two strategies for ANOVA in nonorthogonal designs. *Psychological Bulletin* 90:367-375.

Overall, J. E., and D. K. Spiegel. 1969. Concerning least squares analysis of experimental data. *Psychological Bulletin* 72:311-350.

Overall, J. E., D. K. Spiegel, and J. Cohen. 1975. Equivalence of orthogonal and nonorthogonal ANOVA. *Psychological Bulletin* 82:182-186.

Parker, S. P. (ed.) 1989. *McGraw-Hill Dictionary of Scientific and Technical Terms.* New York: McGraw-Hill.

Pendleton, O. J., M. Von Tress, and R. Bremer. 1986. Interpretation of the four types of analysis of variance tables in SAS. *Communications in Statistics—Theory and Methods* 15:2785-2808.

Poirier, D. J. 1988. Causal relationships and replicability. *Journal of Econometrics* 39:213-234.

Rubin, D. B., T. W. F. Stroud, and D. T. Thayer. 1981. Fitting additive models to unbalanced two-way data. *Journal of Educational Statistics* 6:153-178.

SAS Institute Inc. 1988. *SAS IML User’s Guide, Release 6.03 Edition*. Cary, NC: author.

SAS Institute Inc. 1990. *SAS/STAT User’s Guide, Version 6,* 4th ed. Vol 1. Cary, NC: author.

Scheffé, H. 1959. *The Analysis of Variance*. New York:

Wiley.

Schmoyer, R. L. 1984. Everyday application of the cell means model. *The American Statistician* 38:49-52.

Searle, S. R. 1971. *Linear Models*. New York: Wiley.

Searle, S. R. 1980. Arbitrary hypotheses in linear models with unbalanced data. *Communications in Statistics—Theory and Methods* 9:181-200.

Searle, S. R. 1981. Quirks in linear model calculations with unbalanced data. *Proceedings of the Sixth Annual SAS Users Group International Conference.* 38-45.

Searle, S. R. 1987. *Linear Models for Unbalanced Data.* New York: Wiley.

Searle, S. R. 1988. Best linear unbiased estimation in mixed models of the analysis of variance. In *Probability and Statistics Essays in Honor of Franklin A. Graybill,* edited by J. Srivastava. Amsterdam: North-Holland.

Searle, S. R., F. M. Speed, and H. V. Henderson. 1981. Some computational and model equivalences in analysis of variance of unequal-subclass-numbers data. *The American Statistician* 35:16-33.

Singh, K. P., and U. Singh. 1989. A technique for the analysis of unbalanced data. *Biometrical Journal* 31:3-17.

Spector, P. E., N. H. Voissem, and W. L. Cone. 1981. A Monte Carlo study of three approaches to nonorthogonal analysis of variance. *Journal of Applied Psychology* 66:535-540.

Speed, F. M. 1969. A new approach to the analysis of linear models. NASA Technical Memorandum NASA TM X-58030. Houston, TX: NASA Manned Spacecraft Center.

Speed, F. M., and R. R. Hocking. 1976. The use of R()-notation with unbalanced data. *The American Statistician* 30:30-33.

Speed, F. M., R. R. Hocking, and O. P. Hackney. 1978. Methods of analysis of linear models with unbalanced data. *Journal of the American Statistical Association* 73:105-112.

SPSS, Inc. 1990. *SPSS Reference Guide.* Chicago, IL: author.

Steinhorst, R. K. 1982. Resolving current controversies in analysis of variance. *The American Statistician* 36:138-139.

Steinhorst, R. K., and D. O. Everson. 1980. Nonorthogonal analysis of balanced data. *Proceedings of the Fifth Annual SAS Users Group International Conference,* 219-223. Cary, NC: SAS Institute.

Tukey, J. 1977. Discussion of an article by J. A. Nelder. *Journal of the Royal Statistical Society, Ser. A* 140:72.

Tukey, J. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.

Tukey, J. 1989. "SPES in the Years Ahead". In *Proceedings of the American Statistical Association Sesquicentennial Invited Paper Sessions.* Alexandria, Virginia: The American Statistical Association.

Turner, D. L. 1990. An easy way to tell what you are testing in analysis of variance. *Communications in Statistics—Theory and Methods* 19:4807-4832.

Urquhart, N. S., and D. L. Weeks. 1978. Linear models in messy data: Some problems and alternatives. *Biometrics* 34:696-705.

Urquhart, N. S., D. L. Weeks, and C. R. Henderson. 1973. Estimation associated with linear models: A revisitation. *Communications in Statistics* 1:303-330.

Velleman, P. F. n.d. *Data Desk Quickstart Guide.* Northbrook, IL: Odesta Corp.

Wald, A. 1942. On the power function of the analysis of variance test. *Annals of Mathematical Statistics* 13:434-439.

Wilkinson, L. 1990. *SYSTAT: The System for Statistics.* Evanston, IL: Systat, Inc.

Winer, B. J. 1971. *Statistical Principles in Experimental Design* 2nd ed. New York: McGraw-Hill.

Yates, F. 1933. The principles of orthogonality and confounding in replicated experiments. *Journal of Agricultural Science* 23:108-145.

Yates, F. 1934. The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association* 29:51-56.

(version of November 28, 1997, except that the note before the abstract was added and typographical errors were corrected on May 10, 1998.)

Mathematical expressions in this document were generated with MathType by Design Science, *http://www.mathtype.com/mathtype/*

CONTENTS

1. Introduction 1
2. History 1
3. Entities 2
4. Properties of Entities 2
5. Values of Properties of Entities 2
6. Relationships Between Properties (Relationships Between Variables) 3
    6.1 Science as a Study of Relationships Between Properties 3
    6.2 Properties as Variables 4
    6.3 Response Variables and Predictor Variables 5
    6.4 A Definition of a Relationship Between Properties (Variables) 5
    6.5 The Null and Alternative Hypotheses 5
    6.6 Experiments and Causation 6
    6.7 ANOVA 6
    6.8 Does ANOVA Detect Relationships Between Variables? 7
    6.9 Why We Need Statistical Tests 7
    6.10 Interactions and Simple Relationships Between Variables 8
    6.11 Summary and Preview 8
7. Group Treatment Tables 8
8. Fully Crossed Experiments, Unbalanced Experiments, and Empty Cells 9
9. Model Equations 10
    9.1 The Cell-Means Model Equation 10
    9.2 The Error Term and the Underlying Assumptions 10
    9.3 The Overparameterized Model Equation 10
    9.4 Using Model Equations To Make Predictions 11
    9.5 The Estimates of the Values of the Parameters In An Overparameterized Model Equation Are Not Unique 11
    9.6 Sigma Restrictions 11
    9.7 The Use of Overparameterized Model Equations 12
10. Relationships Among Parameters 12
    10.1 Review of Relationships Between Variables 12
    10.2 Relationships Among Parameters 12
    10.3 Are Relationships Among Parameters the Same as Relationships Between Variables? 13
    10.4 Why Scientists Test For Relationships Among Parameters 13
11. Summary and Preview 13
12. Residual Sums of Squares 14
13. Two Methods of Computing ANOVA Sums of Squares 14
    13.1 The HTO Method of Computing Sums of Squares 14
    13.2 The HTI Method of Computing Sums of Squares 15
    13.3 General Comments 15
14. The HTO Method Is Generally Inappropriate When Testing For Relationships Among Parameters 17
15. But the HTI Method Is Appropriate When Testing For Relationships Among Parameters 17
16. Both the HTO and HTI Methods Are Appropriate When Testing For Relationships Between the Response Variable and the Predictor Variables 18
17. The HTO Method Is Generally More Powerful Than the HTI Method When Testing For Relationships Between Variables 18
18. Summary 19

Appendix A: Proof of Theorem 1 19

Appendix B: Higher-Level Interactions 20
    B.1 A Reasonable Approach To ANOVA 20
    B.2 Lower-Level Relationships Cannot Be Present In the Presence of a Higher-Level Associated Interaction 21

B.3 The Final Form of the Model Equation: Practical Considerations 22
B.4 The Final Form of the Model Equation: Theoretical Considerations 22
B.5 A Graphical Argument 22
B.6 Exceptions 22
B.7 Valid Means of Rebuttal 23
B.8 Breakdown Analyses 24
B.9 Extant But Non-Significant Higher-Level Interactions 24
B.10 Summary 24
Appendix C: Proof of Theorem 2 24
Appendix D: An Extension To the HTO Approach For Three-Way and Higher Experiments 26
D.1 The HTOS Approach To Computing ANOVA Sums of Squares 26
D.2 The HTOS Cutoff Point 27
D.3 How To Compute HTOS Sums of Squares 28
Appendix E: Other Approaches To ANOVA Statistical Tests 28
References 28