Behavior Research Methods
2009, 41 (3), 901-908
doi:10.3758/BRM.41.3.901

Nominal analysis of "variance"

David J. Weiss
California State University, Los Angeles, California

Nominal responses are the natural way for people to report actions or opinions. Because nominal responses do not generate numerical data, they have been underutilized in behavioral research. On those occasions in which nominal responses are elicited, the responses are customarily aggregated over people or trials so that large-sample statistics can be employed. A new analysis is proposed that directly associates differences among responses with particular sources in factorial designs. A pair of nominal responses either matches or does not; when responses do not match, they vary. That analogue to variance is incorporated in the nominal analysis of "variance" (NANOVA) procedure, wherein the proportions of matches associated with sources play the same role as do sums of squares in an ANOVA. The NANOVA table is structured like an ANOVA table. The significance levels of the *N* ratios formed by comparing proportions are determined by resampling. Fictitious behavioral examples featuring independent groups and repeated measures designs are presented. A Windows program for the analysis is available.

A nominal response is a verbal label. Asking for a nominal response is often the natural way to elicit an opinion. The mechanic making a diagnosis, the potential purchaser choosing a brand, and the possibly prejudiced person expressing a preference are likely to be thinking of a label as they consider their response. Actions that people intend or report having executed are also expressed nominally. In survey research, nominal responses arise frequently.

Experimentalists, on the other hand, generally try to avoid tasks that generate nominal data. In the laboratory, a familiar task may be altered so that numerical data can be provided. For example, suppose an investigator wants to study how physicians evaluate hypothetical patients whose symptoms are varied systematically. So that the judgments can be expressed numerically, the physician might be asked to predict how many months the patient can be expected to survive, or the probability that the patient will survive until a particular point in time. Although the latter judgments are medically pertinent, they are more complex than merely identifying the disease and, perhaps, recommending a regimen, tasks that arise routinely in medical practice.

To be sure, experimenters have been trained to gather numerical data for good reasons. Numerical data provide more information and more analytic power than do nominal data. Quantitative theories are more interesting than qualitative ones, and it seems obvious that numerical data are needed to test a quantitative theory. Simply reporting nominal responses may pose a challenge. How are they to be aggregated? Can nominal data supply more information than can be summarized by a table of frequencies?

Perhaps no single writer has had more influence on the data-gathering proclivities of behavioral researchers than S. S. Stevens (1946, 1951). His discussions of the constraints placed on statistical operations by the scale properties of the responses, although vigorously challenged (e.g., Anderson, 1961; Lord, 1953), still lead experimentalists largely to restrict nominal data to the demographics section of their reports. Even if a researcher were willing to ignore Stevens's proscription, nominal data do not seem suitable for sophisticated experimental work. The elegant factorial designs to which experimenters apply ANOVAs require numerical data. Yet there are many tasks for which nominal responses are the most appropriate, and requiring subjects to use other modes smacks of the carpenter who pounds in screws because the only available tool is a hammer.

In this article, I present a method for analyzing nominal responses collected using factorial designs. The goal is to allow the analytic power afforded by factorial designs to be extended to studies in which nominal responses are the natural way to express a person's actions or judgments. That is, I echo Keppel's (1991) assertion that the intimate connection between design and analysis is conducive to the conceptualization of incisive studies.

Of course, nominal responses do not exhibit variance in the usual sense, since there is no metric that can support measures of distance. One cannot say by how much two nominal responses differ. Still, it is possible to invoke the concept of disparity between responses, in that two responses either match or do not match. The proposed analysis features an orthogonal partitioning of the potential pairwise matches generated by the experimental design. This partitioning corresponds to the orthogonal partitioning of sums of squares characteristic of the ANOVA.

D. J. Weiss, dweiss@calstatela.edu

Nominal Analysis

The fundamental statistic in the proposed nominal ANOVA (NANOVA) is the NANOVA proportion (NP). For main effects, NP is the proportion of nonmatches, calculated by comparing obtained and potential matches associated with a source. The algebraic expression for the proportion is (potential − obtained)/potential. The number of potential matches associated with each of the sources is inherent in the design structure. The correspondence between the NANOVA and the ANOVA is evidenced by the fact that the number of potential matches for each source is the product of the source's *df* and one half the total number of stimulus combinations. Thus, for a design with four levels in Factor A, three levels in Factor B, and two levels in Factor C, the number of potential matches associated with Factor A is 3 (= the *df* for A) × 12 (= 1/2 × 4 × 3 × 2) = 36. The number of potential matches for the BC interaction is 2 × 1/2 × 4 × 3 × 2 = 24. The number of obtained matches is an empirical outcome. Every potential match is assigned to a unique source; this is the sense in which the partitioning is orthogonal.

The rationale for placing nonmatches (the complement of obtained matches) in the numerator of the NP is that when responses to the various levels of a factor are the same (the nominal version of the usual null hypothesis), that factor does not affect the response. Accordingly, the obtained proportion of nonmatches ought to be small if that null hypothesis is true. NPs range between 0 and 1 and are analogous to effect sizes.

Significance questions are addressed in the NANOVA table. NPs play a role similar to that of mean squares in the ANOVA, in that they are combined in a ratio format to yield the test statistic, the *N* ratio. The *N* ratio compares the NP for a substantive source with the NP for the error term associated with that source. The selection of the error term for a source follows the traditional rules of the ANOVA. It should be noted that although proportions play a key role, the NANOVA is not a factorial analysis of proportions as presented by Dyke and Patterson (1952). The data they analyzed were proportions, whereas the data for the NANOVA are individual responses.

Because there are no underlying distributional assumptions, significance tests on *N* ratios are carried out using a resampling procedure (Edgington & Onghena, 2007). The observed scores are randomly permuted without replacement (Rodgers, 2000) a "large" number of times; I use 100,000 as the default large number. After each permutation, *N* ratios for each substantive source are calculated. The proportion of times the *N* ratio derived from permuted data exceeds the *N* ratio from the original data is an estimate of the probability of obtaining an *N* ratio at least that large, given that the null hypothesis is true, thereby corresponding to a *p* value in the ANOVA. The logic is that the resampled data were not truly generated by the factors, so any patterns emerging in the proportions were fortuitous.

Antecedents

A definition of variance for nominal data was first proposed by Gini (1939), who suggested examining all n^2 pairs of ordered responses in a set. A pair that matches generates a difference of 0, and a pair that does not match generates a difference of 1. Gini defined the sum of squares for the set of *n* responses to be the sum over all pairs of the squared differences divided by 1/2*n*. Gini's definition is related to the NANOVA's proportion of nonmatches. Because Gini cleverly chose 0 and 1 as the seemingly arbitrary constants for describing matches and nonmatches ($0^2 = 0$ and $1^2 = 1$), Gini's sum of squares increases as the proportion of nonmatches increases.

Light and Margolin (1971) built a true ANOVA procedure for categorical data on Gini's (1939) definition, partitioning the total sum of squares into between and within terms. They compared their *F* tests to equivalent chi-square contingency statistics and found the former to be more powerful under some circumstances. Onukogu (1985) also constructed an ANOVA for nominal data, but his sums of squares incorporate a slightly different divisor than do those of Gini and of Light and Margolin. Both of these published analyses were illustrated with a large data set.

The NANOVA's key quantities, the NPs, are members of the same family as Gini's (1939) sum of squares but do not yield *F* ratios. The Light and Margolin (1971) and Onukogu (1985) procedures are presented for designs with only one substantive factor. The NANOVA is more general, allowing for multiple factors (which may be crossed or nested) and providing estimates of main effects and interactions.

Multiway contingency analyses (Goodman, 1971; Shaffer, 1973) also apply factorial decomposition to nominal responses and may be seen as competitors to the NANOVA. In this multifactor generalization of traditional chi-square tests of independence, all of the classifications have equivalent status. On the other hand, the NANOVA distinguishes between its dependent variable and its independent variables. From the experimentalist's standpoint, this distinction is meaningful, which is one reason why the ANOVA continues to thrive despite Cohen's (1968) demonstration that equivalent insights can be extracted from multiple regression.

Other competitors include loglinear analysis (Agresti, 1990) and multinomial logit analysis (Grizzle, 1971; Haberman, 1982; McFadden, 1974). Although the latter procedure is applicable to choices among factorially composed stimuli, it is considerably more opaque and entails distributional assumptions that the NANOVA avoids. However, the NANOVA analysis allows for free responding as well as multiple choice using preset options. Removing the constraint threatens to yield data matrices featuring unique responses or empty cells, both of which pose problems of estimability for these generalized linear models (Fienberg, 2000).

The Statistical Model

Bock (1975) proposed a multinomial model as plausible for nominal data, viewing a response as a random variable that assumes an integer value *j* with probability p_j. Responses are presumed to be independent. The binomial model is a special case of the multinomial model used for dichotomous responses, where *j* takes on one of

two possible values. For categorical responses where there are more than two options, j takes on a range of values equal to the number of categories. The p_js are estimated with response proportions. A link function, whose contribution is to address how randomness imposes its effect, determines the particular flavor of the model.

The model proposed for the NANOVA is based on the most widely used of the multinomial models, the multinomial logit model. This model has historical roots in Luce's (1959) general model of choice, in which it is assumed that individuals act rationally to maximize utility. Choices were assigned a random element, because they were held to be dependent on random utilities. McFadden (1974) made the connection explicit, modeling expected utilities on attributes of the stimulus options. An additional layer of complexity, allowing for repeated measures and thus extracting information about individual respondents, was introduced by Jain, Vilcassim, and Chintagunta (1994). I present the model equation using the notation of Chen and Kuo (2001):

$$p_{itj|\mathbf{u}_i} = \Pr\left(y_{it} = j \mid \mathbf{u}_i\right) = \frac{\exp\left(\alpha_j + \mathbf{x}'_{itj}\beta + \mathbf{z}'_{itj}\mathbf{u}_i\right)}{\sum_{j=1}^{J}\exp\left(\alpha_j + \mathbf{x}'_{itj}\beta + \mathbf{z}'_{itj}\mathbf{u}_i\right)}. \quad (1)$$

The equation describes the probability that the ith subject will choose alternative j as the response y_{it} at the tth opportunity, having been exposed to the multidimensional vector $\mathbf{x}_{it}$, which constitutes the current stimulus combination. In specifying the model, $\mathbf{z}_{itj}$ is the known design vector. The vector of unknown random effects, $\mathbf{u}_i$, which is conditional on the ith subject, is assumed to be independent and identically distributed according to a multivariate normal distribution. In the Chen and Kuo (2001) formulation, J represents the fixed number of choices available. For the NANOVA, J should be interpreted as the number of responses actually observed; α_j and β are parameters to be estimated from data.

Because fitting is not part of the NANOVA procedure, the value of the model is primarily heuristic. In utilizing a randomization test, the NANOVA can bypass the distributional assumption of multivariate normality and avoid consideration of asymptotic properties. Accordingly, large samples are not required, although large samples might be expected to convey the usual benefits of stability and power.

Examples

I will use fictional experiments and artificial data to illustrate the kinds of studies for which the NANOVA might be suitable and how the matches are assigned to sources. In the illustrations, I will use single letters as the responses to simplify the presentation; but the responses could as well be words spoken or written by respondents, or they might be verbal descriptions of actions. I first will show two repeated measures analyses using the same design, with data constructed to show a stimulus effect in the first example and a subject effect in the second. These results will serve as a check on the validity of the NANOVA, in that the analyses capture the different effects I built into the data. Then I will show a one-way, independent groups design and compare the analysis with a conventional contingency table analysis, using a chi-square test. The last example, a two-factor independent groups design, will illustrate how the NANOVA handles interaction. Additional examples, including one featuring a mixed design with a nested subjects factor, are posted at www.davidjweiss.com/NANOVA.htm.

Although one can count matches manually, the process becomes mind-boggling as the designs get larger and/or more complex. I recommend use of the NANOVA computer program,[1] which also handles the resampling. The number of resampling repetitions is an option within the program; using the default value of 100,000, none of the analyses took more than 1 min on a Core 2, 2.0-GHz computer. However, larger designs can take considerably longer. The program reports degrees of freedom in the NANOVA table to maintain correspondence with the ANOVA, but *df*s are not involved in the computations.

The NANOVA tables in the examples highlight the most important practical advantage of the new technique. The structure built into the study comes through in a manner that is familiar to experimentalists. This clarity is especially valuable for designs with multiple factors. The results look like ANOVA tables, and the contributions of the factors are interpreted in much the same way.

Example 1: Repeated measures design. In Tables 1 and 3, the (fictional) data are clinical diagnoses made by medical students. The responses are unconstrained, in that no set of possible diseases from which to choose was provided. The students named the disease (here, signified by the letters a, b, c, or d) they attributed to a patient who had the designated set of symptoms. This is a 4 (symptom sets) × 5 (medical students) repeated measures design. Whereas in typical studies that examine diagnosis, the outcome measure is likely to be accuracy, here we explore the process question of how variation in the symptoms induces variation in the diagnoses. The data in Table 1 were constructed to show a *symptom* effect. That is, the medical students generally agreed on the diseases suggested by the symptom sets.

The overall number of potential matches is always the combination of the number of responses taken two at a time. There are 190 ($_{20}C_2$) potential matches generated by the 4 × 5 design; 41 occurred. Main effects are assessed by counting the matches within a row (column). The more matches that occur, the less effect that vari-

Table 1
Diagnoses (Artificial Data Constructed to Show Symptom Effect)

Medical Student	Symptom Set 1	Symptom Set 2	Symptom Set 3	Symptom Set 4
1	a	b	b	c
2	b	a	c	d
3	a	b	d	c
4	a	c	b	d
5	a	b	c	d

able has. To address the symptom effect, we examine the 30 ($5 \cdot {}_4C_2$) potential matches across rows, of which 1 occurred. Matches within columns are associated with medical students, which is the subjects source in this study. There are 40 ($4 \cdot {}_5C_2$) potential matches within columns; 15 occurred. Matches relevant to interaction are those not associated with either row or column. More specifically, interaction comparisons are those for which the cell indices for the pair both differ (e.g., 11 vs. 22, 21 vs. 12, 31 vs. 12, etc.). The more of those matches that occur, the greater the interaction. Therefore, the NP for an interaction is the proportion of matches, rather than the proportion of nonmatches.

The subject × treatment interaction is the usual *error term* in a repeated measures ANOVA. There are 120 of these pairs in the 4 × 5 design; matches occurred 25 times. Because we know the total number of potential and actual matches, those numbers used for the error term might also have been derived by subtraction (potential matches = 190 − 30 − 40; actual matches = 41 − 1 − 15). The obtained *N* ratio is formed by comparing the NPs as though they were mean squares in an ANOVA. As is shown in Table 2, the *N* ratio of 4.64 and the *p* value of .021 capture the symptom effect that I built into the data.

In contrast, the data in Table 3 were constructed to show a *subject* effect. Each medical student tends to give an idiosyncratic diagnosis without much regard for the symptoms.

This time, the *N* ratio for symptoms is smaller and nonsignificant, reflecting the pattern I built into the data.

Thus, the NANOVA analyses shown in Tables 2 and 4 detect what was built into the data. In Table 2, where subjects interpret the differential symptom information in much the same way, the *N* ratio for symptoms is "large." In Table 4, subjects tend to respond the same way regardless of the symptoms, and correspondingly, the *N* ratio for symptoms is "small." This difference is perhaps the strongest evidence for the promise of the proposed technique, in that the *N* ratio is responsive to effects in the data.

Example 2: Independent groups design. In the experiment reported in Table 5, 16 buyers saw one of four Web pages and then purchased product a, b, c, d, e, or nothing (after Fasolo, McClelland, & Lange, 2005). This is a one-way, independent groups design with four scores per cell. The null hypothesis is that what was purchased does not depend on which page the buyer saw; the alternative hypothesis is that the page does influence the purchase. Of course, with nominal data, hypotheses are never directional. The hypothesis is tested by comparing the NP for pages to a within-cells error term.

Over this set of 16 responses, there are 120 potential matches (${}_{16}C_2$). Twenty-three matches occurred. The potential and obtained matches are partitioned according to the factorial structure. Across Web pages, there are 24 potential matches ($4 \cdot {}_4C_2$), of which 3 occurred. Therefore, there are 21 nonmatches, yielding an NP of 21/24 = .875. The remainder of the potential and observed matches are allocated to the within-cells error term.[2] Here, *N* = .875/.313 = 2.80 (see Table 6).

Table 2
Nominal ANOVA, One-Way Repeated Measures Design (Artificial Data)

Source	*df*	Potential	NP	*N* Ratio	*p*
Medical student (M)	4	40	.625		
Symptom (S)	3	30	.967	4.64	.021
M × S	12	120	.208		

Note—Potential matches, 190; obtained matches, 41.

Table 3
Diagnoses (Artificial Data Constructed to Show Subject Effect)

Medical Student	Symptom Set 1	Symptom Set 2	Symptom Set 3	Symptom Set 4
1	a	a	b	a
2	b	b	b	c
3	d	c	c	c
4	d	d	d	d
5	a	b	a	a

Table 4
Nominal ANOVA, One-Way Repeated Measures Design (Artificial Data)

Source	*df*	Potential	NP	*N* Ratio	*p*
Medical student (M)	4	40	.850		
Symptom (S)	3	30	.400	2.82	.999
M × S	12	120	.142		

Note—Potential matches, 190; obtained matches, 41.

Table 5
Product Purchased (Artificial Data)

Page 1	Page 2	Page 3	Page 4
a	c	none	a
b	a	b	a
none	b	d	a
d	e	b	a

In the resampling analysis, the proportion of times the *N* ratio from permuted data exceeded the *N* ratio obtained from the original data (2.80) was .055, corresponding to a *p* value of .055 in an ANOVA. According to standard null hypothesis-testing logic, these results are consistent with the null hypothesis at the .05 level of significance; the page does not affect the purchase.

The usual way in which data suitable for a one-way, independent groups NANOVA are analyzed is with a chi-square test of independence. In Table 7, the data from Table 5 are displayed in a contingency table, wherein the entries are the number of people who bought a particular product after seeing a particular page. The null hypothesis for the chi-square test of independence is logically equivalent to that of a one-way NANOVA. Table 7 appears rather sparse, because this mode of presentation is poorly suited to the structure of the data set, in that some products were never chosen in response to particular pages. Consequently, the data may be inappropriate for a standard chi-square test of independence (a topic debated intensely 60 years ago; see, e.g., Lewis & Burke, 1949). Ignoring

that concern, I calculated the chi-square observed as 18.0. With 15 *df*s, the corresponding *p* value is .26, illustrating a power advantage for the NANOVA in this case.

Example 3: Two-factor independent groups design. Table 8 illustrates a two-factor (3 programs × 2 grades), independent groups design with four scores per cell. In this study, sixth-grade children who had earned either "A" or "C" grades in science last year were assigned to write a synopsis of a specific television program that they were asked to watch. The programs, all featuring scientists of a sort, were shown at 10 p.m. and were not normally seen by these young viewers. One week later, all of the students were asked to list three careers that they were considering. The children's first responses were examined to see whether the program assignment differentially influenced career consideration and whether this effect depended on the child's previous success in science. In this case, the responses are careers.

Each of the 276 ($_{24}C_2$) potential matches generated by the design is associated with exactly one of the sources. The results are shown in Table 9. The student's grade had the largest effect on career choice, but none of the three sources affected the choice significantly.

For the substantive sources, we count matches among corresponding pairs of responses. For the main effect of program, we count across rows. There are 15 matches among the 24 comparisons. For the main effect of grade, we count matches down columns; there is only 1. The null hypothesis for a NANOVA interaction is that the response distribution across the levels of one factor does not differ over the various levels of the other factor. The test checks for matches among cells that are not in the same row or column. Two such matches occurred.

Table 6
Nominal ANOVA, One-Way Independent Groups Design (Artificial Data)

Source	*df*	Potential	NP	*N* Ratio	*p*
Pages	3	24	.875	2.80	.055
Within	12	96	.313		

Note—Potential matches, 120; obtained matches, 23.

Table 7
Product Purchased × Page (Same Artificial Data As in Table 1)

Product	Page 1	Page 2	Page 3	Page 4
a	1	1	0	4
b	1	1	2	0
c	0	1	0	0
d	1	0	1	0
e	0	1	0	0
none	1	0	1	0

Table 8
Career Choices Among Sixth Graders (Artificial Data)

	Program		
Grade	ER	CSI	NUMB3RS
"A"	a a	b a	a b
	a a	a a	a a
"C"	c c	c d	d c
	c c	a c	c d

Table 9
Nominal ANOVA, Two-Way Independent Groups Design (Artificial Data)

Source	*df*	Potential	NP	*N* Ratio	*p*
Program (P)	2	24	.375	0.54	.999
Grade (G)	1	12	.917	1.33	.146
P × G	2	24	.083	0.12	1.000
Within	18	216	.690		

Note—Potential matches, 276; obtained matches, 87.

Preprocessing the Data

Carrying out a NANOVA requires decisions about whether each pair of responses matches. The simplistic interpretation of matching is that the responses must be identical. When constrained response options are offered, that determination is easy to accomplish and can be left to a computer. However, when free responding is permitted, the researcher may have to judge whether a pair of nonidentical responses ought to be counted as a match.

Declaring linguistic equivalents as matches seems innocuous. If two different words are true synonyms, the analyst may enter one of them both times. The use of different languages by respondents also justifies substitution. A more delicate judgment is required when one response is effectively a subset of another. For example, during a study examining the effectiveness of automobile advertising, the subject may be asked to name the kind of car he or she wants to buy. If the response is "Camry," is that a match with "Toyota"? "Camry" is certainly closer to "Toyota" than it is to "Ford." The researcher will have to make a decision. Such fuzzy matches were explored by Oden (1977), who asked people to judge, for example, the extent to which a bat is a bird. It may be feasible to devise a generalization of the NANOVA that incorporates degree of closeness, where 0 means *no match*, 1 means *identical*, and intermediate values capture the extent to which nonidentical responses overlap in meaning.

The Exchangeability Assumption

An important assumption underlying randomization tests is exchangeability. In the NANOVA application, all observations are held to be exchangeable. In particular, in repeated measures designs, one cannot restrict the exchanges on a subject by subject basis. If the responses made by each subject are exchanged separately, the number of matches for a subject does not change, and so the NP term for subjects does not change. The resampling routine allows all responses to be interchanged.

Explorations of Stability

The stability of the *p* values produced by the NANOVA was explored by systematically varying the number of randomizations used for an analysis. I conducted 1,000 separate analyses of the Table 5 data, successively employing 10,000, 50,000, and 100,000 randomizations for each

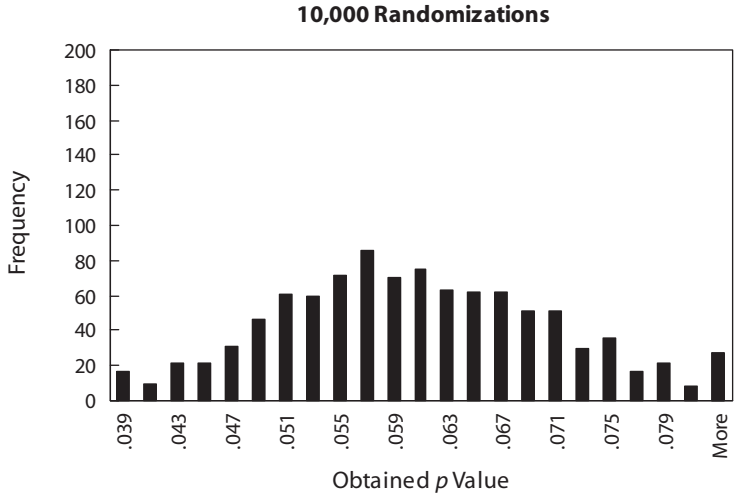

Figure 1. Distribution of 1,000 resampled *p* values using data from Table 5, with each *p* value based on 10,000 randomizations. $M = .060$, $SD = .010$.

analysis. The program reported individual *p* values to the third decimal place. Under all three conditions, the mean of the distribution was .060. As would be expected, the standard deviation decreased as the number of randomizations increased but differed only in the third decimal place. Given that all three conditions employed "large" numbers of randomizations, the similarity in these results is not surprising. However, dramatic differences appear when we examine the distributions graphically in the histograms displayed in Figures 1, 2, and 3. Using bin widths of .002, the modal *p* value of .059 was achieved 172 times when 100,000 randomizations were used to estimate each *p* value, but the modal value of .057 was achieved only 86 times when 10,000 randomizations were used. The more sharply peaked distribution seen in Figure 3 provides justification for using 100,000 randomizations as the default in the NANOVA computer program.

Explorations of Responsiveness

The responsiveness of the NANOVA analysis to the magnitude of effects in the data was examined by manipulating data structured in accord with Table 5. Each column had a unique modal response, capturing the idea that a page drew shoppers to a specific brand. In the first case, two of the four responses in each column were the same, with the other two different from them and from each other. That is, the first column contained a, a, b, c; the second column contained b, b, c, d; and so on. In the second case, the page had a stronger effect. Three of the four responses in each column were the same, with the fourth response different (a, a, a, c; b, b, b, d; etc.). In the third, most extreme case, all of the responses in the first three columns were the modal brand, whereas three of the four responses in the fourth column were the same and the last was different (a, a, a, a; . . . ; d, d, d, e). The *N* ratio for the

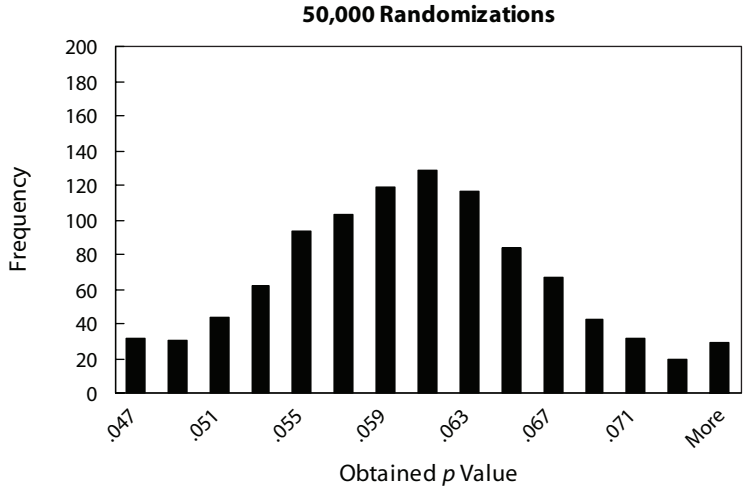

Figure 2. Distribution of 1,000 resampled *p* values using data from Table 5, with each *p* value based on 50,000 randomizations. $M = .060$, $SD = .007$.

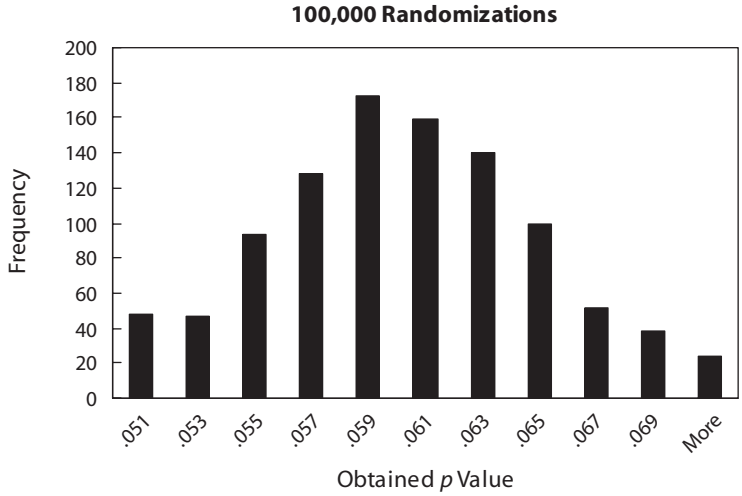

Figure 3. Distribution of 1,000 resampled *p* values using data from Table 5, with each *p* value based on 100,000 randomizations. $M = .060$, $SD = .005$.

first case, a baseline weak effect, was 2.91, with a *p* value of .219. The second case mimicked a strong effect, and the NANOVA duly reported an *N* ratio of 5.33 ($p < .001$). The third case was as strong an effect as possible (without generating an infinite *N* ratio), and the *N* ratio was 32.00 ($p < .001$). This demonstration shows that for a small data set, the NANOVA is not sensitive to weak effects but does detect stronger ones in accord with their magnitude.

Power Considerations

It seems obvious that nominal data afford less power than do numerical data, but how much less? To gain a foothold, I tried two exercises to examine how power in the NANOVA context compares with power in the ANOVA. When the data are homogeneous, one would expect power to increase with the size of the data set. To test this prediction, I increased the size of the data set in Table 5 by duplicating the responses repeatedly. With one duplication ($df = 3,28$; potential and obtained matches = 496, 108), the *N* ratio increased to 3.11. With two duplications ($df = 3,44$; potential and actual matches = 1,128, 255), *N* was 3.21. With three duplications ($df = 3, 60$; potential and actual matches = 2,016, 464), *N* was 3.26. The latter three *N* ratios all yielded a *p* value of $<.001$. In this exercise, the power gain from replication was achieved entirely from increases in the proportion of nonmatches in the within term, because with perfectly replicated responses the proportion of nonmatches associated with pages remains constant.

The other slant on power compared numerical responses with nominal responses. One would expect numerical data to afford more power. I computed an ordinary *F* ratio after replacing the letters in Table 5 with their ordinal positions in the alphabet and the "none" responses with zeros. If the study had been a real one, the numbers might represent time spent looking at the page or amount of money spent after viewing it. The comparison is via the *p* values. To my surprise, the *p* value was much higher (.564), suggesting less power for the numerical data. The resolution is that the advantage of numerical responses is that although they are inherently more sensitive to fine distinctions respondents may make, my substitution was linked to distinctions that had already been made using nominal responses. So, notwithstanding the inappropriateness of this mode of comparison, the result does suggest that the NANOVA is not an inherently weak test.

Power will be affected by the variety of responses the subject chooses. If only a few alternatives are exercised, there will be many matches. In some circumstances, the number of alternatives used will depend on the subject's verbal habits or base rates for particular response options. If we are studying actions, the situation itself may constrain the number of plausible options.

Power also depends on the statistical test employed. Pesarin and De Martini (2002) observed that the power of a permutation test depends on the population distribution. Customary power computations entail assumptions about that distribution. Little information about the distribution of freely emitted nominal responses is currently available. As such knowledge accrues, it may be feasible to make assumptions that support power analyses useful to an experimenter during the design of a study.

Discussion

An intriguing methodological possibility offered by the NANOVA procedure is parallel assessment, by which I mean studies that collect numerical and behavioral responses to the same stimuli. This is really an old idea, going back at least to a classic study in which LaPiere (1934) compared racial attitudes expressed by, and actions taken by, Southern innkeepers. One might examine how medical warnings varying in length and intensity affect the patient emotionally and, at the same time, see whether the same factors inspire behavioral change. Studies of training might look at how instructional innovations affect both knowledge and choice of action.

Nominal responding is the natural mode for expressing a choice, and the study of what underlies choices can perhaps best be accomplished with the analytic power of factorial designs.

The addition of the NANOVA procedure fills a hole in the experimenter's toolbox. The extant techniques for dealing with nominal responses are suited to the structure found in nature, but they do not take full advantage of the factorial structure that one can establish in the laboratory. Just as the ANOVA maintains its value even though it is, in a sense, subsumed by multiple regression (Cohen, 1968), so the NANOVA can be useful even though multinomial logit analysis may be applicable to the same data. The NANOVA analysis is conceptually simpler, makes fewer assumptions, and generates familiar inferences about the efficacy of the factors built into the study.

Nominal data are inherently less informative than quantitative data. Nominal data cannot be averaged, cannot be graphed, and do not convey information about the magnitude of differences. Because averaging is a meaningless operation, it is not clear how to deal with missing data or inequality of cell sizes. However, nominal responses are the natural mode for capturing actions, and a science of behavior ought to be able to make use of them. Despite Stevens's (1946, 1951) negative view, likely based on his prophysics, antipsychology biases (Matheson, 2006), nominal data can provide valuable information in experimental settings.

AUTHOR NOTE

This research was supported by the U.S. Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under Grant 2007-ST-061-000001. However, any opinions, findings, conclusions, or recommendations in this document are those of the author and do not necessarily reflect views of the U.S. Department of Homeland Security. I am grateful to Julia Pounds for suggesting the NANOVA acronym and to Richard John, Jorge Mendoza, and Rick Thomas for valuable comments regarding the manuscript. I also acknowledge the efforts of the reviewers toward improving the statistical aspects of the article. Correspondence regarding this article, including requests for reprints, should be sent to D. J. Weiss, 609 Colonial Circle, Fullerton, CA 92835 (e-mail: dweiss@calstatela.edu).

REFERENCES

AGRESTI, A. (1990). *Categorical data analysis*. New York: Wiley Interscience.

ANDERSON, N. H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, **58**, 305-316. doi:10.1037/h0042576

BOCK, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

CHEN, Z., & KUO, L. (2001). A note on the estimation of the multinomial logit model with random effects. *American Statistician*, **55**, 89-95.

COHEN, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, **70**, 426-443. doi:10.1037/h0026714

DYKE, G. V., & PATTERSON, H. D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics*, **8**, 1-12.

EDGINGTON, E. S., & ONGHENA, P. (2007). *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

FASOLO, B., MCCLELLAND, G. H., & LANGE, K. (2005). The effect of site design and interattribute correlations on interactive Web-based decisions. In C. P. Haugtvedt, K. Machleit, & R. Yalch (Eds.), *Online consumer psychology: Understanding and influencing behavior in the virtual world* (pp. 325-344). Mahwah, NJ: Erlbaum.

FIENBERG, S. E. (2000). Contingency tables and log-linear models: Basic results and new developments. *Journal of the American Statistical Association*, **95**, 643-647.

GINI, C. (1939). *Variabilità e concentrazione: Vol. 1 di. Memorie di metodologia statistica*. Milano: Giuffrè.

GOODMAN, L. A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, **13**, 33-61.

GRIZZLE, J. E. (1971). Multivariate logit analysis. *Biometrics*, **27**, 1057-1062.

HABERMAN, S. J. (1982). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, **77**, 568-580.

JAIN, D., VILCASSIM, N., & CHINTAGUNTA, P. (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business & Economic Statistics*, **12**, 317-328.

KEPPEL, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Prentice Hall.

LAPIERE, R. T. (1934). Attitudes and actions. *Social Forces*, **13**, 230-237.

LEWIS, D., & BURKE, C. J. (1949). The use and misuse of the chi-square test. *Psychological Bulletin*, **46**, 433-489. doi:10.1037/h0059088

LIGHT, R. J., & MARGOLIN, B. H. (1971). An analysis of variance for categorical data. *Journal of the American Statistical Association*, **66**, 534-544.

LORD, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, **8**, 750-751.

LUCE, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

MATHESON, G. (2006). Intervals and ratios: The invariantive transformations of Stanley Smith Stevens. *History of the Human Sciences*, **19**, 65-81. doi:10.1177/0952695106066542

MCFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zaremba (Ed.), *Frontiers in economics* (pp. 105-142). New York: Academic Press.

ODEN, G. C. (1977). Fuzziness in semantic memory: Choosing exemplars of subjective categories. *Memory & Cognition*, **5**, 198-204.

ONUKOGU, I. B. (1985). An analysis of variance of nominal data. *Biometrical Journal*, **27**, 375-383. doi:10.1002/bimj.4710270404

PESARIN, F., & DE MARTINI, D. (2002). On unbiasedness and power of permutation tests. *Metron*, **60**, 3-19.

RODGERS, J. L. (2000). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research*, **34**, 441-456.

SHAFFER, J. P. (1973). Defining and testing hypotheses in multidimensional contingency tables. *Psychological Bulletin*, **79**, 127-141. doi:10.1037/h0033865

STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, **103**, 677-680. doi:10.1126/science.103.2684.677

STEVENS, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-41). New York: Wiley.

WEISS, D. J. (2006). *Analysis of variance and functional measurement: A practical guide*. New York: Oxford University Press.

NOTES

1. The NANOVA computer program is available for free download from www.davidjweiss.com/NANOVA.htm. The Windows program handles designs with as many as four factors, including subjects or replicates. The Web site also features screen shots illustrating design specification and data entry.

2. The NANOVA computer program constructs the within-cells error term by pooling what would have been the *subjects* source along with all interactions involving *subjects* if a repeated measures design had been employed. Accordingly, for a given set of responses, the program generates the identical NANOVA table for an independent groups design as it does for a repeated measures design in which *subjects* are nested under all substantive factors. This correspondence is a property of the ANOVA as well (Weiss, 2006).

(Manuscript received August 12, 2008;
revision accepted for publication March 4, 2009.)