Oecologia (1998) 116:475–481

John R. Lockwood III

On the statistical analysis of multiple-choice feeding preference experiments

Received: 29 November 1997 / Accepted: 20 April 1998

Abstract A stopping rule for an experiment defines when (under what conditions) the experiment is terminated. I investigated the stopping rules used in numerous multiple–choice feeding-preference experiments and also examined a recently proposed method for analyzing the data arising from such experiments. All of the surveyed experiments imposed stopping rules which result in a random total food consumption. If an acceptable quantification of preference is relative consumption of different food types, then the proposed analysis will likely misstate the information about preference conveyed by the data. This is due to the fact that the method may confound differences in preferences among food types with differences in the total consumption across trials. I discuss this issue in detail and present an alternative procedure which is appropriate under all stopping regimes when preference is quantified through relative consumption. The procedure I suggest uses an index which is a multivariate generalization of the preference index suggested by Kogan and Goeden (Ann Entomol Soc 1970; 63: 1175–1180) and Kogan (Ann Entomol Soc 1972; 65: 675–683) and which is analogous to a selection index for discrete food units proposed by Manly (Biometrics 1974; 30: 281–294).

Key words Preference testing · Stopping rules · Relative consumption · Multivariate analysis · Parametric procedures

Introduction

Multiple–choice feeding preference experiments are an integral part of ecological research. For example, in plant–insect ecology, feeding preference tests have been used to investigate a wide array of issues: induced chemical resistance by plants following herbivore damage (Edwards and Wratten 1982; Edwards et al. 1985; Karban 1988; Zangerl 1990) the dependence of induced resistance on injury mechanisms (Lin et al. 1990), the time scales over which induced resistances are maintained (Wratten et al. 1984), insect preference for different food species (Barnes 1963; Grime et al. 1968; Dudgeon et al. 1990), insect preference for different food conditions (Lewis 1984), the relationship between relative species abundance and palatability (Landa and Rabinowitz 1983), the relationship between successional status and palatability (Cates and Orians 1975), the relationship between environmental conditions and insect preference (Coleman and Jones 1988; Jones and Coleman 1988b), and even the dependence of preference on choice test design (Risch 1985).

Complementing the widespread application of multiple–choice preference experiments is great diversity in both the experimental designs and the statistical techniques used in their analysis. Recent papers by Peterson and Renaud (1989), Roa (1992), and Manly (1993) describe general statistical procedures for the analysis of such experiments. I follow Roa (1992) and Manly (1993) by considering only the class of designs where a single consumer is given the choice between equal amounts of K different food types. In addition, Peterson and Renaud (1989), Roa (1992) and Manly (1993) are primarily concerned with the issue of controlling for autogenic changes in the foods (that is, changes due to factors other than animal feeding) which occur during the course of the experiment. Manly (1993) offers modifications to Roa's procedures which are appropriate when autogenic changes are significant. I, in turn, focus on the case where autogenic changes are believed to be negligible, although the issues I discuss are equally relevant in situations where autogenic changes are important.

This study proposes an alternative analysis to that suggested by Roa (1992) which is more appropriate for the types of stopping rules used by most investigators. I begin by defining three classes of stopping rules which

J.R. Lockwood III (✉)
Carnegie Mellon University, Department of Statistics, Baker Hall 132, Pittsburgh, PA 15213–3890, USA
e-mail: jlock@stat.cmu.edu

can be employed in multiple–choice preference tests. I then suppose that an experimenter wishes to quantify preference through the proportion of the total consumption within each feeding arena comprised by each of the K food types. I show that under the stopping regimes found in over 20 papers which have used preference tests, the method proposed by Roa (1992) may misrepresent the evidence for (or against) a preference. Finally, I offer modifications to a parametric procedure suggested by Roa (1992) which is appropriate under all stopping regimes and which is tailored for a relative consumption definition of preference.

Stopping rules

A design issue which has been overlooked in the literature concerning preference testing is the distinction between the various stopping rules that could be imposed in an experiment. By far the most common design is one in which the consumers are allowed to feed for a fixed amount of time, after which the trials are terminated and measurements are taken. I refer to this as the *fixed-time* design. Alternatively, a researcher could employ a design in which the feeding time is variable, but the trials are terminated after a fixed level of total consumption has been reached. I refer to this as the *fixed-consumption* design. Finally, various hybrid stopping rules have been used. For example, a rule which terminates a trial when 50% of one of the food types has been consumed, or when anywhere between 3–6 h of feeding time has elapsed, is neither a fixed-time nor a fixed-consumption design. I collectively refer to these as *mixed* designs.

Table 1 A survey of publications using preference tests along with their experimental stopping-rule classes

Study	Stopping-rule class
Barnes (1963)	Fixed time
Grime et al. (1968)	Fixed time
Edwards and Wratten (1982)	Fixed time
Landa and Rabinowitz (1983)	Fixed time
Lewis (1984)	Fixed time
Edwards et al. (1985)	Fixed time
Risch (1985)	Fixed time
Marquis and Braker (1987)	Fixed time
Coleman and Jones (1988)	Fixed time
Karban (1988)	Fixed time
Dudgeon et al. (1990)	Fixed time
Zangerl (1990)	Fixed time
Bingaman and Hart (1992)	Fixed time
Nylin and Janz (1996)	Fixed time
Cates and Orians (1975)	Mixed
de Boer and Hanson (1987)	Mixed
Jones and Coleman (1988a)	Mixed
Jones and Coleman (1988b)	Mixed
Peterson and Renaud (1989)	Mixed
Lin et al. (1990)	Mixed
Baur, Binder and Benz (1991)	Mixed
Wratten et al. (1984)	Not specified
Lin and Kogan (1990)	Not specified

Table 1 lists 23 papers which have used preference tests along with the design implemented in each. Note that none of the papers used a fixed-consumption design, presumably due to the difficulty of implementation. That is, a fixed consumption design would be tedious in that feeding arenas would have to be continually monitored until the fixed consumption was reached. The fact that the vast majority of experiments are not of the fixed-consumption design is significant because both fixed-time and mixed designs result in random total consumption; therefore, total consumption is likely to vary across arenas. The implications of variable total consumption, especially when preference is quantified through relative consumption, are discussed below.

Quantification of preference

The quantification of preference in a multiple–choice preference experiment is not straightforward. I assume that the experimenter wishes to quantify preference through relative or proportional consumption; for example, in dual-choice testing, one index which satisfies this criterion is the amount of one food type consumed divided by the total consumption (within each arena). I make this supposition for three reasons. First, as noted by Chesson (1983), defining preference in terms of the relative consumptions of different food types agrees well with our intuitive definition of "preference." Second, many of the "preference indices" that have been used in the analysis of dual-choice experiments are based on relative or proportional consumption, indicating that this method has practical appeal (see for example, Barnes 1963; Grime et al. 1968; Cates and Orians 1975; Wratten et al. 1984; Edwards et al. 1985; Lin and Kogan 1990; Lin et al. 1990). Finally, methods which are based solely on amounts convey little information about the *degree* of preference exhibited when the total consumption is random, as is the case in the vast majority of preference testing. For example, knowing that an insect consumed 1 g more of food A than food B may indicate that food A is preferred, but the degree of preference conveyed by this information is not clear without knowledge of the total amount consumed. However, the use of proportional consumptions is also somewhat troublesome because it ignores the total consumption, and one may wish to have large total consumptions carry more weight. For example, a consumption of 1 g of food A and 2 g of food B and a consumption of 10 g of food A and 20 g of food B carry the same information when viewed from the perspective of proportional consumption, but one may consider the second set of measurements as more informative. This difference would be captured if the actual amounts are used. Therefore, neither the total consumption nor the relative consumption viewpoint is flawless, and it is not the purpose of this paper to advocate either. Rather, the goal is to show how the distinction is important, especially when the total consumption is likely to vary across arenas.

Recently proposed methods

Roa (1992) suggests a common multivariate analysis technique for use in the fixed-time K–choice test design. A multivariate analysis is more appropriate than standard ANOVA techniques because, as noted by Roa, the consumptions within an arena may violate ANOVA independence assumptions. For example, data arising from preference testing are ostensibly amenable to randomized complete-block ANOVA where each arena is a block and each food type a treatment. However, randomized complete-block ANOVA assumes that within-block observation errors are independent (Neter et al. 1996), a condition which is unlikely to be met when the treatment observations are made simultaneously. A multivariate analysis, on the other hand, explicitly accounts for these within-block correlations.

The context for the proposed methods is the following. Each arena initially has equal amounts of K different food types. In each arena, the consumer is allowed to feed for a fixed time, after which the remaining amounts of the foods are measured. The datum arising from each arena is a K-dimensional vector of consumptions, each component representing the amount of food consumed for a particular food type. To formalize notation, let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iK})'$ be a random vector of consumptions for the *ith* arena, $i = 1, \ldots, n$ such that the *jth* component of the *ith* vector gives the amount of food j consumed in arena i. Let the observed vectors be $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})'$ (throughout the paper, upper-case letters refer to random variables and lower-case letters refer to realizations of these random variables). Roa assumes that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and identically distributed multivariate normal random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. He then proposes the use of Hotelling's T^2 (Morrison 1990), the multivariate generalization of a t–test, to test the null hypothesis of no preference. Letting $\bar{x}_{..} = \frac{1}{nK}\sum_{i=1}^{n}\sum_{j=1}^{K} x_{ij}$, Roa suggests testing the null hypothesis

$$H_0 : \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_K \end{bmatrix} = \begin{bmatrix} \bar{x}_{..} \\ \vdots \\ \bar{x}_{..} \end{bmatrix} \tag{1}$$

As noted by Manly (1993), this form of the hypothesis statement is not strictly correct because it treats the sample estimate $\bar{x}_{..}$ of the population grand mean as a known quantity. Manly corrects this oversight by suggesting a modification to this test which does not assume that this quantity is known; the modification tests the null hypothesis

$$H_0 : \begin{bmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_K - \mu_{K-1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2}$$

by calculating Hotelling's T^2 for the differenced data $w_{ij} = x_{i,j+1} - x_{ij}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, K - 1$.

In addition, Roa's data vector is not ideal when one wishes to quantify preference through relative consumption and when the total consumption is a random quantity. Because the vector components are amounts and the total consumption is likely to vary across arenas, variability in the vector comes from two sources (ignoring experimental error): variability in the total consumption and variability in the proportional consumption. The latter quantity – the quantity of interest – is confounded with the variability in the random total consumption. Therefore, it is not possible to assess accurately the evidence for preference using the suggested analysis.

A somewhat contrived example reveals the shortcoming of these methods when applied in a design which results in a random total consumption. Consider the fixed-time design when $K = 2$, and imagine a certain species of insect that, when given the choice between equal amounts of food 1 and food 2, always consumes exactly twice as much food 1 as food 2. Clearly, this defines a preference for food 1. Suppose that we perform a fixed-time choice experiment with n replicates, and collect the vectors $\mathbf{x}_i = (x_{i1}, x_{i2})'$ for $i = 1 \ldots n$. It is highly likely that the total amount consumed in each arena will not be the same across arenas; therefore, knowing that the insects always consume in a 2:1 ratio, the vectors $\mathbf{x}_i$ will be distinct. We can therefore carry out either of the hypothesis tests (1) or (2) and produce a P-value, giving us the choice between acceptance and rejection of the null hypothesis. However, the interpretation of this P-value is suspect. The quantity of interest, relative consumption, is the same for all replicates and hence has zero sample variance. The P-value of a sensible test for preference in this case should therefore be zero (i.e., based on the test, one has no choice but to conclude that a preference exists). However, the P-values calculated from either Roa's or Manly's test, as long as not all of the vectors $\mathbf{x}_i$ are the same, will be non-zero. That is, if one's level of significance is low enough, one will fail to reject the null hypothesis of no preference despite the fact that a preference clearly exists.

Although the example is extreme, its implications hold in general for the test described by Roa (or the modification proposed by Manly). The P-value is inflated by the random total consumption, which is purely noise to one who is interested in basing inferences on proportional consumption. A test which is based only on the amounts consumed necessarily misinterprets this noise as substantive information; the degree to which the results, viewed in terms of relative consumption, are affected depends on how severely the total consumptions vary. Therefore, when random total consumption is expected and when one is concerned with relative consumption, tests based on a vector of absolute consumptions are not ideal.

In the remainder of the paper, I offer an analog to the above procedure which is appropriate when the experimenter wishes to quantify preference through proportional consumption, and which is appropriate under all stopping regimes. I give an example of how to implement the technique using a hypothetical data set, and I compare the results of this technique to those produced by Manly's modification to Roa's analysis.

Alternative analysis

When preference is to be quantified through proportional consumption and when the total consumptions are likely to vary across arenas, the K-dimensional vectors of consumptions must be standardized to reflect proportional consumptions. This can be accomplished by dividing each vector by the sum of components within the vector. That is, let $\mathbf{Z}_i = \mathbf{X}_i / \sum_{j=1}^{K} X_{ij}$ for $i = 1 \ldots, n$. Hence, the jth component of the ith new vector is the proportion of all of the food consumed in arena i which is of type j. Note that this interpretation holds regardless of whether total consumption is a fixed or random quantity and so is consistent across stopping rules. Despite the fact that the vectors cannot be multinormal (each is constrained to lie on a simplex; i.e., $\sum_{j=1}^{K} Z_{ij} = 1$), we can still use Hotelling's T^2 to test the null hypothesis of no preference by appealing to the multivariate central-limit theorem. If we assume that these n vectors are independent and identically distributed with mean vector $\boldsymbol{\pi}$ and finite covariance matrix $\boldsymbol{\Sigma}$, then the sample mean vector $\overline{\mathbf{Z}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i$ will be normally distributed with mean vector $\boldsymbol{\pi}$ and covariance matrix $\boldsymbol{\Sigma}/n$ as n goes to infinity (Schervish 1995). The approximate normality of the sample mean vector for finite samples makes the test based on Hotelling's T^2 a reasonable tool even if the data are not normally distributed.

When the vectors are within-arena proportions, a linear constraint exists because the components must sum to unity. Although this does not affect the convergence to normality (Schervish 1995), the limiting normal distribution will be degenerate in the sense that one of the components can be expressed as one minus the sum of the remaining components. Furthermore, the linear constraint will cause the sample covariance matrix to be singular. As Hotelling's T^2 requires an invertible covariance matrix, it is necessary to reduce the dimensionality of the vector by eliminating one of the components (and its corresponding row and column of the covariance matrix). The value of T^2 is invariant to which component is eliminated.

Without loss of generality, suppose that the *Kth* component of each vector is eliminated; therefore, the relevant data are the random vectors $\mathbf{Y}_i = (Z_{i1}, \ldots, Z_{i,K-1})'$ for $i = 1, \ldots, n$ $(n > K)$. The assumptions about the $\mathbf{Z}_i$ imply that these vectors are independent and identically distributed from a multivariate distribution with mean vector $\boldsymbol{\pi}_Y = (\pi_1, \ldots, \pi_{K-1})$ and covariance matrix $\boldsymbol{\Sigma}_\mathrm{Y}$, where $\boldsymbol{\pi}_Y$ consists of the first $K-1$ components of $\boldsymbol{\pi}$ and $\boldsymbol{\Sigma}_\mathrm{Y}$ the upper-left $(K-1) \times (K-1)$ corner of $\boldsymbol{\Sigma}$. Let $\overline{\mathbf{Y}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i$ be the sample mean vector and let $\mathbf{S}_\mathbf{Y} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{Y}_i - \overline{\mathbf{Y}}_n)(\mathbf{Y}_i - \overline{\mathbf{Y}}_n)'$ be the sample covariance matrix. We apply Hotelling's T^2 to test the null hypothesis of no preference; i.e.,

$$H_0 : \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_{K-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{K} \\ \vdots \\ \frac{1}{K} \end{bmatrix} \tag{3}$$

To test this hypothesis, let $\boldsymbol{\pi}_\mathbf{0} = (1/K, \ldots, 1/K)'$ and form the statistic

$$T^2 = n(\overline{\mathbf{Y}}_n - \boldsymbol{\pi}_\mathbf{0})'\mathbf{S}_\mathbf{Y}^{-1}(\overline{\mathbf{Y}}_n - \boldsymbol{\pi}_\mathbf{0}) \tag{4}$$

Under H_0, the quantity $\frac{n-K+1}{(K-1)(n-1)}T^2$ is distributed as an F random variable on $K-1$ and $n-K+1$ degrees of freedom. Hence, the natural decision rule is to reject H_0 if and only if

$$T^2 > \frac{(K-1)(n-1)}{n-K+1} F_{\alpha;K-1,n-K+1} \tag{5}$$

where $F_{\alpha;K-1,n-K+1}$ is the $(1-\alpha)$ quantile of the F-distribution on $K-1$ and $n-K+1$ degrees of freedom (Morrison 1990).

If this null hypothesis is rejected, it will typically be of interest to determine which of the components deviated most significantly from the null hypothesis. In this case, since no matrix inversion is required, we can use the full K-dimensional mean vector of proportions $\bar{\mathbf{z}}$ and $K \times K$ sample covariance matrix of the proportions $\mathbf{s_z}$ to build confidence intervals for quantities of interest. The general form of a symmetric $(1-\alpha)\%$ confidence interval for a linear combination $\mathbf{a}'\boldsymbol{\pi}$ for $\mathbf{a}$ any $K \times 1$ vector of constants is

$$(\mathbf{a}'\bar{\mathbf{z}}_\mathbf{n} - Q_\alpha, \mathbf{a}'\bar{\mathbf{z}}_\mathbf{n} + Q_\alpha) \tag{6}$$

If more than one interval is to be constructed, which is typically the case in a multivariate setting, the choice of Q_α must reflect a correction for performing multiple comparisons. A common choice is

$$Q_\alpha = \left[\frac{(n-1)(K-1)}{n(n-K+1)}(\mathbf{a}'\mathbf{s_Z}\mathbf{a})(F_{\alpha;\,K-1;\,n-K+1})\right]^{\frac{1}{2}} \tag{7}$$

The advantage of using this particular value is that one can form confidence intervals for an arbitrary number of linear combinations of the parameters and all will *simultaneously* have the confidence coefficient $(1-\alpha)$. Other multiple-comparison procedures are detailed in Morrison (1990).

Remarks

The alternative analysis is analogous to the procedure suggested by Manly (1974) for use in selection experiments where discrete food units are chosen. In this case, we can imagine that the food presented to the consumer consists of a large number of very small discrete units. It is also of interest to note that when $K = 2$, the test reduces to performing a traditional two-sided t-test on the ratio of the amount of one of the food types consumed to the total consumption within each arena, a procedure that has been used extensively (e.g., Edwards et al. 1985; Lin and Kogan 1990; Lin et al. 1990). Thus, the suggested analysis extends a popular technique used in dual-choice testing to experiments of arbitrary dimension.

In addition, while the multivariate central-limit theorem motivates the test as asymptotically correct, small sample sizes should not hinder its validity. The T^2 test, like the univariate t-test, is quite robust to violations of normality assumptions and is tailored for use when sample sizes are small. As the number of trials n goes to infinity, the T^2-test is equivalent to a particular χ^2-test appropriate when the covariance matrix is known (Morrison 1990). However, using T^2 instead of the χ^2 results in more conservative confidence intervals and hypothesis tests, especially when n is small. This can protect inferences against errors introduced by both the normal approximation of the distribution of the sample mean vector and the use of an estimated covariance matrix.

Another issue regarding the validity of the test is the assumption that the proportional consumption data can be adequately modeled as independent and identically distributed (iid). This is the only assumption necessary for the test to be asymptotically correct as stated; however, it is something which should be examined carefully in practice due to the complex nature of animal feeding behavior. One advantage of the assumption used in this procedure compared to that used in Roa (1992) is that it allows animal feeding scale to vary across arenas. That is, while Roa's procedure requires that the absolute consumption vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid, the analysis described here allows the consumptions to be distributed as $c_1\mathbf{X}_1^*, \ldots, c_n\mathbf{X}_n^*$ for arbitrary positive scalars $c_1, \ldots, c_n$ and iid random vectors $\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*$. The additional flexibility provided by allowing arena distributions to differ by scale factors may help to control for inherent differences in consumer feeding habits.

Finally, there are at least two non-parametric procedures which may be appropriate for the analysis of these data. Both replace the observed consumption vectors with vectors of within-arena consumption ranks; note that this does not distinguish between absolute and proportional consumption. While the procedures are conceptually straightforward, they are somewhat notationally cumbersome and are not detailed here. The first is Friedman's test, described in detail in Conover (1980). Although the test is meant for use in the analysis of data arising from a randomized complete-block design, tests based on ranks are generally robust to violations of distributional assumptions (Hora and Iman 1988) and so the test may be appropriate. The other test is given by Anderson (1959) and is designed specifically for detecting preferences for certain treatments. However, the test cannot be used when there are ties in the rankings within any arena.

Example

Table 2 contains hypothetical consumption data from a fixed-time preference test with $N = 20$, $K = 3$, and where each arena started with 50 mass units of each of the three food types. The arenas are listed in order of increasing total consumption. It seems that some degree of preference was expressed for food type 3, which had the highest proportional consumption in 15 of the 20 arenas. Proceeding with Manly's modification to Roa's analysis, the sample mean vector and sample covariance matrix for the differenced data $w_{ij} = x_{i,j+1} - x_{ij}$ for $i = 1, \ldots, 20$ and $j = 1, 2$ are

$$\bar{\mathbf{w}} = \begin{bmatrix} -0.115 \\ 3.02 \end{bmatrix} \qquad \mathbf{s_w} = \begin{bmatrix} 46.39187 & -33.06705 \\ -33.06705 & 55.32063 \end{bmatrix} \tag{8}$$

These lead to $T^2 = 5.44$, and the 0.95 quantile of the F-distribution with 2 and 18 degrees of freedom is $F_{\text{crit}} = 3.55$. Taking $\alpha = 0.05$, we reject the null hypothesis of no preference if and only if $T^2 > \frac{38}{18}F_{\text{crit}} = 7.49$, which does not hold. Hence, an analysis based on absolute consumption fails to reject the null hypothesis. The P-value of the test is 0.103.

The analysis based on the vectors of proportional consumptions proceeds as follows. The mean vector and covariance matrix for the normalized data $\mathbf{z}_i = \mathbf{x}_i / \sum_{j=1}^{3} x_{ij}$ for $i = 1, \ldots, 20$ are

$$\bar{\mathbf{z}} = \begin{bmatrix} 0.3050322 \\ 0.3107815 \\ 0.3841863 \end{bmatrix}$$

Table 2 Absolute consumption data for example

Arena	Type 1	Type 2	Type 3	Total
1	1.0	0.9	1.5	3.4
2	3.5	4.5	4.8	12.8
3	4.1	5.1	7.6	16.8
4	5.6	7.7	6.8	20.1
5	5.3	6.2	9.6	21.1
6	5.5	6.3	10.2	22.0
7	8.1	7.2	8.5	23.8
8	9.9	8.0	14.6	32.5
9	11.6	7.5	14.7	33.8
10	14.4	26.5	13.7	54.6
11	23.2	24.2	32.1	79.5
12	26.1	23.7	34.0	83.8
13	27.1	22.1	34.9	84.1
14	38.4	20.1	30.8	89.3
15	30.0	24.6	35.5	90.1
16	32.3	24.0	34.0	90.3
17	30.2	33.3	35.0	98.5
18	29.7	39.3	30.1	99.1
19	24.0	35.8	41.4	101.2
20	39.2	39.9	27.5	106.6

$$\mathbf{s_z} = \begin{bmatrix} 0.0023224006 & -0.001330092 & -0.0009923086 \\ -0.0013300919 & 0.004235507 & -0.0029054152 \\ -0.0009923086 & -0.002905415 & 0.0038977238 \end{bmatrix} \quad (9)$$

(Despite the insignificance of most of the digits, it is advisable to avoid rounding until the end of the calculation, especially when inverting matrices.) Note that the mean vector sums to unity and that the covariance matrix has only rank 2. We redress this problem by eliminating the third component of the mean vector and the third row and column of the covariance matrix, which is equivalent to eliminating the third food type from the proportional data matrix and calculating the mean vector and covariance matrix of the remaining two types. To test the null hypothesis that all three proportions are equal to one-third, we calculate Hotelling's T^2 using the reduced mean vector and covariance matrix (Eq.4) to produce $T^2 = 15.55$. As before, $F_{\text{crit}} = F_{0.05,2,18} = 3.55$. We reject the null hypothesis if and only if $T^2 > \frac{38}{18} F_{\text{crit}} = 7.49$, which holds. The P-value of this test is 0.0046. Thus, at level 0.05, there is a statistically significant difference in proportional consumption of the three food types.

After rejection of the null hypothesis, it is common to perform all of the pairwise comparisons to decide which food types were preferred differently. That is, letting $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)'$ be the underlying proportions, we can test whether $\pi_i - \pi_j = 0$ for $i < j \leq 3$. We can carry out simultaneous level $\alpha = 0.05$ tests by building three 95 confidence intervals for the differences $\pi_i - \pi_j$ and checking which intervals contain zero. Using Eq. 6 and 7, the three intervals are

$$\begin{aligned} \pi_1 - \pi_2 &: \quad (-0.065, \quad 0.053) \\ \pi_1 - \pi_3 &: \quad (-0.135, \quad -0.024) \\ \pi_2 - \pi_3 &: \quad (-0.146, \quad -0.001) \end{aligned}$$

Based on these intervals, we are 95 confident that food 3 is preferred over food 1 and marginally 95 confident that food 3 is preferred over food 2 in terms of proportional consumption. There does not appear to be a significant difference in preference for foods 1 and 2.

Summary

If relative consumption of different food types is an acceptable measure of preference and if total consumption varies across feeding arenas, tests based on vectors of absolute consumptions are not ideal. In particular, Hotelling's T^2-test performed with these vectors confounds variability in total consumption with variability in relative consumption, leading to less efficient estimation than is possible if inferences are based on vectors of proportional consumptions. Since most preference test designs result in random total consumption and many investigators quantify preference in terms of relative consumption, it is of considerable practical interest to have a statistical technique which is not affected by the noise introduced by random total consumption. The analysis proposed in this paper is one such technique, independent of the stopping rule used in the experiment.

Acknowledgements I thank William F. Morris, George MacKenzie, and Nora Underwood for their input on the project idea. I also thank Kevin R. Gross, William F. Morris, Mark J. Schervish, Michael J. Daniels, Dick Green and one anonymous reviewer for their helpful comments on the manuscript.

References

Anderson RL (1959) Use of contingency tables in the analysis of consumer preference studies. Biometrics 15: 582–590

Barnes OL (1963) Food-plant tests with the differential grasshopper. J Econ Entomol 56: 396–399

Baur R, Binder S, Benz G (1991) Nonglandular leaf trichomes as short-term inducible defense of the grey alder, *Alnus incana* (L.), against the crysomelid beetle, *Agelastica alni* L. Oecologia 87: 219–226

Bingaman BR, Hart ER (1992) Feeding and oviposition preferences of adult cottonwood leaf beetles (Coleoptera: Chrysomelidae) among *Populus* clones and leaf age classes. Environ Entomol 21: 508–517

Boer G de, Hanson FE (1987) Feeding responses to solanaceous allelochemicals by larvae of the tobacco hornworm, *Manduca sexta*. Entomol Exp Appl 45: 123–131

Cates RG, Orians GH (1975) Successional status and the palatability of plants to generalized herbivores. Ecology 56: 410–418

Chesson J (1983) The estimation and analysis of preference and its relationship to foraging models. Ecology 64: 1297–1304

Coleman JS, Jones CG (1988) Plant stress and insect performance: cottonwood, ozone and a leaf beetle. Oecologia 76: 57–61

Conover WJ (1980) Practical nonparametric statistics, 2nd edn. Wiley, New York

Dudgeon D, Ma HHT, Lam PKS (1990) Differential palatability of leaf litter to four sympatric isopods in a Hong Kong forest. Oecologia 84: 398–403

Edwards PJ, Wratten SD (1982) Wound-induced changes in palatability in birch (*Betula pubescens*: ehrh. ssp. *pubescens*). Am Nat 120: 816–818

Edwards PJ, Wratten SD, Cox H (1985) Wound-induced changes in the acceptability of tomato to larvae of *Spodoptera littoralis*: a laboratory bioassay. Ecol Entomol 10: 155–158

Grime JP, MacPherson-Stewart SF, Dearman RS (1968) An investigation of leaf palatability using the snail *Cepaea nemoralis* L. J Ecol 56: 405–420

Hora SC, Iman RL (1988) Asymptotic relative efficiencies of the rank- transformation procedure in randomized complete block designs. J Am Stat Assoc 83: 462–470

Jones CG, Coleman JS (1988a) Leaf disc size and insect feeding preference: implications for assays and studies on induction of plant defense. Entomol Exp Appl 47: 167–172

Jones CG, Coleman JS (1988b) Plant stress and insect behavior: cottonwood, ozone and the feeding and oviposition preference of a beetle. Oecologia 76: 51–56

Karban R (1988) Resistance to beet armyworms (*Spodoptera exigua*) induced by exposure to spider mites (*Tetranychus turkestani*) in cotton. Am Mid Nat 119: 77–82

Landa K, Rabinowitz D (1983) Relative preference of *Arphia sulphurea* (Orthoptera: Acrididae) for sparse and common prairie grasses. Ecology 64: 392–395

Lewis AC (1984) Plant quality and grasshopper feeding: effects of sunflower condition on preference and performance in *Melanoplus differentialis*. Ecology 65: 836–843

Lin H, Kogan M (1990) Influence of induced resistance in soybean on the development and nutrition of the soybean looper and the Mexican bean beetle. Entomol Exp Appl 55: 131–138

Lin H, Kogan M, Fischer D (1990) Induced resistance in soybean to the Mexican bean beetle (Coleoptera: Coccinellidae): comparisons of inducing factors. Environ Entomol 19: 1852–1857

Manly BFJ (1974) A model for certain types of selection experiments. Biometrics 30: 281–294

Manly BFJ (1993) Comments on design and analysis of multiple-choice feeding-preference experiments. Oecologia 93: 149–152

Marquis RJ, Braker HE (1987) Influence of method of presentation on results of plant-host preference tests with two species of grasshopper. Entomol Exp Appl 44: 59–63

Morrison DF (1990) Multivariate statistical methods, 3rd edn. McGraw-Hill, New York

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models, 4th edn. Irwin, Chicago

Nylin S, Janz N (1996) Host plant preferences in the comma butterfly (*Polygonia c-album*): Do parents and offspring agree? Ecoscience 3: 285–289

Peterson CH, Renaud PE (1989) Analysis of feeding preference experiments. Oecologia 80: 82–86

Risch SJ (1985) Effects of induced chemical changes on interpretation of feeding preference tests. Entomol Exp Appl 39: 81–84

Roa R (1992) Design and analysis of multiple-choice feeding-preference experiments. Oecologia 89: 509–515

Schervish MJ (1995) Theory of statistics, 2nd edn. Springer, Berlin Heidelberg New York

Wratten SD, Edwards PJ, Dunn I (1984) Wound-induced changes in the palatability of *Betula pubescens* and *B. pendula*. Oecologia 61: 372–375

Zangerl AR (1990) Furanocoumarin induction in wild parsnip: evidence for an induced defense against herbivores. Ecology 71: 1926–1932