



## Archaeofaunal Relative Abundance, Sample Size, and Statistical Methods

Michael D. Cannon\*

Department of Anthropology, Box 353100, University of Washington, Seattle, WA 98195-3100, U.S.A.

(Received 30 November 1999, revised manuscript accepted 29 February 2000)

Tests for correlations between sample size and taxonomic relative abundance are commonly used in zooarchaeological analyses to determine whether observed trends in relative abundance might simply be the result of sampling error. Monte Carlo simulations designed to evaluate the utility of this method indicate that it is inadequate as a means of detecting errors resulting from the incorporation of small samples in an analysis. Among simulated sets of sample assemblages, significant correlations between sample size and relative abundance are distributed randomly with respect to whether or not Type II errors concerning trends in relative abundance are present, and are underrepresented in cases in which Type I errors are present. This is because the conditions that are most conducive to correlations between sample size and relative abundance are quite different from the conditions that will lead to erroneous conclusions about the presence or absence of a trend in relative abundance. An alternative chi-square-based statistical method for evaluating trends in sample relative abundance, Cochran's test of linear trend, results in lower rates of both Type I and Type II errors than is the case with previously used correlation-based methods.

© 2001 Academic Press

**Keywords:** ZOOARCHAEOLOGY, QUANTITATIVE METHODS, SAMPLE SIZE, MONTE CARLO SIMULATION.

Quantitative analyses of temporal and spatial trends in taxonomic relative abundance have become common in zooarchaeological treatments of issues relating to subsistence and palaeoecology (e.g. Szuter & Bayham, 1989; Grayson, 1991; Szuter, 1991; Broughton, 1994a, b, 1997, 1999; Janetski, 1997; Quirt-Booth & Cruz-Uribe, 1997). Admirably, many who have performed this sort of analysis have recognized that it is important to consider the sizes of the samples that are used and have attempted to show that observed trends in relative abundance are not simply artefacts of sampling but are truly present in the populations being sampled. The most common method by which this has been done involves evaluating whether a significant correlation exists between the sizes of samples and the relative abundance values observed in those samples. This approach (hereafter, the "sample size correlation method") was introduced by Grayson (1984: 116–130; 1989), and has recently seen increasing use (e.g., Broughton, 1994a, b, 1997, 1999; Janetski, 1997; Grayson & Cannon, 1999).

The use of this method is based on two related assumptions: (1) that the presence of a significant correlation between sample size and relative abundance suggests that sample size may be responsible for an observed trend in relative abundance; and (2) that

the absence of such a correlation indicates that variability in observed relative abundance values is independent of variability in sample size. Both of these assumptions, however, are problematic.

The second assumption provides an insufficient basis for concluding that sample sizes are adequate in an analysis of trends in relative abundance. There may be no correlation between sample size and relative abundance among a set of sample assemblages, but samples may still be too small to allow much confidence in the inference that relative abundance truly varies among the populations that are being sampled (Cannon, 2000). The first assumption rests on an incomplete understanding of the relationship between the relative abundance values observed in a set of samples and the sizes of those samples (Grayson, 1984: 126–127). Exploring this relationship in greater detail is the purpose of this paper.

Using sets of sample assemblages drawn from hypothetical sets of population assemblages through the technique of Monte Carlo simulation, I show that correlations between sample size and relative abundance are not useful as an indicator of sample size-related problems in analyses of trends in relative abundance. Among these simulated sets of sample assemblages, significant correlations between sample size and relative abundance are distributed randomly with respect to whether or not Type II errors concerning trends in relative abundance are present, and are

\*E-mail: [mcannon@u.washington.edu](mailto:mcannon@u.washington.edu); Tel.: (+1) 206-543-7712.

underrepresented in cases in which Type I errors are present. This is because the circumstances that are most conducive to correlations between sample size and relative abundance are quite different from the circumstances that are most likely to lead to erroneous conclusions about the presence or absence of a trend in relative abundance. I conclude by discussing an alternative chi-square-based statistical method for evaluating trends in sample relative abundance that can be shown to result in lower rates of both Type I and Type II errors than occur with previously used correlation-based methods.

### Evaluation of the Sample Size Correlation Method through Monte Carlo Simulation

The most useful way in which to think about the problem at hand is in terms of hypothesis testing. In analyses of trends in archaeofaunal relative abundance there is often an explicit hypothesis that is being evaluated; for instance, that the relative abundance of large mammals shows a decline through time, perhaps ultimately due to increasing rates of harvest by humans (e.g., Szuter & Bayham, 1989; Broughton, 1994a, b, 1999; Janetski, 1997; Cannon, 2000; see also Grayson & Cannon, 1999). The assemblages that are used to test such an hypothesis, of course, should be treated as samples drawn from one or more population assemblages (e.g., Leonard, 1997), and there are two kinds of errors that are possible in this sort of test: (1) concluding that a trend in relative abundance is present among population assemblages when in fact one is not (Type I error); and (2) concluding that no trend in relative abundance is present among population assemblages when in fact one is (Type II error).

The sample size correlation method has been employed in analyses of trends in relative abundance primarily as a means of determining whether a Type I error might be present; that is, it has been used to evaluate whether an observed trend in relative abundance might be nothing more than an artefact of sampling. In this paper, I test the utility of the method in this regard, as well as its utility for detecting Type II errors, through the technique of Monte Carlo simulation.

#### *Simulation methodology*

Monte Carlo simulation is used here to generate a large number of sets of sample assemblages drawn randomly from several sets of population assemblages. These sets of simulated sample assemblages can then be used to explore the conditions under which trends in relative abundance are more or less likely to be correctly identified, and the conditions under which correlations between sample size and relative abundance are more or less likely to occur.

Figure 1 presents a schematic representation of the way in which the simulations discussed here were

carried out. Five distinct population assemblages were used. These were ordered along an arbitrary scale of 1 through 5, which might be thought of as a time scale with 1 being the oldest assemblage and 5 the youngest. Two variables were manipulated: the proportion of the taxon of interest in each population assemblage\* and the size of the sample assemblage that was drawn from each population assemblage. Population size was 200 specimens for each of the five population assemblages.

Four sets of assemblage population proportions, or "Population Sets", were employed (Table 1). The first two (Population Sets 1 and 2) both display declining linear trends in relative abundance, while the second two (Population Sets 3 and 4) both consist of assemblages that do not vary at all in relative abundance. In addition, within both the "declining trend" and the "no trend" pairs of Population Sets, the assemblages in the first Population Set have higher relative abundances than do the assemblages in the second Population Set. In the analysis that follows, the "no trend" Population Sets will be used to assess the utility of the sample size correlation method for detecting Type I errors in analyses of trends in relative abundance, and the "declining trend" Population Sets will be used to assess its utility for detecting Type II errors.

To generate sets of sample assemblages, 20 "Sample Size Sets" were created (Table 2). For each Sample Size Set, five sample sizes were drawn randomly from an array containing the values 10, 50, 100, 150, and 190 (reflecting the sampling percentages 5, 25, 50, 75, and 95%, respectively), and these sample sizes were assigned to the positions 1 through 5 in the order in which they were drawn. To ensure that adequate variability in sample size would be present in each Sample Size Set, the method used to draw sample sizes would not allow any single value to be chosen more than twice.

Each of these 20 Sample Size Sets was then combined with each of the four Population Sets to comprise a simulation run, for a total of 80 runs. Each run entailed drawing a random sample of the specified size from each of the five population assemblages 100 times. Individual samples were drawn without replacement, although the same "specimens" might be drawn multiple times in subsequent samples within a run. The 100 sample cases within each of the 80 simulation runs thus each consist of five sample assemblages ordered along the arbitrary 1–5 "time" scale. Each of the five sample assemblages within each sample case possesses its own sample size, which is the same for all of the 100 sample assemblages of a given scale position in a given run. Each of the five sample assemblages in each

\*The results discussed here apply equally to the proportions of single taxa in assemblages composed of many taxa and to measures of the abundance of one taxon relative to another one like the "Artiodactyl Index" (e.g. Szuter & Bayham, 1989). For the latter, however, the relevant populations and samples consist only of the numbers of specimens of the two taxa that are incorporated into the index, rather than the total number of specimens of all of the taxa that might be present.

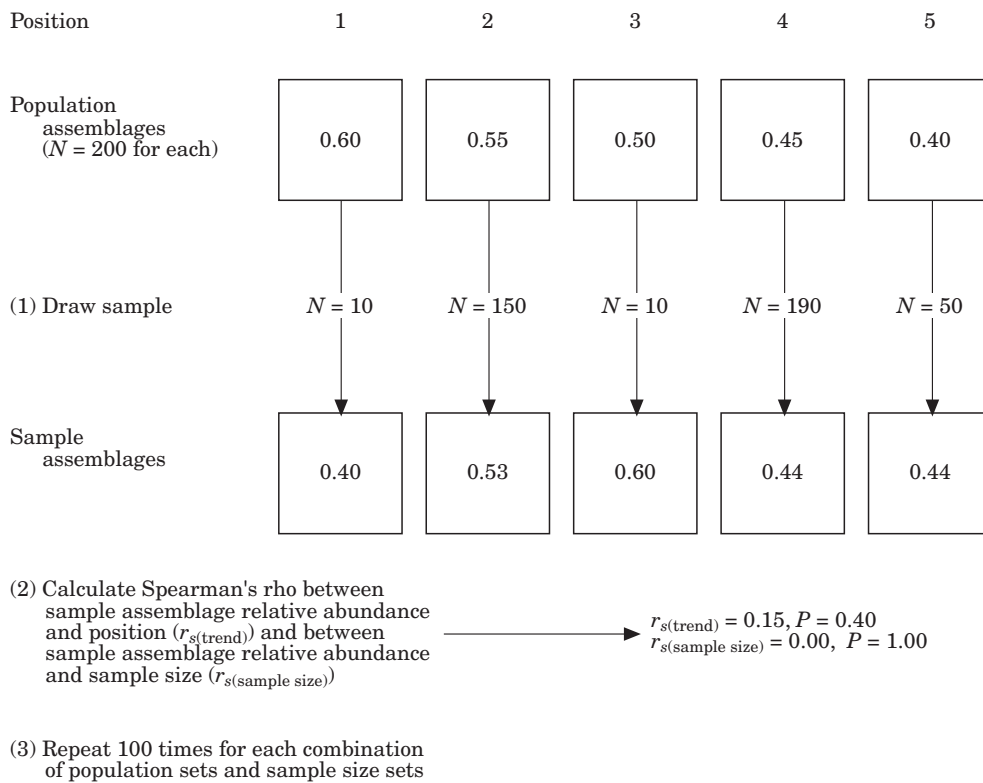


Figure 1. Schematic of the simulations used to evaluate the sample size correlation method for assessing sample size adequacy. The population assemblage relative abundance values in this example (the numbers inside the population assemblage boxes) are from Population Set 1, and the sample size values are from Sample Size Set 1. Sample relative abundance values (the numbers inside the sample assemblage boxes) and Spearman's rho results are those observed in the first sample case from the simulation run that used Population Set 1 and Sample Size Set 1.

sample case also possesses its own relative abundance value, but this value varies for a given position between sample cases within a run because the random samples drawn from each population assemblage will vary from one sample case to the next.

Finally, Spearman's rank order correlation coefficients (Spearman's rho) were calculated between sample assemblage relative abundance value and position in the time scale, and between sample assemblage relative abundance and sample size, for each of the 100 sample cases within each of the 80 runs. These Spearman's rho measurements will be the focus of my analysis here. Correlation coefficients between sample relative abundance and sample size are, of course, the means by which sample size adequacy has been

evaluated in analyses that have employed the sample size correlation method, and correlation coefficients between sample relative abundance and age are the means that have been used to determine whether temporal trends in taxonomic relative abundance are present in these analyses (e.g., Grayson, 1984, 1989; Broughton, 1994a, 1994b, 1997, 1999; Janetski, 1997; Grayson & Cannon, 1999). I use Spearman's rho because numbers of identified specimens, and relative abundance values derived from them, are best treated as ordinal scale measures (Grayson, 1984).\*

*Type II errors*

The simulations involving Population Sets 1 and 2, in which declining trends in relative abundance are present among the population assemblages, allow evaluation of whether correlations between sample size and relative abundance are useful for identifying the presence of Type II errors in analyses of trends in relative abundance. If sample size–relative abundance correlations are indicative of Type II errors resulting from the incorporation of small samples in such an

Table 1. Population proportions used in Monte Carlo simulations of samples taken from sets of five ordered population assemblages

Population set	Population assemblage proportion				
	1	2	3	4	5
1	0.60	0.55	0.50	0.45	0.40
2	0.20	0.15	0.10	0.05	0.01
3	0.50	0.50	0.50	0.50	0.50
4	0.10	0.10	0.10	0.10	0.10

\*The simulations presented in this paper were carried out through simple programs written for the Stata statistical package (version 5.0). The code for these programs is available from the author upon request.

Table 2. Sample sizes used in Monte Carlo simulations of samples taken from sets of five ordered population assemblages

Sample size set	Sample assemblage size					Sample size-position correlation <sup>a</sup>	Mean sample size	Mean sample size of assemblages 1 and 5
	1	2	3	4	5			
1	10	150	10	190	50	0.410	82	30
2	190	10	10	50	150	-0.051	82	170
3	50	150	190	50	10	-0.462	90	30
4	50	10	50	100	150	0.821	72	100
5	150	10	100	150	100	-0.053	102	125
6	150	50	10	190	150	0.308	110	150
7	50	10	190	150	150	0.564	110	100
8	100	10	190	50	190	0.410	108	145
9	50	150	100	150	10	-0.205	92	30
10	50	150	190	50	150	0.211	118	100
11	150	50	50	190	100	0.154	108	125
12	190	50	50	100	190	0.158	116	190
13	150	100	150	50	190	0.205	128	170
14	10	100	10	100	50	0.316	54	30
15	100	190	150	50	10	-0.700	100	55
16	10	150	50	100	150	0.564	92	80
17	190	50	50	150	150	-0.105	118	170
18	10	10	150	50	190	0.872	82	100
19	100	10	50	190	150	0.600	100	125
20	150	10	100	100	150	0.158	102	150

<sup>a</sup>Spearman's rho.

analysis, then they should occur most often in instances in which sample assemblages fail to display trends in relative abundance that are truly present in population assemblages.

It can be shown quite clearly that this is not the case. Table 3 is a contingency table in which the 4000 simulation sample cases that use Population Sets 1 and 2 are categorized according to whether or not Spearman's rho is significant at the  $P=0.05$  level for the relationship between sample size and sample relative abundance, and whether or not Spearman's rho is significant at the 0.05 level for the relationship between sample relative abundance and position along the "time" scale.\*† A chi-square test on this table suggests that significant correlations between sample size and relative abundance are distributed entirely randomly with respect to the presence or absence of significant declines in relative abundance ( $\chi^2=0.14$ ,  $P=0.71$ ). In

\*Since analyses of trends in relative abundance are usually approached with specific hypotheses in mind about the direction that those trends should take, one-tailed significance levels are the appropriate ones to use. "Significant decline in relative abundance" in this paper thus means that Spearman's rho is significant at the one-tailed 0.05 level. For correlations between sample size and relative abundance, on the other hand, there is generally no *a priori* reason to think that the correlation should take one direction rather than the other. Two-tailed significance levels are thus appropriate for the analysis of this relationship, and "significant correlation between sample size and relative abundance" in this paper means that Spearman's rho is significant at the two-tailed 0.05 level.

†No significant positive correlations between sample relative abundance and scale position—that is, significant increases in relative abundance—were observed in sample cases from the simulations involving Population Sets 1 and 2 in which the population assemblages displayed declines in relative abundance.

other words, sample cases that correctly show significant declines in relative abundance are just as likely to exhibit significant correlations between sample size and relative abundance as are sample cases that incorrectly fail to show significant declines in relative abundance. This result also holds when the sample cases in Table 3 are grouped by Population Set (Population Set 1:  $\chi^2=1.10$ ,  $P=0.29$ ; Population Set 2:  $\chi^2=1.68$ ,  $P=0.19$ ).

Population Sets 1 and 2 can also be used to explore the effects that variability in mean population relative abundance has on the probability that a Type II error will occur in analyses of trends in relative abundance. Table 4 categorizes sample cases by Population Set and the presence or absence of a significant decline in relative abundance, and Table 5 does the same for the presence or absence of significant correlations between sample size and relative abundance.

The first thing to notice is that many of the sample cases fail to find declines in relative abundance that are significant at the 0.05 level, even though such a trend is present among the population assemblages; the Type II error rate here is 29.5% overall (Table 4). Population Set 2, however, in which the relative abundance of the taxon of interest is much lower in all of the population assemblages, contains many fewer cases that fail to find this decline in relative abundance (18.9%, as opposed to 40.1% for Population Set 1). A chi-square test indicates that the difference between Population Sets is highly significant ( $\chi^2=215.2$ ,  $P<0.01$ ). This result is most likely due to the effects of the population proportion value on the standard error of sample proportion values. The standard error of a proportion provides a measure of the probability that the relative



Table 3. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in relative abundance, and significant and non-significant correlations between sample size and relative abundance: samples from simulations with trends in population relative abundance (Population Sets 1 and 2)

$\chi^2=0.14, P=0.71$	Significant sample size correlation	No significant sample size correlation
Significant decline in relative abundance	100	2721
No significant decline in relative abundance	39	1140

Table 4. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in relative abundance by Population Set: samples from simulations with trends in population relative abundance (Population Sets 1 and 2)

$\chi^2=215.2, P<0.01$	Population Set 1	Population Set 2
Significant decline in relative abundance	1199	1622
No significant decline in relative abundance	801	378

Table 5. Cross-tabulation of the number of sample cases exhibiting significant and non-significant correlations between sample size and relative abundance by Population Set: samples from simulations with trends in population relative abundance (Population Sets 1 and 2)

$\chi^2=4.66, P=0.03$	Population Set 1	Population Set 2
Significant sample size correlation	57	82
No significant sample size correlation	1943	1918

abundance value observed in any given sample will be greatly different from the true population value, and it can be shown that standard error is lowest when the proportion in the population being sampled equals 0 or 1 and highest when it equals 0.5 (e.g., Zar, 1996: 521–524). It may thus be possible to conclude that Type II errors in analyses of trends in relative abundance will also be most common when the proportion of the taxon of interest takes intermediate values in the populations being sampled.\*

\*Population proportions larger than those of Population Set 1 were not used in these simulations because the effects of this variable will be symmetrical around a value of 0.50. Population proportions of 0.80, 0.85, 0.90, 0.95, and 0.99, for example, would give the same results as those for Population Set 2, because a population in which 80% of the specimens belong to “taxon A” is equivalent to one in which 20% of the specimens are “not taxon A”, etc.

Unlike failures to find significant declines in relative abundance (Type II errors), significant correlations between sample size and relative abundance are not at all common among the Population Set 1 and 2 sample cases, occurring in only 3.5% of them overall (Table 5). However, they are slightly more common in Population Set 2 than in Population Set 1 (4.1% versus 2.9%), and this difference is significant at  $P=0.03$  ( $\chi^2=4.66$ ). These results further call into question the utility of the sample size correlation method as a means of assessing sample size adequacy. First, significant sample size correlations are more common in Population Set 2 than in Population Set 1, but Population Set 1 is the one in which Type II errors are most common. Second, significant correlations between sample size and relative abundance occur far less often in these sets of assemblages than do Type II errors, suggesting that most Type II errors would go undetected by the sample size correlation method even if there were an association between sample size–relative abundance correlations and Type II errors.

To understand these results better, the factors responsible for the presence or absence of significant correlations between sample relative abundance and both sample size and “time” scale position can be explored further. It is reasonable to expect that sample size is important in determining whether a significant trend in relative abundance will be correctly identified among a set of samples, and indeed this does seem to be the main factor responsible for the occurrence of significant correlations between sample relative abundance and scale position. Figure 2(a) presents the number of significant correlations between sample relative abundance and scale position observed in each of the 40 Population Set 1 and 2 simulation runs (consisting of 100 sample cases each) plotted against the mean of the five sample sizes used in each run (see Table 2). Runs involving Sample Size Sets with larger mean sample sizes do tend to contain more sample cases in which significant declines in relative abundance were found, and failures to find significant declines in relative abundance (Type II errors) are more common when mean sample size is smaller. A Spearman’s rank order correlation test indicates that this relationship is highly significant, both for the two Population Sets combined and for each one individually (combined:  $r_s=0.55, P<0.01$ ; Population Set 1:  $r_s=0.72, P<0.01$ ; Population Set 2:  $r_s=0.64, P<0.01$ ).

Significant correlations between sample relative abundance and sample size, on the other hand, are more common in runs in which sample sizes are highly correlated with time scale position. Figure 2(b) presents the number of significant correlations between sample size and relative abundance for each simulation run plotted against the absolute value of Spearman’s rho between sample size and scale position for each run (see Table 2). The absolute value of this correlation coefficient is used because it is irrelevant whether sample sizes increases or decreases with scale position.

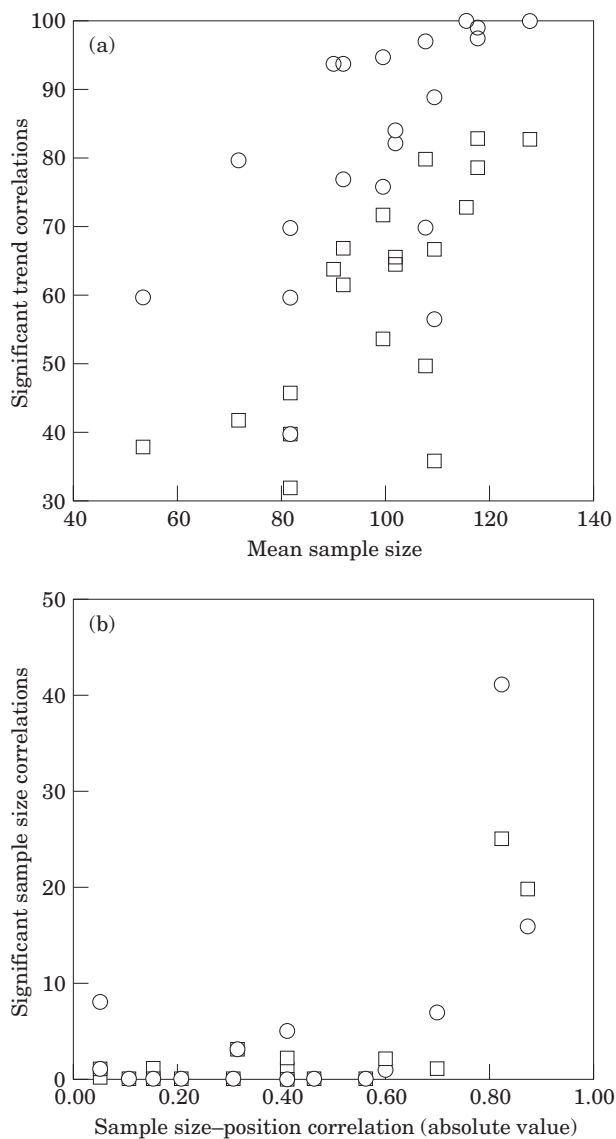


Figure 2. For the “declining trend” simulation runs involving Population Sets 1 ( $\square$ ) and 2 ( $\circ$ ): (a) relationship between the number of significant correlations between sample relative abundance and scale position (“significant trend correlations”) per run and the mean of the sample sizes used in each run; (b) relationship between the number of significant sample size–relative abundance correlations per run and the absolute value of the correlation between sample size and scale position for each run.

Runs using Sample Size Sets in which sample size tends either to increase or decrease with position along the scale produce more significant correlations between sample size and relative abundance. A Spearman’s rank order correlation test indicates that this relationship is highly significant for the two Population Sets combined, and slightly less so for each one individually (combined:  $r_s=0.40$ ,  $P<0.01$ ; Population Set 1:  $r_s=0.44$ ,  $P=0.03$ ; Population Set 2:  $r_s=0.36$ ,  $P=0.06$ ).

The reason for this relationship should be obvious: since relative abundance is correlated with scale

position in the population assemblages used in these simulations, sample cases that pick up this trend will of course show correlations between sample size and relative abundance when sample size is also correlated with scale position. Thus, when significant trends in relative abundance are observed, the presence of a significant correlation between sample size and sample relative abundance may often indicate nothing more than the presence of a correlation between sample size and position in a temporal or spatial ordering, such as might occur if sample sizes tend to increase or decrease with depth in a stratified archaeological site. The finding above that significant correlations between sample size and relative abundance are no more common among individual sample cases that fail to find significant trends in relative abundance than among sample cases that do find such trends (Table 3) strongly suggests, in fact, that a correlation between sample size and relative abundance should not be taken as an indicator of anything other than a correlation between sample size and temporal or spatial scale position.

The most important point here, however, is that significant sample size–relative abundance correlations and Type II errors are most likely to be observed under entirely different conditions: Type II errors appear to be largely a function of overall mean sampling percentage, whereas sample size–relative abundance correlations seem to occur primarily when sample size is correlated with time scale position anyway. Since there is no reason why overall mean sample size should be related in any way to the degree of correlation between sample size and scale position, it should not be surprising to find that there is no significant association between the occurrence of sample size–relative abundance correlations and the occurrence of Type II errors.

#### Type I errors

The sample size correlation method is usually used in an attempt to avoid Type I errors, or concluding that a trend is present among population assemblages when in fact one is not. The simulations involving Population Sets 3 and 4, in which population assemblages are all equal in relative abundance (Table 1), allow evaluation of how useful the sample size correlation method is at detecting this type of error.

In these simulations, finding a significant decline in relative abundance, rather than failing to find one, is what constitutes an error. I should note that I have counted as Type I errors only those sample cases that show *declines* in relative abundance, or *negative* correlations between sample relative abundance and scale position, and have not included the approximately equal number of sample cases that show significant *positive* correlations between sample relative abundance and scale position. I have done this based on the assumption that analyses of trends in relative abundance are most often approached with a specific

Table 6. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in relative abundance by Population Set: samples from simulations in which population relative abundance is equal in all assemblages (Population Sets 3 and 4)

	Population Set 3	Population Set 4
$\chi^2=0.61, P=0.44$		
Significant decline in relative abundance	81	91
No significant decline in relative abundance	1919	1909

Table 7. Cross-tabulation of the number of sample cases exhibiting significant and non-significant correlations between sample size and relative abundance by Population Set: samples from simulations in which population relative abundance is equal in all assemblages (Population Sets 3 and 4)

	Population Set 3	Population Set 4
$\chi^2=0.05, P=0.83$		
Significant sample size correlation	187	183
No significant sample size correlation	1813	1817

hypothesis in mind about the direction that the trend should take. When this is the case, of course, finding a significant trend in the opposite direction of that postulated by an hypothesis does not constitute a Type I error from the standpoint of hypothesis testing.

Before turning to the issue of the utility of the sample size correlation method for detecting Type I errors in analyses of trends in relative abundance, it should first be noted that, in contrast to the previous set of simulations, overall mean population relative abundance does not appear to have much of an effect on either the occurrence of Type I errors or the occurrence of significant sample size–relative abundance correlations when there is no trend present among population assemblages. Table 6 categorizes the sample cases from these “no trend” simulations by Population Set and by the presence or absence of a significant decline in relative abundance (Type I errors), and Table 7 categorizes these sample cases by Population Set and by the presence or absence of a significant correlation between sample size and relative abundance. The occurrence of both Type I errors and significant sample size–relative abundance correlations does not vary much between the two Population Sets, and the results of chi-square tests are not significant for either table (declines in relative abundance:  $\chi^2=0.61, P=0.44$ ; sample size–relative abundance correlations:  $\chi^2=0.05, P=0.83$ ).

The “no trend” simulations do suggest, however, just as the “declining trend” simulations did that the presence or absence of a significant correlation between sample size and sample relative abundance is not useful as a measure of sample size adequacy. Table 8 is a

Table 8. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in abundance and significant and non-significant correlations between sample size and relative abundance: samples from simulations in which population relative abundance is equal in all assemblages (Population Sets 3 and 4)

	Significant sample size correlation	No significant sample size correlation
$\chi^2=2.53, P=0.11$		
Significant decline in relative abundance	10	162
No significant decline in relative abundance	360	3468

contingency table that categorizes the 4000 sample cases from the runs involving Population Sets 3 and 4 according to whether or not significant correlations between sample size and sample relative abundance are present, and whether or not significant negative correlations between sample relative abundance and scale position (Type I errors) are present. If sample size–relative abundance correlations do indicate the presence of Type I errors resulting from the inclusion of small samples in an analysis, then they should be most common among sample cases that incorrectly find significant declines in relative abundance. A chi-square test on the values in Table 8 gives a *P*-value of 0.11 ( $\chi^2=2.53$ ) for a test of the null hypothesis that the distributions of significant sample size correlations and significant declines in relative abundance are independent. When the sample cases in this table are grouped by Population Set, the relationship is slightly more significant for the Population Set 3 cases ( $\chi^2=3.18, P=0.07$ ) and not significant for Population Set 4 ( $\chi^2=0.24, P=0.62$ ).

To the extent that a relationship does exist here, though, the direction that it takes suggests that relying solely on correlations between sample size and relative abundance to assess sample size adequacy may actually do more harm than good. That is, significant correlations between sample size and relative abundance tend to occur *less* often in sample cases in which Type I errors occur. Of the 172 cases in which Type I errors occurred, only 10 of these also exhibited significant correlations between sample size and relative abundance. Likewise, of the 370 cases in which significant sample size–relative abundance correlations were observed, only 10 of these were cases in which Type I errors occurred.

Why should sample size–relative abundance correlations be *less* likely to occur in cases in which Type I errors are present? The occurrence of a Type I error in a test of an hypothesis about a trend in relative abundance is determined largely by the samples at the ends of the ordering of assemblages. For a trend in relative abundance to be observed in sample assemblages when such a trend is not present in population assemblages, it must be the case that the relative

abundance values observed in the samples towards one end of the order be wrong in one direction and that the values observed in the samples at the other end of the order be wrong in the opposite direction. Since large sampling errors in relative abundance are more likely to occur when samples are small (that is, since the standard error of sample relative abundance values is larger for smaller samples; e.g., Zar, 1996: 521–524), it is reasonable to expect that Type I errors will be more likely to occur when the samples at both ends of an ordering of assemblages are small.

Figure 3(a) shows that this is the case. This figure displays the number of significant declines in sample relative abundance per run of 100 sample cases plotted against the mean size of the samples taken from the assemblages in the 1st and 5th positions in each run (see Table 2) for all 40 of the “no trend” simulation runs. Spearman’s rho for this relationship is negative and highly significant for all runs combined ( $r_s = -0.68, P < 0.01$ ), as well as for each Population Set individually (Population Set 3:  $r_s = -0.77, P < 0.01$ ; Population Set 4:  $r_s = -0.60, P < 0.01$ ). Type I errors do occur more often when the samples at the ends of the ordering are smaller.

In cases in which a trend in sample relative abundance is observed, however, the occurrence of a correlation between sample size and relative abundance requires that the samples towards one end of the trend be small and that the samples towards the other end of the trend be large. In other words, the problem of identifying Type I errors requires that we pay particular attention to sets of sample assemblages that contain small samples at both ends. A reliance solely on the sample size–relative abundance correlation method, though, virtually assures that the potentially most problematic sets of assemblages will never be called into question because sets of sample assemblages that contain small samples at both ends are not those that are most likely to produce significant correlations between sample size and relative abundance when a trend in relative abundance is also observed. This can be seen in Figure 3(b), which plots the number of significant sample size–relative abundance correlations per run against the mean size of the 1st and 5th position samples in each run for the Population Set 3 and 4 simulations. There is no relationship here ( $r_s = -0.08, P = 0.32$ ).

Finally, we can also consider the individual sample cases in which significant sample size–relative abundance correlations and Type I errors co-occur (Table 8). All 10 of these cases come from simulation runs that used a Sample Size Set for which the absolute value of Spearman’s rho between sample size and position was among the three highest (Table 2): two come from Sample Size Set 4 ( $r_{s(\text{sample size-position})} = 0.82$ ), two come from Sample Size Set 15 ( $r_{s(\text{sample size-position})} = -0.70$ ), and six come from Sample Size Set 18 ( $r_{s(\text{sample size-position})} = 0.87$ ). These sample cases thus seem to be the exceptions that prove the rule: signifi-

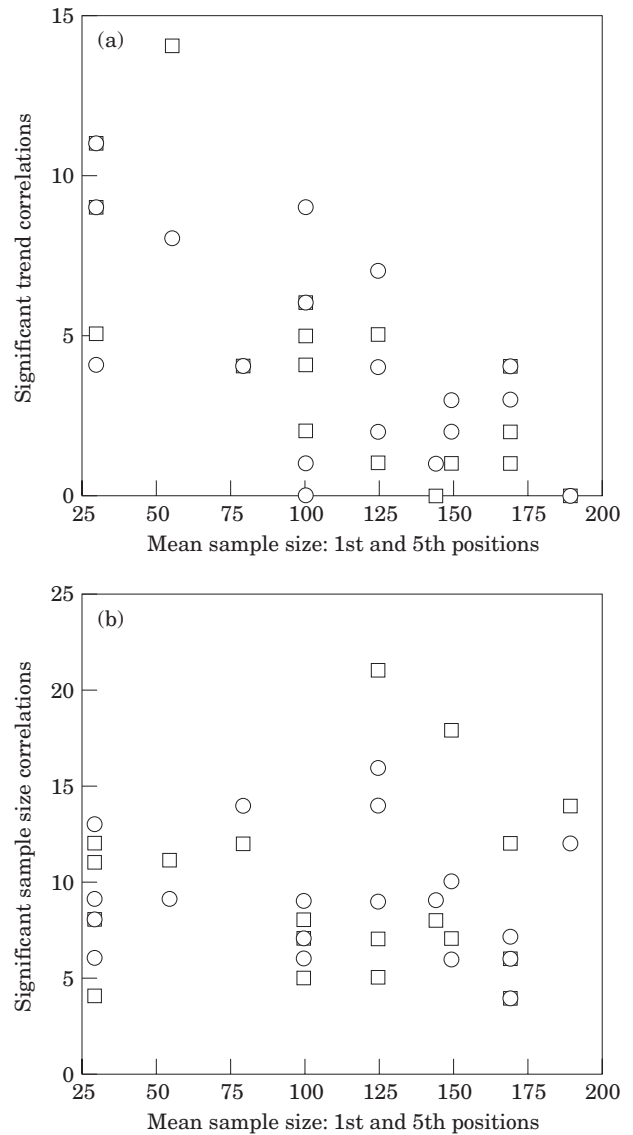


Figure 3. For the “no trend” simulation runs involving Population Sets 3 (□) and 4 (○): (a) relationship between the number of significant correlations between sample relative abundance and scale position (“significant trend correlations”) per run and the mean of the sample sizes used for the 1st and 5th position assemblages in each run; (b) relationship between the number of significant sample size–relative abundance correlations per run and the mean of the sample sizes used for the 1st and 5th position assemblages in each run.

cant sample size–relative abundance correlations and Type I errors are likely to co-occur only in instances in which sample size is correlated with temporal or spatial scale position anyway.

In summary, the sample size correlation method is not useful as a means of detecting errors in tests of hypotheses about trends in relative abundance among a set of ordered samples, and a sole reliance on this method may actually divert attention away from the sets of samples that are potentially the most problematic. The Monte Carlo simulations discussed here



suggest that there is no reason to expect a positive association between the occurrence of significant sample size–relative abundance correlations and the occurrence of either Type I or Type II errors in analyses of trends in relative abundance. This is because each of these kinds of errors are most likely to occur under conditions that are quite different from those that favour the occurrence of significant correlations between sample size and sample relative abundance. Fortunately, there is a better statistical method available for evaluating trends in relative abundance.

### Cochran's Test of Linear Trend

The root of the problem with the sample size correlation method is that it is used in conjunction with a method for evaluating trends in relative abundance that does not directly take sample size into account. Correlation coefficients are calculated between some measure of temporal or spatial variability and a measure of relative abundance that is expressed only as a proportion. Because relative abundance values are expressed only as proportions, all information about sample size is lost: it is not possible to tell from a relative abundance value of 0.57, for example, how many specimens were present in the sample assemblage from which that value was derived. Thus, the only way to address issues of sample size in analyses of this sort is to draw on a method, like the sample size correlation method, that is external to the test that is actually used to assess the trend in relative abundance (i.e., the correlation analysis between relative abundance and position in the time or space scale).

There are, however, methods available for evaluating trends in relative abundance among sets of sample assemblages that do take sample size into account more directly. Elsewhere, I have suggested the use of Cochran's test of linear trend (Cannon, 2000). This test is a form of chi-square analysis that partitions the total chi-square value for a two-row contingency table into a portion that is associated with a linear trend in relative abundance and a portion that reflects departure from that trend; the test can also be thought of as a form of regression analysis in which relative abundance values are weighted by sample size (e.g., Zar, 1996: 562–565). Cochran's test is sensitive to sample size in that the chance of finding a significant trend in relative abundance declines as the sizes of the samples included in an analysis decrease.

To contrast the performance of Cochran's test with that of the correlation method for evaluating trends in relative abundance, this test was performed on a subset of the sample cases generated in the simulations discussed above. Five sample cases were drawn from each of the 100 sample cases in each of the 80 simulation runs, resulting in a random 5% sample of all sample cases stratified by run. This allows comparison of the proportions of Type I and Type II errors that occur

Table 9. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in relative abundance by statistical test: samples from simulations with trends in population relative abundance (Population Sets 1 and 2)

	Spearman's rank order correlation test	Cochran's test of linear trend
$\chi^2=49.5, P<0.01$		
Significant decline in relative abundance	2821	187
No significant decline in relative abundance	1179	13

Table 10. Cross-tabulation of the number of sample cases exhibiting significant and non-significant trends in relative abundance by statistical test: samples from simulations in which population relative abundance is equal in all assemblages (Population Sets 3 and 4)

	Spearman's rank order correlation test	Cochran's test of linear trend
$\chi^2=8.97, P<0.01$		
Significant decline in relative abundance	172	0
No significant decline in relative abundance	3828	200

when Cochran's test is used to evaluate a trend in relative abundance with the proportions of these errors that occur when Spearman's rho is used.

For the simulation runs involving Population Sets 1 and 2, in which there is a decline in relative abundance among the population assemblages, Cochran's test of linear trend fails to find a decline in relative abundance that is significant at the 0.05 level in 6.5% of the sample cases (Table 9). This is much lower than the 29.5% Type II error rate that occurred when Spearman's rho was used to evaluate trends in relative abundance in the preceding section, and the difference in Type II error rates between the two tests is highly significant ( $\chi^2=49.5, P<0.01$ ).

The rate of Type I errors is also lower with the use of Cochran's test than with Spearman's rho (Table 10). Whereas Type I errors occurred in 4.3% of the sample cases when Spearman's rho was used to evaluate trends in relative abundance, Cochran's test found no declines in relative abundance that were significant at the 0.05 level, and this difference between the two tests is significant ( $\chi^2=8.97, P<0.01$ ). Of course, this result should not be interpreted as indicating that Type I errors will never occur when Cochran's test is used.

\*Interestingly, all 13 of the Type II errors that occurred using Cochran's test appeared in sample cases from runs involving Population Set 1, in which mean population relative abundance was higher than in Population Set 2. This is consistent with the finding discussed above that Type II errors in analyses of trends in relative abundance will be more likely to occur when the taxon of interest is present in intermediate proportions in population assemblages (Table 4).

Rather, the important point is that, when evaluating trends in relative abundance among the same set of simulated sample cases, the use of Cochran's test results in lower rates of both Type I and Type II errors than does the use of Spearman's rho. And, as demonstrated in the preceding section, when Spearman's rho is used to evaluate a trend in relative abundance, supplementing it by using the sample size correlation method will do nothing to increase the probability that a Type I or Type II error will be detected.

## Discussion

The sample size correlation method for evaluating sample size adequacy has been useful to the extent that it has called attention to the issue of sample size in zooarchaeological analyses of trends in relative abundance. The Monte Carlo simulations presented here, however, indicate that it is not useful as a general method for detecting errors in such analyses that are the result of sample size effects. The simulations suggest that the method will typically fail to find those sets of sample assemblages that are potentially the most problematic, and that it will often call into question sets of sample assemblages that do truly reflect underlying population trends.

The sample size correlation method has its roots in methods that were designed to explore, and take advantage of, relationships between sample size and assemblage richness, or the number of taxa or artefact classes present (Grayson, 1984). In this realm, correlation- and regression-based techniques for evaluating differences in richness between assemblages have proven to be quite powerful (e.g., Grayson & Cole, 1998; Grayson & Delpech, 1998). The effects of sample size on relative abundance are different from the effects of sample size on richness, however (e.g., McCartney & Glass, 1990), and different methods are necessary for addressing potential sample size problems in analyses of trends in relative abundance, as the simulation results presented here have shown.

In addition to the chi-square-based test that I have argued is superior to the previously used correlation-based methods, there exist alternative, or perhaps even complementary, methods that might also be applied to problems of this sort. These include incremental subsampling techniques, such as have been used by Lipo *et al.* (1997) to assess sample size adequacy in regards to assemblage richness in the context of frequency seriation (see also Kintigh, 1984, 1989; McCartney & Glass, 1990), and the Bayesian methods discussed by Robertson (1999). Whichever method is chosen in any particular analysis of temporal or spatial trends in relative abundance, however, more detailed attention should be given to the specific ways in which sample size can cause errors in tests of hypotheses about such trends. Only when this is done should it be concluded that trends in archaeological data truly reflect changes

that occurred in the past, rather than simply the effects of archaeological sampling.

## Acknowledgements

Donald K. Grayson, Robert D. Leonard, and an anonymous reviewer provided very helpful comments on earlier versions of this paper. Dale S. Cannon originally suggested to me that Cochran's test was the appropriate statistic for the analyses that I was conducting.

## References

- Broughton, J. M. (1994a). Declines in mammalian foraging efficiency during the Late Holocene, San Francisco Bay, California. *Journal of Anthropological Archaeology* **13**, 371–401.
- Broughton, J. M. (1994b). Late Holocene resource intensification in the Sacramento Valley, California: the vertebrate evidence. *Journal of Archaeological Science* **21**, 501–514.
- Broughton, J. M. (1997). Widening diet breadth, declining foraging efficiency, and prehistoric harvest pressure: ichthyofaunal evidence from the Emeryville Shellmound, California. *Antiquity* **71**, 845–862.
- Broughton, J. M. (1999). Resource depression and intensification during the late Holocene, San Francisco Bay: evidence from the Emeryville Shellmound vertebrate fauna. *University of California Publications: Anthropological Records* **32**. Berkeley: University of California Press.
- Cannon, M. D. (2000). Large mammal relative abundance in pit-house and pueblo period archaeofaunas from southwestern New Mexico: resource depression among the Mimbres–Mogollon? *Journal of Anthropological Archaeology*, **19**, 317–347.
- Grayson, D. K. (1984). *Quantitative Zooarchaeology*. New York: Academic Press.
- Grayson, D. K. (1989). Sample size and relative abundance in archaeological analysis: illustrations from spiral fractures and seriation. In (R. D. Leonard & G. T. Jones, Eds) *Quantifying Diversity in Archaeology*. Cambridge: Cambridge University Press, pp. 79–84.
- Grayson, D. K. (1991). Alpine faunas from the White Mountains, California: adaptive change in the prehistoric Great Basin? *Journal of Archaeological Science* **18**, 483–506.
- Grayson, D. K. & Cole, S. C. (1998). Stone tool assemblage diversity across the Middle-to-Upper Paleolithic transition in France. *Journal of Archaeological Science* **25**, 927–948.
- Grayson, D. K. & Delpech, F. (1998). Changing diet breadth in the early Upper Palaeolithic of southwestern France. *Journal of Archaeological Science* **25**, 1119–1129.
- Grayson, D. K. & Cannon, M. D. (1999). Human paleoecology and foraging theory in the Great Basin. In (C. Beck, Ed.) *Models for the Millennium: Great Basin Anthropology Today*. Salt Lake City: University of Utah Press, pp. 141–151.
- Janetski, J. C. (1997). Fremont hunting and resource intensification in the eastern Great Basin. *Journal of Archaeological Science* **24**, 1075–1088.
- Kintigh, K. W. (1984). Measuring archaeological diversity by comparison with simulated assemblages. *American Antiquity* **49**, 44–54.
- Kintigh, K. W. (1989). Sample size, significance, and measures of diversity. In (R. D. Leonard & G. T. Jones, Eds) *Quantifying Diversity in Archaeology*. Cambridge: Cambridge University Press, pp. 25–36.
- Leonard, R. D. (1997). The sample size–richness relation: a comment on Plog and Hegmon. *American Antiquity* **62**, 713–716.
- Lipo, C. P., Madsen, M. E., Dunnell, R. C. & Hunt, T. (1997). Population structure, cultural transmission, and frequency

- seriation. *Journal of Anthropological Archaeology* **16**, 301–333.
- McCartney, P. H. & Glass, M. F. (1990). Simulation models and the interpretation of archaeological diversity. *American Antiquity* **55**, 521–536.
- Quirt-Booth, T. & Cruz-Urbe, K. (1997). Analysis of leporid remains from prehistoric Sinagua sites, northern Arizona. *Journal of Archaeological Science* **24**, 945–960.
- Robertson, I. G. (1999). Spatial and multivariate analysis, random sampling error, and analytical noise: empirical Bayesian methods at Teotihuacan, Mexico. *American Antiquity* **64**, 137–152.
- Sanchez, J. L. J. (1996). A re-evaluation of Mimbres faunal subsistence. *Kiva* **61**, 295–307.
- Szuter, C. R. (1991). *Hunting by Prehistoric Horticulturalists in the American Southwest*. New York: Garland.
- Szuter, C. R. & Bayham, F. E. (1989). Sedentism and prehistoric animal procurement among desert horticulturalists of the North American Southwest. In (S. Kent, Ed.) *Farmers as Hunters: The Implications of Sedentism*. Cambridge: Cambridge University Press, pp. 80–95.
- Zar, J. H. (1996). *Biostatistical Analysis*. Upper Saddle River, New Jersey: Prentice Hall.