Chapter 13

Analysis of Frequencies

Within recent years there appears to have been an increasing awareness on the part of archaeologists that certain statistical techniques offer economical methods of extracting information of cultural significance from archaeological data. Albert Spaulding, 1953, in *Statistical Techniques for the Discovery of Artifact Types.*

With this opening sentence, Albert Spaulding introduced a now commonly used statistical procedure to archaeology called chi-square analysis. Spaulding was dissatisfied by typological procedures in common use at the time, and sought to make typology rigorous, replicable, and scientific. Most archaeologists at the time created archaeological types by putting artifacts together on a table, and then putting similar objects together in a pile. What constituted “similar” was up to the individual building the piles. Of course, what one individual determined to be similar differed from what another might see. This led to no end of problems—one of the most significant being that a select group of individuals who originally created types in a region became the “expert” on the type they created. As a consequence, if a student didn’t learn types at the feet of the Master (or one of the Master’s apprentices), their work was forever suspect.

Spaulding sought to put an end to this procedure, and offered statistical procedures as the answer. Spaulding felt that by demonstrating non-random associations between attributes of artifacts we would be able to “discover” the meaningful emic types that the original makers had in mind. Chi-square analysis does indeed identify such non-random associations. Let us consider the calculation of chi-square by using Spaulding’s historical example. Table 13.1 presents two hypothetical variables that Spaulding sought to use to discover ceramic types based on sherd frequencies. These data create a *matrix* of data in which the correspondence of two or more variables is measured.

Table 13.1. Matrix of a hypothetical frequencies of sherds classified by their surface treatment and temper type.

	Grit Temper	Shell Temper	Total
Stamped Surface	25	25	50
Smooth Surface	25	25	50
Total	50	50	100

In this example we have equal observations in each cell. That is, there are 25 sherds that have grit temper and a stamped surface, 25 with shell temper and a stamped surface, etc. This is the pattern we would expect to see if we had 100 sherds and if there were no association between temper type and surface treatment.

Table 13.2 presents the opposite—a perfect association between temper type and surface treatment.

Table 13.2. Hypothetical frequencies of sherds classified by their surface treatment and temper type in which there is a strong association between the two variables.

	Grit Temper	Shell Temper	Total
Stamped Surface	50	0	50
Smooth Surface	0	50	50
Total	50	50	100

Here, grit temper is always associated with a stamped surface, and shell temper with a smooth surface. To Spaulding, these would constitute emic types, a function of the will of the potters involved in making the original ceramics. The associations here are quite unambiguous, and we don't really need any statistics to identify them.

Let us consider another situation that is not quite so clear. Consider the data presented in Table 13.3.

Table 13.3. Another set of hypothetical frequencies of sherds classified by their surface treatment and temper type.

	Grit Temper	Shell Temper	Total
Stamped Surface	12	134	146
Smooth Surface	222	24	246
Total	234	158	392

While we definitely see positive associations between a stamped surface and shell temper, and smooth surface and shell temper, is it beyond what we could expect due to chance? Chi-square analysis provides us with a means of answering this question by comparing what we observe to what we would expect due to chance and then determining if the difference is more than we would expect.

Table 13.3 presents our observed examples, symbolized as O. We determine our expected values, symbolized as E, for each observed value through this simple calculation:

$$E = \frac{RT \times CT}{GT}$$

where RT = Row Total, CT = Column Total, and GT = Grand Total.

Note that in all of our tables, only four cells contain observed values. By convention we label these cell's a, b, c and d as illustrated in Table 13.4.

Table 13.4. Designations used for 2 x 2 matrix tables.

a	b
c	d

So, for cell a in Table 13.3 we have 12 sherds that are of grit temper and stamped surface. To calculate E for cell a in Table 13.3 we proceed as follows.

$$E = \frac{RT \times CT}{GT} = \frac{146 \times 234}{392} = 87.15$$

Given our row and column totals, we would expect cell a to have approximately 87 sherds. Values for the remaining three cells can be calculated in the same way. Table 13.5 presents both observed and expected values for all cells.

Table 13.5. Expected and observed sherd frequencies for the hypothetical data presented in Table 13.3.

	Grit Temper	Shell Temper	Total
Stamped Surface	O=12 E=87.15	O=134 E=58.85	146
Smooth Surface	O=222 E=146.85	O=24 E=99.15	246
Total	234	158	392

Do the deviations of observed from expected values differ in such a manner that we could conclude that there is a meaningful association between surface treatment and temper type? Rephrasing this question in the form of a null hypothesis, we are asking if $H_O : O_{ij} = E_{ij}$. This hypothesis can be tested by calculating the chi-square value, which is symbolized by χ^2 and is calculated as follows:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Table 13.6 presents these calculations in table form.

Table 13.6. Chi-square test evaluating $H_O : O_{ij} = E_{ij}$.

Cell	Observed	Expected	(O-E)	$(O-E)^2$	$(O-E)^2/E$
a	12	87.15	-75.15	5647.52	64.80
b	134	58.85	75.15	5647.52	95.96
c	222	146.85	75.15	5647.52	38.46
e	24	99.15	-75.15	5647.52	56.96
$\sum$	392	392	0		256.18

χ^2 = 256.18. We can now compare this value to the critical value corresponding with our chosen significance level listed in Appendix D. The degrees of freedom are calculated as:

df=(R-1)(C-1)

where R is the number of rows in a matrix and C is the number of columns. In our example above (Table 13.3), we have two columns and two rows. The degrees of freedom consequently are:

$$df = (R - 1)(C - 1) = (2 - 1)(2 - 1) = 1 \times 1 = 1.$$

The critical value for a level of significance of .05 and one degree of freedom is 3.841. The Chi-square value of 256.18 listed on Table 13.6 exceeds this value. We therefore reject the null hypothesis and conclude that the observed values are indeed different than the expected values. But what does this mean? Well that depends entirely on one's hypothesis of interest. Let us consider a real archaeological example to illustrate the utility of the Chi-square test.

Table 13.7 presents the frequencies of flaked stone artifacts recovered from three room blocks (Conjuntos) and a plaza area at Galeana, a large pueblo-like site in northwestern Mexico, during our initial survey of the site. We are interested in determining whether there are any differences in the frequencies of flaked stone artifacts grouped by their raw material, which could in turn reflect possible differences in activities performed at each location.

Table 13.7. Frequencies of flaked stone artifacts grouped by provenience and raw material.

Raw Material	Conjunto2	Conjunto 3	Conjunto 4	Plaza	Total
Chert	86	21	38	97	242
Chalced.	39	10	12	13	74
Obsidian	4	3	3	8	18
Quartzite	16	8	7	6	37
Igneous	217	63	81	353	714
Total	362	105	141	477	1085

Looking at the table, there are clearly differences in the frequencies of flaked stone artifacts. For example, 86 chert artifacts were recovered from Conjunto 2 but only 21 were recovered from Conjunto 3. However, 362 flaked stone artifacts were recovered from Conjunto 2 compared to 105 from Conjunto 3. Given the differences in the number of flaked stone artifacts recovered from the four proveniences we would naturally expect some differences in the actual frequencies of each raw material type, so we really aren't concerned about difference in the absolute number of chert artifacts.

Instead, our hypothesis of interest is whether there are differences in the number of flaked stone artifacts of each raw material type expected based on the sample size from each provenience. This is a perfect application for the Chi-Square test.

Our null hypothesis is $H_O : O_{ij} = E_{ij}$. The expected values are determined as previously illustrated. The level of significance is set at .05, and the results of the analysis are presented in Table 13.8.

Table 13.8. Chi-square test comparing the relative frequencies of lithic raw materials from four areas of Galeana, Chihuahua, Mexico.

Provenience	Raw Material	Observed	Expected	Chi-Square Value
Conjunto2	Chert	86	80.7	0.34
	Chalcedony.	39	24.7	8.29
	Obsidian	4	6.0	0.67
	Quartzite	16	12.3	1.08
	Igneous	217	238.2	1.89
Conjunto 3	Chert	21	23.4	0.25
	Chalcedony	10	7.2	1.13
	Obsidian	3	1.7	0.91
	Quartzite	8	3.6	5.45
	Igneous	63	69.1	0.54
Conjunto 4	Chert	38	31.4	1.36
	Chalcedony	12	9.6	0.59
	Obsidian	3	2.3	0.19
	Quartzite	7	4.8	1.00
	Igneous	81	92.8	1.50
Plaza	Chert	97	106.4	0.83
	Chalcedony	13	32.5	11.73
	Obsidian	8	7.9	0.00
	Quartzite	6	16.3	6.48
	Igneous	353	313.9	4.87
		Chi-Square value=		49.10
		Critical Value (.05, 12)=		21.03

The chi-square value from the test is 49.10, which exceeds the critical value of 21.03 for 12 degrees of freedom. We reject the null hypothesis and conclude that at least some of the raw materials are present in different relative frequencies at each provenience.

But which raw materials? The analysis in Table 13.8 demonstrates differences, but it doesn't tell us what those differences are. Is there more chert than expected at Conjunto 2? Well, yes; 86 chert artifacts were recovered but only 80.7 were expected. Does this difference reflect a statistically significant difference between the expected and observed values? We can't tell by looking at Table 13.8. All of the observed and expected values differ somewhat, some by very little and some by a bit more. Which of these differences are significant?

We could reason that the largest differences are more likely to be responsible for the statistical difference we determined through the test, a conclusion that is supported by the fact that these pairs of observed and expected values also produced the highest chi-square values. Based on the size of the chi-square value associated with each pair, we might even be able to identify those differences that contributed the most to the total chi-square value, and thereby be able to determine which differences are statistically significant.

For example, the difference between the observed and expected frequencies of chert at Conjunto 2 produced a chi-square value of only .34, which seems small when compared to the chi-square values associated with chalcedony artifacts from the plaza (11.73) or quartzite from Conjunto 3 (5.45). However, we don't really know at what point the differences shift from being small to significant. We could guess, but that seems unnecessarily arbitrary.

Instead of guessing, we would ideally like to be able to calculate some sort of standardized measure of variation for each chi-square value, which in turn would allow us to determine whether each observed and expected value differed significantly at some significance level. Fortunately, and as you have probably already surmised, there is such a measure. It is called the residual and is calculated using the following formula:

$$e_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}} .$$

The chi-square residual is the mathematical equivalent of the square root of the chi-square value with the exception that the residual can be either positive or negative. Positive values indicate that the observed frequency was greater than expected whereas negative values indicate fewer observed than expected. Once e_{ij} has been calculated, it can be compared to standard deviation units from the standardized normal distribution that correspond to a given significance level (e.g., 1.96 for α=.05; see Appendix A). A residual greater than the critical value (e.g., e_{ij}>1.96) or less than the negative version of the critical value (e.g., e_{ij}<-1.96) indicates a significant difference at the specified significance level.

The chi-square residual has been demonstrated to be biased such that it tends to underestimate the significance of differences for small samples. As a result, it is best to calculate the adjusted residual as follows:

$$d_{ij} = \frac{e_{ij}}{\sqrt{\left(1 - \frac{CT}{GT}\right)\left(1 - \frac{RT}{GT}\right)}}$$

CT, RT, and GT again stand for the column total, row total, and grand total for a given value in the matrix.

To illustrate the calculation of the adjusted residual, consider Table 13.9, which presents the observed and expected values from the Galeana survey.

Table 13.9. Observed and expected values for the Galeana data.

Raw Material	Conjunto 2	Conjunto 3	Conjunto 4	Plaza	Total Observed
Chert	O=86 E=80.7	O=21 E=23.4	O=38 E=31.4	O=97 E=106.4	242
Chalcedony	O=39 E=24.7	O=10 E=7.2	O=12 E=9.6	O=13 E=32.5	74
Obsidian	O=4 E=6.0	O=3 E=1.7	O=3 E=2.3	O=8 E=7.9	18
Quartzite	O=16 E=12.3	O=8 E=3.6	O=7 E=4.8	O=6 E=16.3	37
Igneous	O=217 E=238.2	O=63 E=69.1	O=81 E=92.8	O=353 E=313.9	714
Total Observed	362	105	141	466	1085

The residual for the first cell (i=1, j=1) in the matrix is calculated as:

$$e_{1,1} = \frac{O_{1,1} - E_{1,1}}{\sqrt{E_{1,1}}} = \frac{86 - 80.7}{\sqrt{80.7}} = \frac{5.3}{8.98} = .59 .$$

The adjusted residual is calculated as:

$$d_{1,1} = \frac{e_{1,1}}{\sqrt{\left(1 - \frac{CT}{GT}\right)\left(1 - \frac{RT}{GT}\right)}} = \frac{.59}{\sqrt{\left(1 - \frac{362}{1085}\right)\left(1 - \frac{242}{1085}\right)}} = \frac{.59}{\sqrt{.66 * .78}} = .81 .$$

Because .81 is not larger than 1.96 or smaller than –1.96 (the critical values for α=.05), we conclude that the differences between the observed and expected frequencies of chert artifacts from Conjunto 2 are not significant. The adjusted residuals for the rest of the cells in Table 13.9 can be calculated in the same way.

Table 13.10 presents the chi-square test previously reported in Table 13.8 with the addition of the adjusted residuals. Bolded values indicate adjusted residuals corresponding with differences significant at α=.05.

Table 13.10. Adjusted residuals for the chi-square analysis of flaked stone raw materials recovered from four proveniences.

Provenience	Raw Material	Observed	Expected	Chi-Square Value	Adjusted Residuals
Conjunto 2	Chert	86	80.7	0.34	0.81
	Chalcedony.	39	24.7	8.29	**3.66**
	Obsidian	4	6.0	0.67	-1.01
	Quartzite	16	12.3	1.08	1.30
	Igneous	217	238.2	1.89	**-2.88**
Conjunto 3	Chert	21	23.4	0.25	-0.60
	Chalcedony	10	7.2	1.13	1.16
	Obsidian	3	1.7	0.91	1.01
	Quartzite	8	3.6	5.45	**2.50**
	Igneous	63	69.1	0.54	-1.32
Conjunto 4	Chert	38	31.4	1.36	1.42
	Chalcedony	12	9.6	0.59	0.85
	Obsidian	3	2.3	0.19	0.47
	Quartzite	7	4.8	1.00	1.09
	Igneous	81	92.8	1.50	**-2.24**
Plaza	Chert	97	106.4	0.83	-1.38
	Chalcedony	13	32.5	11.73	**-4.74**
	Obsidian	8	7.9	0.00	0.04
	Quartzite	6	16.3	6.48	**-3.46**
	Igneous	353	313.9	4.87	**5.04**

Chi-Square value=49.10
Critical Value (.05, 12)=21.03

Given the adjusted residuals, it is no longer difficult to determine what differences underlie the rejection of the null hypothesis. We now can conclude that we recovered more chalcedony artifacts from Conjunto 2, more quartzite artifacts from Conjunto 3, and more crystalline ingenious artifacts from the plaza area than expected by change. We likewise collected fewer igneous artifacts from Conjuntos 2 and 4, and fewer chalcedony and quartzite artifacts from the plaza than expected. Based on these differences, we can further conclude that the materials used in the plaza tended to be more crystalline, which result in more durable tools with dull edges when compared to the other raw materials, than was common in the room blocks. Furthermore, differences between the room blocks might provide additional information about changes in raw material selection through time, differences in the use of raw materials, or differences in access to raw materials among the inhabitants of Galeana. These differences then could (and did) help frame our future research questions and field research.

We note that although the adjusted residual is very useful when the chi-square test results in the rejection of the null hypothesis, it should not be used when a significant differences was not present, i.e., when one fails to reject the null hypothesis. When the chi-square does not produce a significant result, we can conclude that NONE of the expected and observed values in the table were sufficiently large to conclude that the differences between them are not the result of chance. Use of the adjusted residuals to try to identify significant differences are unnecessary at best and downright misleading at worst.

Assumptions of Chi-Square Analysis

The chi-square test is extremely powerful and can be very useful to archaeologists, who frequently deal with counts of artifacts and other materials. It does have assumptions that must be met before it can be used, however.

Two of these assumptions are obvious. First, the chi-square test requires that data be organized at nominal or ordinal scale. Although ordinal and interval level data can be transformed into nominal scale, it is seldom worthwhile to do so. Very powerful statistical tools such as the t-test and correlation are available for use with these data.

Second, the data must be derived as part of an independent random sample. Put another way the variables considered during the analysis should be independent of one another such that one is not a direct product of the other. For example, it would be of limited utility to compare the frequency of crystalline, cryptocrystalline, and noncrystalline artifacts by raw material in the Galeana example presented above. The coarseness of flaked stone artifacts is directly correlated with raw material type, especially in the case of materials such as obsidian. These variables are not independent, and therefore violate the assumptions of the chi-square test.

The third and most significant assumption reflects the nature of the chi-square distribution, upon which the chi-square test is built. Just as there is a normal distribution, so there is a chi-square distribution. This distribution is skewed heavily to the right, and the critical values reported in Appendix D relate to the areas under the curve of the chi-square distribution. However, the shape of the distribution changes, just as the T distribution does, with changes in sample size.

Statisticians realized that with a small enough sample size, the chi-square distribution would loose its characteristic shape and in fact begin to resemble a normal distribution. When this happened, the critical values calculated for the chi-square would no longer be accurate, and the results of a chi-square test would become spurious. The point at which this happened was unknown when the chi-square test came into common usage but a commonly accepted convention, called the rule of 5, was developed that guaranteed its applicability.

The rule of 5 holds that at least 80% of the expected frequencies must be 5 or more before chi-square is appropriate. This rule was agreed upon as a 'safe bet' based on the fact that statisticians were sure that the chi-square test would be applicable under these conditions (see Cochran 1954). This rule was widely accepted and has been enshrined in statistics books ever since.

Recent research such as Larntz (1978), Lewontin and Felsenstein (1965), Roscoe and Byars (1971), Slakter (1966), and Yarnold (1970) have demonstrated that the Chi-Square Test is generally applicable even if a significant proportion of the expected values are less than 5 (see also Everitt, 1992:39). Lewontin and Felsenstein (1965:31) in fact argued based on the results of a Monte Carlo simulation consisting of 1,500 Chi-Square values derived using randomly created data that the Chi-Square statistic will be correctly distributed as long as all of the expected values are 1 or greater, and, by extension, that the Chi-Square Test will produce valid statistical results with these small samples (see also Roscoe and Byars 1971:758). However, traditions die hard. Most archaeologists (and others) continue to apply the rule of 5, and look very skeptically at chi-square tests applied to samples with small expected values.

It is reasonable to question whether small samples do in fact accurately reflect the parent population, but the research noted above does suggest that the chi-square test is as appropriate for these small samples as are the other statistical methods proposed as alternatives. We therefore note that you will wish to seriously consider: 1) whether a small sample resulting in many expected values less than 5 accurately reflects the parent population, and 2) whether you wish to apply the chi-square test (or any other statistical method) to such small samples.

If one wishes, there are several ways to increase the size of expected values. The first and perhaps easiest way is to eliminate categories represented by small samples. For example, if one is analyzing pottery frequencies of sites and there is a particular trade ware that occurs infrequently (n<5) at all of the site, then perhaps it would be worthwhile to eliminate this trade ware from one's analysis. This is a viable alternative as long as the elimination of the category does not impact one's hypothesis of interest. If one is trying to examine the general distribution of pottery types among the site, then eliminating the trade ware may not damage one's analysis at all, because it occurs in such low frequencies at all of the sites. However, if one's hypothesis of interest focuses on the frequency of exotica in the sites, then eliminating the trade ware directly impacts one's analysis and is thus not desirable.

A second possibility is to collapse categories. For example, instead of having many multiple trade ware types in one's analysis that each occur in low frequencies, one could collapse them into fewer (or a single) category. This would increase their frequencies and would thereby increase the expected values. Again, care must be exercised to make sure that one does not change the structure of the data such that it is no longer possible to evaluate the hypothesis of interest.

Finally, larger samples will increase the frequencies of even rare categories. Although it is not always possible to increase one's sample size, this is perhaps the best method to deal with small expected frequencies. After all, as sample size increases, so does our certainty that our sample accurately reflects the population we are trying to study. Increasing sample size thus has the benefit of improving both the strength of the statistical test and our confidence in the conclusions derived through the test.

Nonparametric ANOVA

As previously discussed, ANOVA is a very powerful tool for comparing multiple data sets, but it does require that each of the data sets be normally distributed. Many of the distributions that archaeologists study are not normally distributed, though. For example, measurements of artifact length, width, thickness, and weight are frequently skewed to the right, with a few large artifacts preventing a normal distribution. Also we sometimes deal with samples that are too small to accurately determine the shape of the underlying distribution, even though they may be adequate to provide estimates of certain population parameters. What do we do in such cases when we would like to apply ANOVA, but are uncertain if its underlying assumptions are applicable?

Fortunately there are nonparametric versions of ANOVA that do not require assumptions of the underlying shapes of the distributions. One means of completing a nonparametric Analysis of Variance is to use median values as measures of central tendency. Unlike means that are heavily impacted by outliers, medians are robust measures of a distribution's central tendencies in skewed distributions, because they are impervious to the absolute size of outliers. Differences in medians can be evaluated by computing the median value of all of the data under consideration, and then determining the number of members of each class that are greater than and less than the median value.

The total number of variates that are greater than the median will be equal to the number of variates that are less than the median across all of the groups, an obvious result of how the median is calculated, but some classes may have an inordinate number of variates either above or below the group median. Using a Chi-square test, the presence of classes that tend to be smaller or larger than the group median can be detected. Such an approach, which effectively compares the medians of the classes, makes no assumptions concerning the shape of the distribution of the data.

Consider the following data in Table 13.11, which are the length measurements of four classes of arrow points from the Hohokam period occupation of Ventana Cave. We might be interested in determining if any of the classes tend to be longer than the others. We could accomplish this by comparing the means of each sample of points. Is the parametric form of ANOVA introduced in Chapter 10 appropriate for analyzing these data? We don't know. The samples are large enough to provide reasonable estimates of population parameters such as μ, but we don't really know enough about the underlying distributions of point length to determine the shape of the distribution. We could apply parametric ANOVA, and hope that it is applicable. Or we could use a nonparametric ANOVA. If we are good archaeologists, we will likely opt for the second alternative.

Table 13.11. Length measurements for four classes of arrowheads from Ventana Cave, Arizona.

Side-notched	Corner-notched	Straight Stemmed	Triangular
1.98	2.71	2.1	2.31
2.2	2.08	3.07	2.01
2.39	2.55	2.25	2.57
2.61	2.45	2.4	2.8
2.4	2.25	2.81	2.4
2.59	2.98	2.27	3
3.28	2.61	2.72	2.4
3.41		3.02	2.89
3.81		3.72	2.82
2.81			3.32
2.55			

For the nonparametric ANOVA, the null hypothesis is $H_0 : M_1 = M_2 = ... = M_i$, where M (the Greek letter) is the symbol of the population parameter for the median of each group. In this case the null hypothesis is $H_0 : M_1 = M_2 = M_3 = M_4$. We will set the level of rejection at .05 again.

The first step in calculating the nonparametric ANOVA is to determine the group median of all of the data in Table 13.11. In this case, n=37. The group median is 2.59, which is the 19th (the center of 37) largest value in Table 13.11.

The next step is to determine which of the values in Table 13.11 are larger and which are smaller than the median. The magnitude of the difference is unimportant; only the direction matters. Table 13.12 provides this information. Note that the number of plus and minus signs must be equal through the entire matrix, because there must be an equal number of variates above and below the median by definition. The variate(s) that directly correspond with the median are excluded.

Table 13.12. The direction of difference between each variate listed in Table 13.11 and the group median. Note that "+" indicates that the variate is greater than the group median whereas "-" indicates it is smaller.

Side-notched	Corner-notched	Straight Stemmed	Triangular
-	+	-	-
-	-	+	-
-	-	-	-
+	-	-	+
-	-	+	-
NA	+	-	+
+	+	+	-
+		+	+
+		+	+
+			+
-			

Table 13.13 presents the counts for each class.

Table 13.3. Counts of the number of variates greater than and less than the median.

	Side-notched	Corner-notched	Straight Stemmed	Triangular
Great than Median	5	3	5	5
Less than Median	5	4	4	5

If the null hypothesis is true and the medians are equal, than each class should have an equal number of variates greater than and less than the group median. This proposition can be evaluated with a Chi-square test. The results of the Chi-square test are presented in Table 13.14.

Table 13.14. Chi-square test comparing the frequencies of observed and expected values of arrow points that are greater than and less than the group median.

Direction of Difference	Point Class	Observed	Expected	Chi-square value
Larger	Side-notched	5	5	0.00
	Corner-notched	3	3.5	0.07
	Straight Stemmed	5	4.5	0.06
	Triangular	5	5	0.00
Smaller	Side-notched	5	5	0.00
	Corner-notched	4	3.5	0.07
	Straight Stemmed	4	4.5	0.06
	Triangular	5	5	0.00
		Chi-square value=		0.25
		Critical value (.05,3)=		7.81

In this case the Chi-square value of .025 is not greater than the critical value of 7.81 for 3 degrees of freedom. As a result we cannot reject the null hypothesis and we conclude that there are no significant differences in the median point length among the four classes of points. This conclusion is in fact similar to the conclusion that there are no apparent differences in the mean lengths of each of the classes that we would have reached using a parametric ANOVA, but we are certain that we are not violating any critical assumptions by using the nonparametric ANOVA.

The nonparametric ANOVA requires interval or ratio level data. Such scales of measurement are necessary if one is to identify a median. At times, it is possible to apply some of the same principles to ordinal level data using categories such as large, medium, and small. When doing so, one cannot determine a median. Instead, one would need to count the number of members in each class that fall within each of the ordinal categories and then compare their frequencies using a Chi-square test (e.g., compare the number of large, medium, and small artifacts within each class). This can be very effective, but it is more properly considered an application of the Chi-square test instead of a nonparametric ANOVA.

But what about Spaulding's suggestion for using the Chi-square test to define artifact types, which we used to begin this chapter? Spaulding's suggestion initiated one of the most interested debates in archaeology, one that is in many ways continuing to this day. His exchange in James Ford reflected and affected the ways that anthropologists conceptualize both the archaeological record and the use of statistics to study it. In fact, archaeologists have come to use certain statistics in certain ways so commonly that any professional archaeologist is guaranteed to come across them. Unfortunately, the ubiquity of these statistical applications often causes archaeologists to ignore important underlying issues, and instead to use the statistics without consideration of their attributes and limitations. We turn to these "special" applications in the final chapter.