

Society for American Archaeology

Alternative Procedures for Assessing Standardization in Ceramic Assemblages

Author(s): Kenneth L. Kvamme, Miriam T. Stark, William A. Longacre

Source: *American Antiquity*, Vol. 61, No. 1 (Jan., 1996), pp. 116-126

Published by: [Society for American Archaeology](#)

Stable URL: <http://www.jstor.org/stable/282306>

Accessed: 02/12/2010 15:37

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=sam>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Society for American Archaeology is collaborating with JSTOR to digitize, preserve and extend access to *American Antiquity*.

ALTERNATIVE PROCEDURES FOR ASSESSING STANDARDIZATION IN CERAMIC ASSEMBLAGES

Kenneth L. Kvamme, Miriam T. Stark, and William A. Longacre

Interest in the material correlates of economic specialization has led to numerous quantitative studies of standardization (or the lack of it) in craft production, particularly of ceramics. Most of this research has focused on measures of variation and conventional statistical procedures in the treatment of the empirical data, many of which are dependent on unrealistic assumptions (such as normal populations), yielding results that can be questioned. To resolve this crisis more robust statistical methods are investigated including the "jackknife" method (for confidence interval construction) and a multigroup test for homogeneity of variance known as the Brown-Forsythe Test. Computer simulations show that the latter is robust under a variety of distributional forms. These new methods are used to reanalyze ethnoarchaeological ceramic data from the Philippines; it is shown that markedly different conclusions can be reached when compared with the results of more conventional procedures.

Interés en las correlaciones de especialización económica ha conducido a numerosos estudios cuantitativos de la estandarización (o la falta de) en la producción de artesanía, particularmente de cerámica. La mayor parte de esta investigación se ha enfocado en las medidas de variación y procedimientos estadísticos convencionales usados en el análisis de datos empíricos, muchos de los cuales dependen de suposiciones irreales (tal como poblaciones normales), produciendo resultados que pueden ser dudosos. Para resolver esta crisis, métodos estadísticos alternativos y más robustos son investigados, incluyendo el "jackknife" (para construcciones de intervalos de confianza) y una prueba multi-grupo para la homogeneidad de varianza conocida como la prueba Brown-Forsythe. Simulaciones computarizadas demuestran que la segunda prueba es robusta en una variedad de formas distribucionales. Estos nuevos métodos son usados para re-analizar datos etno-arqueológicos de cerámica proveniente de las Filipinas y se demuestra que las conclusiones son marcadamente diferentes en comparación con los resultados obtenidos con procedimientos más convencionales.

Studies of ceramic standardization have gained prominence in recent years, in part because of interest in the material correlates of economic specialization (Brumfiel and Earle 1987; Rice 1981, 1987, 1991; Tosi 1984). Underlying such studies is "the standardization hypothesis" (Blackman et al. 1993), the assumption that production intensity is reflected through increased product uniformity (e.g., Costin 1991; Davis and Lewis 1985; Feinman et al. 1981; Hagstrum 1988; Kramer 1985; Rice 1991). The focus of this paper is on the methods that archaeologists have employed to examine standardization issues in empirical research.

Quantitative studies of ceramic standardiza-

tion generally focus on measures of variation, notably the sample variance, standard deviation, or the coefficient of variation. That is, statistical tests are commonly employed to assess whether significant differences in standardization exist between groups of interest as reflected by these measures. Most standardization studies use an F -test based on the ratio of two sample variances as a basis for inference, and confidence intervals derived from the chi-square distribution are sometimes employed (e.g., Allen 1992; Arnold 1991; Arnold and Nieves 1992; Benco 1988; Hagstrum 1988; Longacre et al. 1988; Sinopoli 1988; M. Stark 1991). These procedures were certainly adequate as a starting point, but they are

Kenneth L. Kvamme ■ Department of Archaeology and Center for Remote Sensing, Boston University, 675 Commonwealth Avenue, Boston, MA 02215

Miriam T. Stark ■ Department of Anthropology, University of Hawaii, 2424 Maile Way, Porteus 346, Honolulu, HI 96822

William A. Longacre ■ Department of Anthropology, University of Arizona, Tucson, AZ 85721

American Antiquity, 61(1), 1996, pp. 116–126.
Copyright © by the Society for American Archaeology

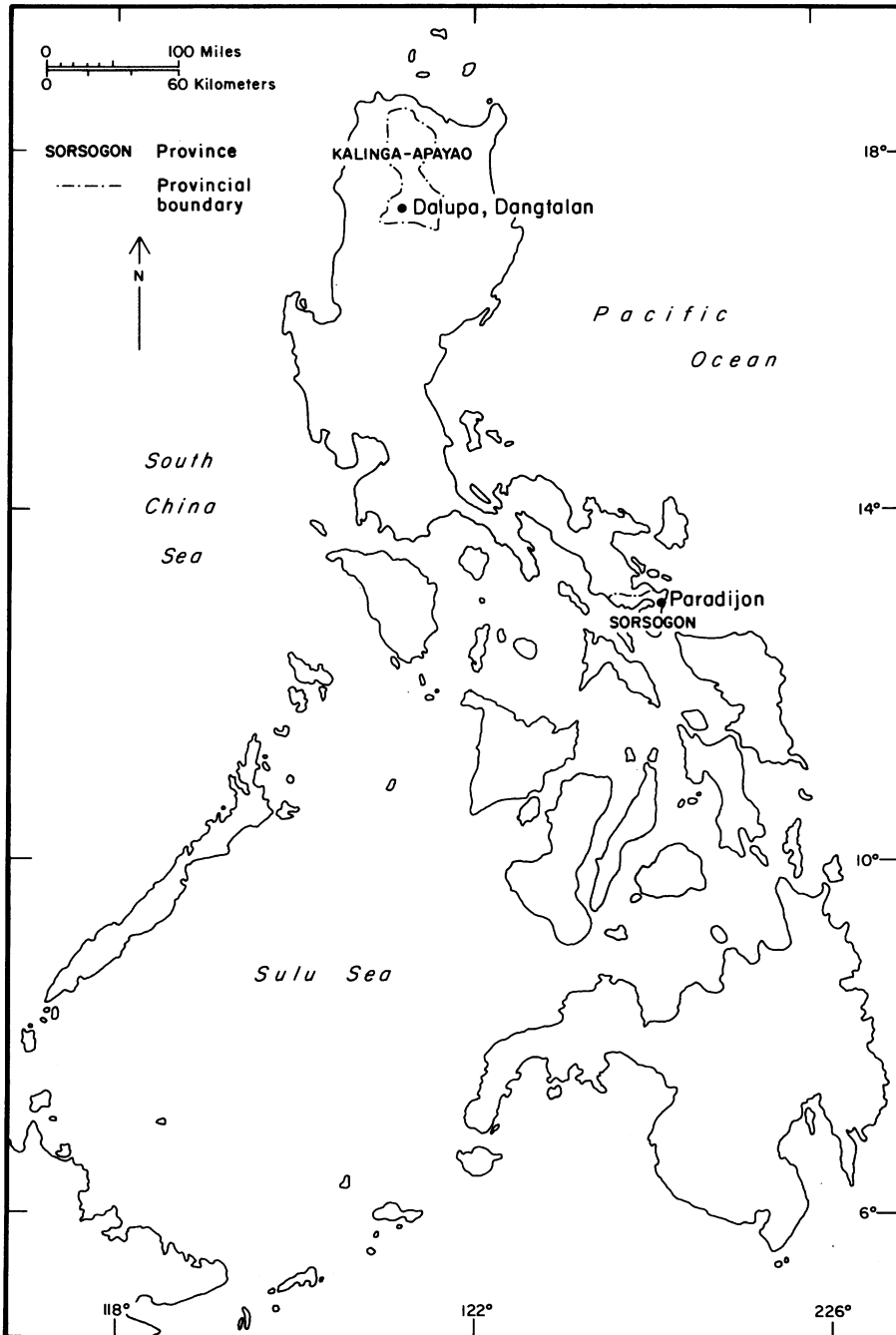


Figure 1. The Philippines with study areas highlighted.

problematic because they depend on a normality assumption. Although many statistical tests based on this assumption are relatively robust in the face of mild departures from normality, this tends *not* to be the case with tests concerned with variance (Brown and Forsythe 1974; Conover et

al. 1981; Miller 1986). While quantitative studies of the standardization issue have been undertaken for some time by a host of researchers, little progress has been made in the statistical methods used to assess this phenomenon in ceramic assemblages.

Today, many alternative statistical tests exist for examining homogeneity of variance, and there are new procedures that offer robust confidence intervals for population variances and standard deviations. In fact, as long ago as 1981 Conover et al. published a comparative study that examined nearly 60 tests for equality of variance. Importantly, several of these tests are quite robust and perform well even when the data illustrate marked deviations from normality, a situation which we will show is the rule rather than the exception in one data set.

In this paper, as with our previous work (Longacre et al. 1988; M. Stark 1991, 1992), we return to the Philippines and, in light of the foregoing, provide examples of alternative statistical procedures by reexamining evidence for ceramic standardization (or the lack of it) in several pottery making communities. We use two Kalinga communities in northern Luzon—Dangtalan and Dalupa—and one in southern Luzon in Sorsogon Province known as Paradijon. In many ways the Philippines data are ideal for this because of their quality (derived from ethnographic populations under controlled conditions) and the sizes of the samples.

Given significant departures from normality, which can be determined through statistical testing, it is appropriate to employ robust statistical methods. For confidence intervals for the population standard deviation we employ jackknifed standard deviations that yield more realistic limits in comparison to the normal-theory-based chi-square approach (Miller 1968; Mosteller and Tukey 1977:139). For a robust statistical test to compare sample variances several excellent candidates are available (Conover et al. 1981). We employ here a test described by Brown and Forsythe (1974) that generally performs better than other available tests (Conover et al. 1981; O'Brien 1978), and is relatively easy to implement. The Brown-Forsythe test, when compared with the standard *F*-test on the same data, often yields markedly different results under non-normality, underscoring the need for a careful and thoughtful approach to research and data analysis.

Differences in Organizational Scales: Three Philippine Communities

Crosscultural research has demonstrated that ceramic production varies widely in the ethno-

graphic record; previous researchers have characterized this variability in terms of organizational modes (Peacock 1982; van der Leeuw 1977) or parameters (Costin 1991; Pool 1992) of production. Philippine pottery systems also vary widely in organizational mode, from household production systems (Longacre 1981) to market-oriented industries (e.g., London 1991). The locations of each of our three study communities, two in northern Luzon and one in southern Luzon, are shown in Figure 1.

The three Philippine communities differ in their organizational modes of ceramic production. The Kalinga village of Dangtalan lies at the "simple" end of the continuum: Dangtalan pottery is made primarily for household use and for restricted exchange (Longacre 1981). In Dalupa, a second Kalinga village, part-time potter specialists exchange their pots in an extensive, nonmarket-based distributional network to supplement farming returns (M. Stark 1991, 1992). In Paradijon (Sorsogon Province), full-time specialists sell their goods to shopkeepers locally and in the provincial capital 16 km distant (London 1991; Longacre et al. 1988:102).

These three Philippine communities share both construction techniques and emic conceptions of vessel size categories. All three communities of potters employ a combination of coil-and-scrape and paddle-and-anvil techniques to fashion their cooking pots (e.g., Longacre 1981, Longacre et al. 1988). The turntable (*bayangan*) that Paradijon potters use differs only slightly from the wooden platters (*chuyas*) that Kalinga potters employ in the initial phases of manufacture. Specific size classifications are based on vessel volume in all three communities, reckoned through the *chupa*, which is generally the volume of a can of evaporated milk (ca. 100 ml). The *chupa* provides a pan-Philippines system of measurement: in all three communities pots are manufactured according to *chupa* sizes. One is able to order a two-, three-, or four-*chupa* pot, for example. Although the *chupa* measurement is relatively constant throughout the Philippines, *chupa*-based size class assignments (e.g., "small pot," "medium pot") vary from one community to the next.

Data Collection

Data from the three communities were collected in

field seasons between 1980 and 1990. Dangtalan data were recorded from household inventories (only vessels recorded as obtained after 1980 were included) and from vessels made in 1987. Dalupa data were collected on newly fired vessels throughout 1988. Paradijon data were obtained on newly manufactured vessels during 1980 and 1981. The vessels measured were either part of households' stored assemblages (Dangtalan), or collections that were intended for exchange or market distribution (Dalupa, Paradijon). Information was collected on each vessel's producer and functional class, and quantitative data were collected to assess vessel dimensions. The measurements were made with a measuring tape to obtain data on aperture diameter, vessel height, and maximum circumference. (Paradijon width, rather than circumference, measurements were recorded. These data were converted into circumference measurements for comparability with other communities with the following formula: circumference = $\pi \times$ width, where $\pi \approx 3.14159$. We recognize that this may affect somewhat the apparent variability of this data set.)

Ethnoarchaeologists are often better equipped than are prehistoric archaeologists to control factors that introduce variability into ceramic assemblages. These include vessel function, vessel size, and producer. Since statistics that describe variation are dependent on how broadly or narrowly the ceramic classes are defined, ethnoarchaeological data sets should generally provide a more reliable means for examining a variety of questions related to ceramic standardization and specialization. This is shown quite clearly in Longacre et al. (1988) where the same data from the Philippines were treated as if they were archaeological, without knowledge of emic size categories. Based on the measured data alone, on which archaeologists must rely, it was shown that various pot size groupings might easily arise. Consequently, variance statistics could conceivably vary widely depending on how such groupings are defined. Nevertheless, there certainly is no shortage of standardization studies based on archaeological collections (e.g., Allen 1992; Benco 1988). In archaeological settings with well-controlled ceramic wares and types, vessel functions, dates, and contexts, meaningful conclusions might possibly be obtained.

This study restricts its focus to the functional

class that is the most frequently used earthenware vessel in all three communities: meat and vegetable cooking pots. Unambiguous size classes for the data sets were defined by *chupa* measurement. We focus on one data set of three-*chupa* pots and strengthen comparability in the multigroup analyses by focusing on pots only in the two-*chupa* size class. Our total sample size is 1,068 vessels.

Alternative Techniques for Assessing Ceramic Standardization

Most ceramic standardization studies that statistically test for the equality of variance generally compute a simple ratio of two sample variances, which is compared against an *F*-distribution with appropriate degrees of freedom (Arnold and Nieves 1992; Blackman et al. 1993; M. Stark 1991). The basic problem with this *F*-test, as well as with confidence intervals based on the chi-square distribution, is that they are sensitive to departures from a normality assumption. In fact, these procedures are not just sensitive, they react extremely to even mild departures from normality, with generally disastrous consequences as will be shown below through computer simulation (Brown and Forsythe 1974; Conover et al. 1981; Miller 1968). This means that the actual significance of obtained results can be several times larger or smaller than the nominal level, for example. Archaeological studies of ceramic standardization therefore need to test for normality and, if necessary, employ procedures that are robust against this assumption.

Testing for Normality

For this exercise we employ an excellent data set from Dalupa, focusing on three-*chupa* pots used for cooking meat and vegetables (*oppaya*). Here an unusually large sample of 725 vessels is available. We first examine histograms for two of our three variables: aperture diameter and circumference (Figure 2a, b). Visual inspection of these histograms indicates that the distributions are bell shaped. However, some nonsymmetry is visible and note, in particular, the straggling right tail on the circumference variable.

Probability plots are a useful means for graphically examining departures from normality, where any deviation of the data from a straight line suggests a non-normal distribution. Such a

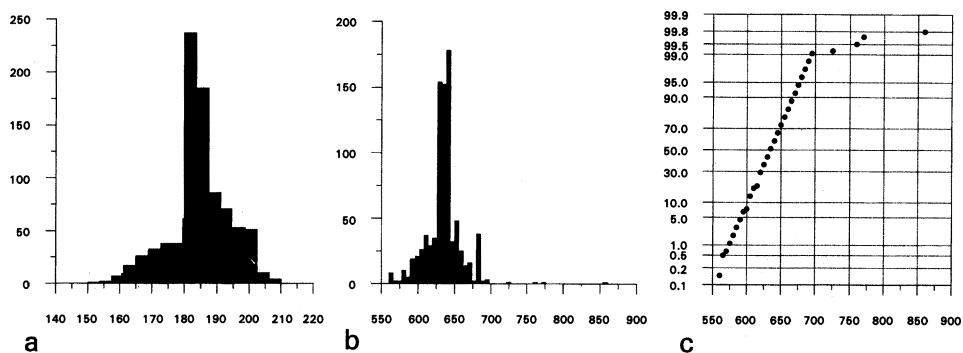


Figure 2. Histograms of (a) aperture diameter and (b) circumference in three-*chupa* meat/vegetable cooking pots from Dalupa ($n = 725$), and a probability plot (c) of circumference. Measurements are in millimeters.

plot for the Dalupa circumference variable is given in Figure 2c, which suggests that the data may not be normal. Further evidence comes from the sample skewness and kurtosis statistics, which deviate from values expected for a Gaussian distribution (Table 1).

We employ the Shapiro-Wilk (1965) statistic, W , a common and robust test for normality, to verify that the three-*chupa* data distributions are not normal (Table 1). In all cases the null hypothesis of a normal population is rejected at the $\alpha = .0001$ level. In view of this finding we can assume that normal-theory-based confidence intervals or statistical tests are inappropriate for these data. Later sections will show that most of our other Philippines data sets are not normally distributed either.

Confidence Intervals for the Standard Deviation

Tukey's jackknife method is a general and robust technique that can be employed to provide parameter estimates and confidence intervals (Mosteller and Tukey 1977; Tukey 1958). The importance of this method in the present context is best described by Miller (1968:567): "Its

approximate confidence intervals are a godsend in problems where messy distribution theory prohibits the formation of exact confidence limits."

The basic idea behind the jackknife technique is that for each of the n observations, that observation is temporarily eliminated and the remaining $n-1$ are used to compute the statistic of interest, here the sample standard deviation. This produces n estimates that are linearly combined with the actual estimate (based on all n cases) to yield a sampling distribution of what are termed "pseudo-values." A robust estimate of σ (the population standard deviation) is simply the average of these pseudo-values (see Kvamme 1988:396 for a partially worked illustration of the jackknife in an archaeological context). Additionally, confidence intervals for σ may be derived through use of the t -distribution and the empirical standard error of the pseudo-values (Mosteller and Tukey 1977:139-140).

The foregoing statistics are illustrated in Table 2 for the Dalupa circumference data where they are compared with the conventional estimator of σ , the sample standard deviation of all n cases, and the chi-square derived confidence limits.

Table 1. Dalupa Descriptive Statistics and Normality Tests on Three-*chupa* Pots.

	Skewness	Kurtosis	W	p
Aperture	-.044	3.787	.933	<.0001
Circumference	1.076	12.530	.921	<.0001
Height	.166	5.265	.929	<.0001

Note: $n = 725$. For normally distributed data, skewness = 0, kurtosis = 3, and the W test statistic is unity.

Table 2. Confidence Intervals of 95 Percent Derived from Normal Theory (Chi-square) and the Jackknife Method.

Derivation	Standard Deviation	Lower Limit	Upper Limit	Width
Chi-Square	25.9	24.7	27.3	2.6
Jackknife (raw)	26.0	22.7	29.2	6.5
Jackknife (log)	26.0	22.9	29.6	6.7

Notes: $n = 725$ three-*chupa* pots from Dalupa: circumference data. Measurements are in millimeters.

Although the jackknifed estimate of σ is about the same as the conventional estimate (this need not be the case), the resultant confidence interval is substantially wider (in fact, more than twice as wide)! Wider intervals, of course, reduce the certainty we can place in estimates making them less desirable in comparisons with other samples, for example. The end result is a much more conservative perspective on findings.

Mosteller and Tukey (1977:141) argue that working with the logarithms of the pseudo-values is a superior tactic because the resultant sampling distribution tends to be better behaved (more symmetrical with less straggling tails, allowing closer approximation by the t -distribution). This outcome, too, is shown in Table 2 where a similarly wide interval is achieved. Because the jackknife is robust and bias reducing, we can expect that the jackknifed intervals shown in Table 2 are more realistic.

We emphasize that with smaller samples, the usual case, intervals that are even more disparate can be obtained between these methods under conditions of non-normality. This means that a 95 percent interval based on the chi-square distribution might really be a 70 percent interval or worse, for example. The jackknifed interval will presumably perform closer to the specified level: i.e., a 95 percent interval should bound the true value of σ 95 percent of the time (in 95 percent of samples).

The Brown-Forsythe Test for Homogeneity of Variances

The Brown-Forsythe test, hereafter BF-test, is a robust procedure for statistically evaluating the equality of variances between two or more samples (Brown and Forsythe 1974). In addition to its ability to yield approximately correct inferences under conditions of non-normality, in contrast to the standard F -ratio of sample variances, the BF-test is a multigroup test that potentially frees us from the latter's two-group limitation.

Alternative tests for homogeneity of variance typically are based on robust estimators of dispersion, such as jackknifed variances (Miller 1968), or of location, such as a trimmed mean or sample median (Levene 1960). The BF-test employs the latter tactic where the data are first transformed, or standardized, by obtaining the absolute deviation

from the class sample median (M_j) for each observation (x_{ij}) within each class: $z_{ij} = |x_{ij} - M_j|$. Obviously, more variable classes will tend to possess higher average values of z than less variable classes. The BF-test focuses on these class differences in the transformed z -values by performing a one-way analysis of variance between the groups under study yielding:

$$B = \frac{\sum_j n_j(z_j - z_*)^2 / (k - 1)}{\sum_i \sum_j (z_{ij} - z_j)^2 / (N - k)}$$

where z_j is the mean z -value for the j th class, z_* is the grand mean over all cases, n_j is the j th class sample size, and N is the size of the combined samples for the k classes. As a ratio of two estimates of variance (between-class and within-class), B follows an F -distribution with $(k-1)$ and $(N-k)$ degrees of freedom (Brown and Forsythe 1974).

Performance of the Brown-Forsythe Test

Several studies have shown, through computer simulation, that the BF-test generally performs much better than other available tests for homogeneity of variance. Brown and Forsythe (1974), in strictly two-group comparisons, demonstrate the reliability of their test under asymmetric, long-tailed, and Gaussian distributional forms. In other words, rejection rates were very close to nominal levels in each of the simulations. Conover et al. (1981), in a comparison of nearly 60 tests for equality of variance including the BF-test presented here, found several that performed at expected levels under a variety of distributions and sample size scenarios in multigroup contexts. Importantly, the BF-test was consistently at or near the top of their list in terms of stability of error rates and power.

To further investigate the performance of the BF-test under conditions of normality and non-normality, and to clearly demonstrate the fragility of the standard F -test in the latter context, Monte Carlo simulation techniques are again employed in this study. The computer simulations are performed under a variety of sample size configurations using pseudo-random numbers generated from the normal (using a sum of uniform deviates algorithm [Kvamme 1993]), the uniform, and log-normal (a heavily skewed distribution that is nor-

Table 3. Simulation Results for the F -test and BF-test under Various Distributional Scenarios.

	F -test			BF-test		
$n = n_2 = 10$						
Nominal α level:	.10	.05	.01	.10	.05	.01
Normal	.088	.057	.010	.083	.046	.008
Uniform	.027	.009	.000	.078	.033	.007
Lognormal	.303	.252	.166	.096	.043	.007
$n_1 = 10, n_2 = 20$						
Nominal α level:	.10	.05	.01	.10	.05	.01
Normal	.088	.045	.012	.102	.053	.012
Uniform	.028	.012	.000	.076	.045	.007
Lognormal	.259	.221	.151	.105	.057	.015

Note: All data are based on 1,000 replications.

mal only in logarithmic form) distributions. Equal samples of $n_1 = n_2 = 10$ are explored, followed by unequal samples of $n_1 = 10$ and $n_2 = 20$, each obtained from the same distribution under a null hypothesis of $H_0 : \sigma_1^2 = \sigma_2^2$. A total of $N = 1,000$ trials is utilized for each run.

In the Gaussian simulation of the F -test, for example, 10 pseudo-random normal deviates were obtained for the first sample, 10 for the second sample (from the same normal population), and the respective sample variances and F -ratio were computed. This was repeated 1,000 times yielding a sampling distribution of 1,000 F -ratios. Under the null hypothesis, and sampling from normal populations, the resultant distribution should closely parallel a theoretical F -distribution with nine and nine degrees of freedom (Hays 1988:331). The theoretical $F_{9,9}$ indicates that approximately .05 of the distribution possesses values beyond 3.18; the computer simulation yielded .057 (Table 3) showing that this test *does* perform as expected under normality (note that the theoretical standard error for $N = 1,000$ trials can be closely approximated by $[p(1-p)/1,000]^{1/2}$; for the five percent level this is roughly .7 percent). Identical procedures were performed under the other distributional forms and sample size schemes for this test and the BF-test (Table 3).

The results of the simulations as demonstrated in Table 3 indicate that while the standard F -test performs well under normality, the outcome is poor when sampling from non-normal populations, as expected. This is particularly true when the parent distributions are heavily skewed, as when sampling from the lognormal distribution.

The BF-test, however, performs remarkably well, very close to nominal levels in all cases, even when the sampled populations are highly skewed. The results also show that the BF-test may be more or less conservative than the conventional F -test, depending on the form of the parent populations. While paralleling the findings of previous simulation work, these results demonstrate the trustworthiness of the BF-test in these contexts and the fallibility of the simple F -test.

Applications to Philippines Data

Armed with some confidence in the BF-test, we now provide examples of its use through a reexamination of the Philippines ceramic data. We employ two-*chupa* meat and vegetable cooking pot data from three villages in this application and use aperture data as an example. Visual inspection of the histograms for Dangtalan, Dalupa, and Paradijon alone suggests differing degrees of variability (Figure 3). Data from Dangtalan (the least specialized potters) seem clearly more dispersed than are data from either Dalupa (part-time specialist potters) or Paradijon (full-time specialists), the latter producing the least variable vessels. Box-and-whisker plots, shown above the histograms in Figure 3, provide a means of visual comparison where a continuum from the most variable (Dangtalan) to the least variable (Paradijon) cooking-pot aperture diameters is suggested.

Descriptive statistics for the two-*chupa* samples from three communities are given in Table 4. The sample standard deviations and coefficients of variation suggest differences in variation, and therefore standardization, between these commu-

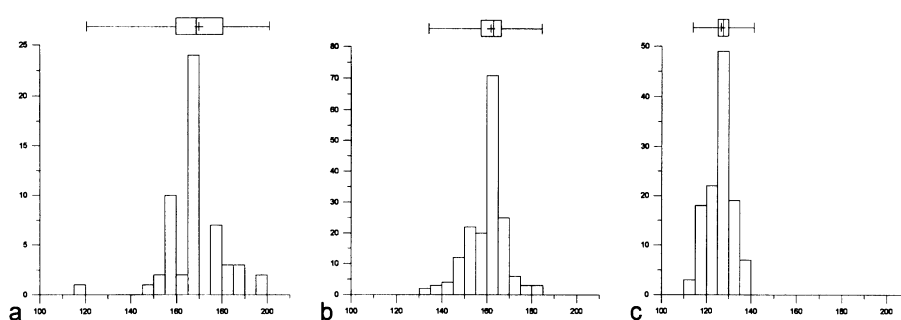


Figure 3. Histograms and box-and-whisker plots of aperture data from (a) Dangtalan, (b) Dalupa, and (c) Paradijon for two-*chupa* meat and vegetable cooking pots. Measurements are in millimeters.

nities. It is therefore appropriate to employ statistical tests to assess the significance of these observed differences. We first note the results of the Shapiro-Wilk (1965) tests for normality, which indicate that non-normality is the general rule in these morphological data. Only two of the nine analyses suggest Gaussian distributions, and one of these is marginal. Use of the standard *F*-test in a statistical assessment of these data would thus be inappropriate, and we instead employ the BF-test to further examine differences in variability in our samples. The results are given in Table 5 (*F*-test results also are given in this table for illustrative and comparison purposes).

The BF-test results show major differences in variation between each village and variable combination, with the possible exception of the height variable where, in two of the comparisons,

Dangtalan-Paradijon and Dalupa-Paradijon, the obtained *p*-values are in that indeterminate area between .05 and .10 that calls for additional study (Table 5). These findings, however, generally parallel the results of earlier work with these data (Longacre et al. 1988).

In comparing findings between the BF-test and the *F*-test (Table 5), note that the critical α 's are strongly different (the critical α is the smallest probability at which the null hypothesis of no difference can be rejected). This means that for any of the tests a different statistical conclusion could be reached depending on the actual decision α selected. In particular, the tests suggest markedly different conclusions for the Dangtalan-Dalupa comparison on circumference (with $p = .056$ and $p = .366$, BF-test given first), and the Dalupa-Paradijon comparisons on aperture ($p = .048$ and

Table 4. Descriptive Statistics and Normality Tests on Two-*chupa* pots from Three Villages.

Variable	Mean	S.D.	C.V.	Skewness	Kurtosis	<i>W/p</i>	Decision
DANGTALAN: Least specialized, <i>n</i> = 55							
Aperture	170.4	12.74	7.47	-.60	6.48	.898/<.0001	Not Normal
Circumference	553.0	34.69	6.27	-.31	3.17	.969/.310	Normal
Height	115.1	9.60	8.34	.16	2.80	.953/.061	Normal ^a
DALUPA: Partly specialized, <i>n</i> = 171							
Aperture	163.0	8.13	4.99	-.38	4.16	.922/<.0001	Not Normal
Circumference ^b	558.9	22.67	4.06	-.53	3.61	.926/<.0001	Not Normal
Height	130.1	7.23	5.56	-.37	3.76	.936/<.0001	Not Normal
PARADIJON: Most specialized, <i>n</i> = 117							
Aperture	128.6	5.83	4.53	-.19	2.67	.904/<.0001	Not Normal
Circumference	608.6	26.18	4.30	-.18	3.16	.929/<.0001	Not Normal
Height	112.8	7.81	6.92	-.42	3.04	.937/<.0001	Not Normal

Notes: S.D. = standard deviation; C.V. = coefficient of variation; *W/p* = *W*-test statistic/significance. For normally distributed data skewness = 0, kurtosis = 3, and the *W*-test statistic is unity.

^aThis result is marginal.

^bOne case omitted due to faulty measurements (*n* = 170).

Table 5. Brown-Forsythe and *F*-ratio of Sample Variances Test Results for Each Village Pair and Variable.

	BF-test	df	<i>p</i>	<i>F</i> -ratio	df	<i>p</i>
DANGTALAN-DALUPA COMPARISON						
Aperture	7.01	1,224	.00869	2.46	54,170	.000123
Circumference*	3.69	1,224	.0560	1.24	170,54	.3660
Height	13.13	1,224	.000360	1.76	54,170	.00682
DANGTALAN-PARADIJON COMPARISON						
Aperture	15.01	1,170	.000152	4.77	54,116	<10 ⁻¹⁰
Circumference	5.87	1,170	.0164	1.75	54,116	.00613
Height	3.25	1,170	.0732	1.51	54,116	.0664
DALUPA-PARADIJON COMPARISON						
Aperture*	3.93	1,286	.0482	1.94	170,116	.000160
Circumference	3.53	1,285	.0614	1.33	116,169	.0436
Height*	2.98	1,286	.0853	1.16	116,170	.365
THREE GROUP ANALYSES: DANGTALAN-DALUPA-PARADIJON						
Aperture	7.65	2,340	.000561			
Circumference	9.11	2,339	.000139			
Height	6.26	2,340	.00214			

Notes: Three-group analyses are given in the bottom group. An asterisk (*) indicates analyses where different conclusions might be reached by the two tests.

$p = .00016$) and height ($p = .085$ and $p = .365$). While for the most part the differences between the tests seem inconsequential in this application, greatly different outcomes could potentially be reached in practice when dealing with other non-normal populations, as was clearly illustrated in the foregoing simulation study (Table 3).

We conclude our examination by demonstrating a BF-test for homogeneity of variance among the three groups taken together. This multigroup capability provides freedom from the confines of strictly two-group comparisons that have been the rule in standardization studies owing to limitations of the *F*-test (see B. Stark 1991, however). With this test we are assessing the significance of differences in variability indicated by the statistics in Table 4 between all three groups simultaneously. As we might expect from the previous results, the three-group analyses suggest pronounced differences in variation between the villages for each of the variables examined (Table 5). The sample statistics (Table 4) and previous analyses (Table 5) point to the nature and source of these differences.

Discussion

We have suggested several alternative methodologies that should strengthen studies of ceramic or other forms of standardization. These include

testing for normality, using robust jackknife techniques for parameter estimation and the construction of confidence intervals, and employing the Brown-Forsythe (1974) test for comparing multiple sample variances. Use of these methods will improve standardization studies when sampling from populations that are not normally distributed, a situation we expect is the rule rather than the exception.

Other refinements in ceramic standardization studies are necessary. These include increasing our analytical control over vessel shape, function, and especially over size classes in archaeological assemblages. Variation in statistical outcomes is closely related to size class definition (Allen 1992:153). For example, in our previous study of Kalinga vessels the use of emic classes ("medium," "large") produced high coefficients of variation (Longacre et al. 1988). This is because each emic size class in Dangtalan and Dalupa encompasses a wide range of *chupa* volumes. When the same data are analyzed with more restrictive size classes, the coefficients of variation naturally become smaller. Ethnoarchaeology therefore provides an important avenue for exploring the extent to which economies of scale affect the degree of product standardization. Unlike fragmentary archaeologi-

cal assemblages (e.g., Benco 1988), ethnographic contexts provide complete vessel information, and aspects of specialization and product standardization can be controlled.

Correlations between standardization and specialization reflect complex networks of factors that elude easy explanation. Differences in raw materials, potters' levels of expertise (London 1981:214), market demand, manufacturing techniques (e.g., Arnold and Nieves 1992), local traditions or customs (e.g., Reina and Hill 1978), and types of measurement aids used to produce vessel types (Arnold 1991:367; Arnold and Nieves 1992) all affect the relative degree of standardization in a ceramic product. Alternative, more effective means of measuring and assessing variability in ceramics can only help to untangle relationships between production intensity and product standardization.

Acknowledgments. The Kalinga data used in this analysis were collected in research supported by a grant from the National Science Foundation (BNS 87-10275) to William A. Longacre. We wish to thank Michael W. Graves and an anonymous reviewer for helpful comments, and Ronald Beckwith who drafted Figure 1. Thanks are also extended to Philippine assistants who assisted in the data collection: Josephine Bommogas, Amy Awing, Narcisa Waggawag, Cristina Tima, Rosalina Busog, and Brenda Dumayag. Computer programs (DOS-based) are available from the first author that produce jackknifed standard deviations and confidence intervals and compute the Brown-Forsythe test for homogeneity of variance.

References Cited

- Allen, K. M. S.
1992 Iroquois Ceramic Production: A Case Study of Household-Level Organization. In *Ceramic Production and Distribution: An Integrated Approach*, edited by G. J. Bey III and C. A. Pool, pp. 133–154. Westview Press, Boulder, Colorado.
- Arnold, D., and A. L. Nieves
1992 Factors Affecting Ceramic Standardization. In *Ceramic Production and Distribution: An Integrated Approach*, edited by G. J. Bey III and C. A. Pool, pp. 113–214. Westview Press, Boulder, Colorado.
- Arnold, P. J., III
1991 Dimensional Standardization and Production Scale in Mesoamerican Ceramics. *Latin American Antiquity* 2:363–370.
- Benco, N.
1988 Morphological Standardization: An Approach to the Study of Craft Specialization. In *A Pot for All Reasons: Ceramic Ecology Revisited*, edited by C. Kolb and L. Lackey, pp. 157–172. Laboratory of Anthropology, Temple University, Philadelphia.
- Blackman, M. J., G. J. Stein, and P. B. Vandiver
1993 The Standardization Hypothesis and Ceramic Mass Production: Technological, Compositional and Metric Indexes of Craft Specialization at Tell Leilan, Syria. *American Antiquity* 58:60–80.
- Brown, M. B., and A. B. Forsythe
1974 Robust Tests for Equality of Variances. *Journal of the American Statistical Association* 69:364–367.
- Brumfiel, E., and T. K. Earle
1987 Specialization, Exchange and Complex Societies: An Introduction. In *Specialization, Exchange and Complex Societies*, edited by E. Brumfiel and T. K. Earle, pp. 1–9. Cambridge University Press, Cambridge.
- Conover, W. J., M. E. Johnson, and M. M. Johnson
1981 A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics* 23:351–361.
- Costin, C. L.
1991 Craft Specialization: Issues in Defining, Documenting, and Explaining the Organization of Production. In *Archaeological Method and Theory*, vol. 3, edited by M. B. Schiffer, pp. 1–56. University of Arizona Press, Tucson.
- Davis, J. L., and H. Lewis
1985 Mechanization of Pottery Production: A Case Study from the Cycladic Islands. In *Prehistoric Production and Exchange: the Aegean and Eastern Mediterranean*, edited by A. B. Knapp and T. Stech, pp. 79–92. Monograph XXV. Institute of Archaeology, University of California, Los Angeles.
- Feinman, G. M., S. Upham, and K. G. Lightfoot
1981 The Production Step Measure: An Ordinal Index of Labor Input in Ceramic Manufacture. *American Antiquity* 46:871–884.
- Hagstrum, M.
1988 Ceramic Production in the Central Andes, Peru: An Archaeological and Ethnographic Comparison. In *A Pot for All Reasons: Ceramic Ecology Revisited*, edited by C. Kolb and L. Lackey, pp. 127–145. Laboratory of Anthropology, Temple University, Philadelphia.
- Hays, W. L.
1988 *Statistics*. 4th ed. Holt, Rinehart and Winston, New York.
- Kramer, C.
1985 Ceramic Ethnoarchaeology. *Annual Review of Anthropology* 14:77–102.
- Kvamme, K. L.
1988 Development and Testing of Quantitative Models. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. J. Judge and L. Sebastian, pp. 325–428. U.S. Department of the Interior, Bureau of Land Management, U.S. Government Printing Office, Washington, D.C.
- 1993 A Simple Pseudo-Random Normal Deviate Generator. *Archaeological Computing Newsletter* 37:8–10.
- Levene, H.
1960 Robust Tests for Equality of Variances. In *Contributions to Probability and Statistics*, edited by I. Olkin, pp. 278–292. Stanford University Press, Palo Alto, California.
- London, G.
1981 *Decoding Designs: The Late Third Millennium B.C.*

- Pottery from Jebel Qa' Aqir*. Unpublished Ph.D. dissertation, Department of Oriental Studies, University of Arizona, Tucson.
- 1991 Standardization and Variation in the Work of Craft Specialists. In *Ceramic Ethnoarchaeology*, edited by W. A. Longacre, pp. 182–204. University of Arizona Press, Tucson.
- Longacre, W. A.
1981 Kalinga Pottery: An Ethnoarchaeological Study. In *Pattern of the Past: Essays in Honor of David Clarke*, edited by I. Hodder, G. Isaac, and N. Hammond, pp. 49–66. Cambridge University Press, Cambridge.
- Longacre, W., K. Kvamme, and M. Kobayashi
1988 Southwestern Pottery Standardization: An Ethnoarchaeological View from the Philippines. *Kiva* 53:101–112.
- Miller, R. G., Jr.
1968 Jackknifing Variances. *The Annals of Mathematical Statistics* 39:567–582.
- Mosteller, F., and J. W. Tukey
1977 *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, Massachusetts.
- O'Brien, R.
1978 Robust Techniques for Testing Heterogeneity of Variance Effects in Factorial Designs. *Psychometrika* 43:327–342.
- Peacock, D. S.
1982 *Pottery in the Roman World: An Ethnoarchaeological Approach*. Longman, London.
- Pool, C.
1992 Interpreting Ceramic Production and Distribution. In *Ceramic Production and Distribution: An Integrated Approach*, edited by G. J. Bey III and C. A. Pool, pp. 275–314. Westview Press, Boulder, Colorado.
- Reina, R. E., and R. M. Hill II
1978 *The Traditional Pottery of Guatemala*. University of Texas Press, Austin.
- Rice, P.
1981 Evolution of Specialized Pottery Production: A Trial Model. *Current Anthropology* 22(3):219–240.
1987 *Pottery Analysis: A Sourcebook*. University of Chicago Press, Chicago.
- 1991 Specialization, Standardization, and Diversity: A Retrospective. In *The Ceramic Legacy of Anna O. Shepard*, edited by R. L. Bishop and F. W. Lange, pp. 257–279. University Press of Colorado, Boulder.
- Shapiro, S., and M. Wilk
1965 An Analysis of Variance Test for Normality. *Biometrika* 52:591–611.
- Sinopoli, C.
1988 The Organization of Craft Production at Vijayanagara, South India. *American Anthropologist* 90:580–597.
- Stark, B.
1991 Standardization and Specialization in Pottery. Paper presented at the 62nd Annual Meeting of the Southwestern Anthropological Association, Tucson, Arizona.
- Stark, M. T.
1991 Ceramic Production and Community Specialization: A Kalinga Ethnoarchaeological Study. *World Archaeology* 23(1):64–78.
1992 From Sibling to Suki: Social Relations and Spatial Proximity in Kalinga Pottery Exchange. *Journal of Anthropological Archaeology* 11:125–136.
- Tosi, M.
1984 The Notion of Craft Specialization and Its Representation in the Archaeological Record of Early States in the Turanian Basin. In *Marxist Perspectives in Archaeology*, edited by M. Spriggs, pp. 22–52. Cambridge University Press, Cambridge.
- Tukey, J. W.
1958 Bias and Confidence in Not-Quite Large Samples. *Annals of Mathematical Statistics* 29:614.
- van der Leeuw, S.
1977 Towards a Study of the Economics of Pottery Making. In *Ex Horreo*, edited by B. L. van Beek, R. W. Brandt and W. Groenman-van Waateringe, pp. 68–76. Institute of Pre- and Proto-History, University of Amsterdam, The Netherlands.

Received March 9, 1994; accepted October 15, 1995.