Replication in Archaeological Science: Shape as a Measure of Weapon Standardisation

Corey Christopherson, Zachary Garcia, Richard Todd, Thomas Winegarden

2/24/2020

## Introduction

The traditional approach to evaluating the level of manufacturing standardization in a society has involved the comparison of standard dimensional measurements of the object through the coefficient of variation (CV) metric [1]. Computed as a simple ratio of the standard deviation to the mean, and scaled by 100 as shown below, the coefficient of variation provides an easily interpretable measure of variation that can be compared to other measurements with different mean values.

However, in the paper *Shape as a measure of weapon standardization: From metric to geometric morphometric analysis of the Iron Age ‘Havor’ lance from Southern Scandinavia* by Birch and Martinon-Torres [-@citationkey], a new approach is proposed. Instead of relying on dimensional comparison of different CVs, they propose using a Geomorphic Morphometric (GMM) analysis as a general measure of overall shape [2].

In this work, we are attempting to confirm that the underlying data provided [2] is the same as that used in the original study by replicating the figure showing the pair-wise correlation of all lance dimensions across the three underlying datasets (original study Fig 8). We are also attempting to replicate the original finding that there was found to be an association between overall centroid size and shape that could not be explained through site difference via an ANOVA analysis.

## Methods

The original analysis was conducted in an unspecified version of R. However, specific versions of geomorph(v3.0.6) and shapes(v1.2.3), two of the most important packages, were directly stated. Our reproduction attempt began with importing the packages explicitly identified in the paper as well as those required in the technical code.

Our next step was to set the working directory. Although a trivial step, the original code used setwd() and a static path string while our reproduction attempt opted to use the more robust here() library as shown below.

Next, we read in and shaped the data obtained from the ScienceDirect archive of the paper using the same code as the original paper, slightly modified to account for our modified approach to setting the working directory.

## [1] 0.5207909  
## [1] 0.5179726  
## [1] 0.5261531  
## [1] 0.5796619  
## [1] 0.466887

With the R session properly configured, correct packages installed, and data loaded and shaped according to the paper and provided R file, we were then able to proceed with our attempt at replicating the correlation matrix of lance dimensions shown in Figure 8 of the original paper. This is where we encountered our first issue, as the code used to create the figure was not included in the R file. Instead a series of five corr() statements (shown below) comparing a subset of lance features were included with the resulting correlation coefficients included in the comments.

Pairwise correlations of lance features cor(socketsock\_bdiam,use=“pairwise.complete.obs”) # 0.52 cor(socketsock\_sdiam,use=“pairwise.complete.obs”) # 0.52 cor(lances\_Mthickness,use=“pairwise.complete.obs”) # 0.53 cor(bladewidth,use=“pairwise.complete.obs”) # 0.58 cor(bladethickness,use=“pairwise.complete.obs”) # 0.47

We then attempted to replicate the figure with the ggpairs() library and perform a visual comparison to Figure 8.

Finally, we attempted to replicate the original finding that there was found to be an association between overall centroid size and shape that could not be explained through site difference via an ANOVA analysis. This test was described on page 44 of the original paper as well as in the R document and relies on the procD.allometry() function from the geomorph package.

## Results

Unfortunately, the pairwise correlation coefficients noted in the comments were not the same as the ones shown in Figure 8. This inconsistiency made it more difficult to confirm the success of our replication attempt.

Luckily, we were able to replicate the finding that there was found to be an association between overall centroid size and shape that could not be explained through site difference via an ANOVA analysis on page 44 of the original paper

### Reproducing Figure 8 and associated analysis

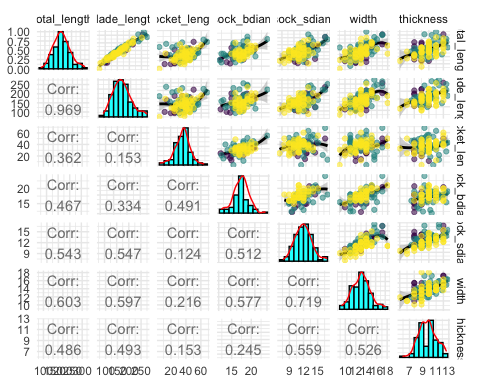
The figure that we are attepting to replicate is Figure 8, captioned as a “Correlation matrix of Havor lance dimensions showing Pearson’s r(lowerpanel), accompanying bivariate plots (upperpanel) and distributions (histograms with kernel density estimation, on the diagonal). Red=Ejsbøl, green=Nydam, blue=Illerup.”

Data appear to be drawn from the main lances\_M dataframe, but the figure labels include “socket (max)” and “socket (min)” which do not have obvious analogues in the column names.

A closer review of the paper leads us to the following line: “Although socket ‘diameters’ were measured, they are referred to here as minimum and maximum socket ‘thickness’ (p.38.” This points us to two fields, which we can test against the summary table that is Table 1:

## # A tibble: 3 x 3  
## site mean n  
## <ord> <dbl> <int>  
## 1 Ejsbol 12.5 15  
## 2 Nydam 12.2 42  
## 3 Illerup 10.9 54  
## # A tibble: 3 x 3  
## site mean n  
## <ord> <dbl> <int>  
## 1 Ejsbol 16.6 12  
## 2 Nydam 17.0 44  
## 3 Illerup 17.9 43  
## # A tibble: 3 x 3  
## site mean n  
## <ord> <dbl> <int>  
## 1 Ejsbol 189 6  
## 2 Nydam 178. 29  
## 3 Illerup 159. 51

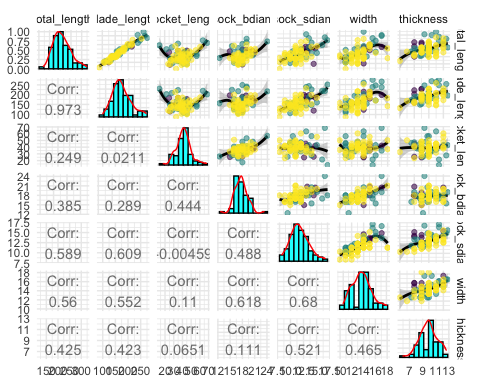
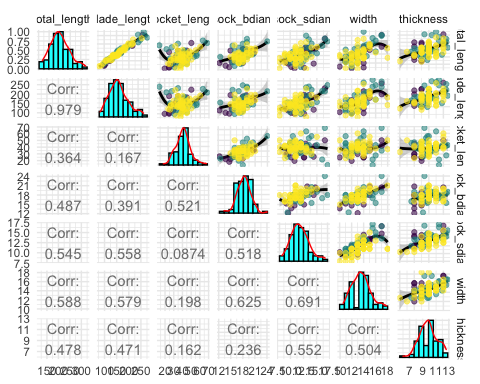
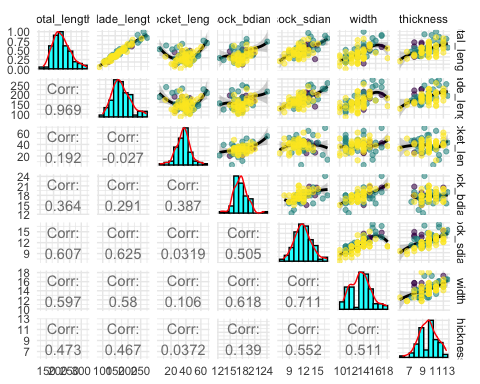
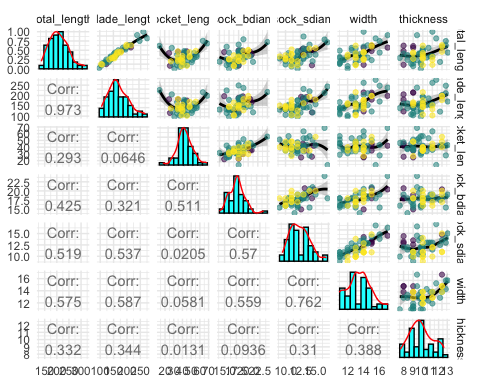
We can be confident therefore that “sock\_bdiam” and “sock\_sdiam” are the sources of “socket (max) [thickness (mm)]” and “socket (min) [thickness (mm)]” respectively. Using some custom functions and the excellent ggally package, we can produce a first replication of the figure.



The correlation values do not exactly match that of the paper. A quick test with the stats passage verifies the results of our plot and confirms that the issue is not a mis-labelling of a spearman coefficient.

## [1] 0.4906856  
## [1] 0.4402349  
## [1] 0.7187674  
## [1] 0.7463491

Below we recreate the figure on different data subsets:



None of the author-defined data subsets (or a manual calculation based on on complete cases) allows us to replicate the correlation results.

### Reproducing ANOVA analysis

As of geomorph version 3.1, the function procD.allmoetry has been deprecated (see <https://rdrr.io/cran/geomorph/man/procD.allometry.html>). This function underpinned the ANOVA tests provided in the technical code appendix. Thankfully, the specificity of the authors’ instructions allowed a previous version of geomorph to be installed!

We replicate the results of p.44 on the ANOVA tests used to establish whether there is a significant relationship between centroid size and shape.

##   
## Homogeneity of Slopes Test  
##   
## Allometry Model  
##   
## Call:  
## procD.allometry(f1 = shape ~ size, f2 = ~site, print.progress = FALSE,   
## data = lances\_geomorph, method = "PredLine")   
##   
##   
##   
## Homogeneity of Slopes Test  
## Df RSS SS Rsq F Z Pr(>F)  
## Common Allometry 74 0.76683   
## Group Allometries 72 0.74137 0.025466 0.030754 1.2366 0.73767 0.234  
##   
## The null hypothesis of parallel slopes is supported  
## based on a significance criterion of alpha = 0.05   
##   
## Based on the results of this test, the following ANOVA table is most appropriate  
##   
## Type I (Sequential) Sums of Squares and Cross-products  
## Randomized Residual Permutation Procedure Used  
## 1000 Permutations  
## ANOVA effect sizes and P-values based on empirical F distributions  
##   
##   
## Df SS MS Rsq F Z Pr(>F)   
## log(size) 1 0.05502 0.055024 0.06645 5.3098 1.87975 0.012 \*  
## site 2 0.00619 0.003096 0.00748 0.2987 -0.83913 0.812   
## Residuals 74 0.76683 0.010363 0.92607   
## Total 77 0.82805   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The results match. We note that no seed is specified in the provided code, making the random resampling approach used volatile.

We can explore the effect of other parameters in the allometry ANOVA, to test whether the core conclusion holds. In the example below, we do not log-transform the size variable and we set a significance threshold (alpha) of 0.1, rather than the default 0.05.

##   
## Homogeneity of Slopes Test  
##   
## Allometry Model  
##   
## Call:  
## procD.allometry(f1 = shape ~ size, f2 = ~site, logsz = FALSE,   
## seed = 42, alpha = 0.1, print.progress = FALSE, data = lances\_geomorph,   
## method = "PredLine")   
##   
##   
##   
## Homogeneity of Slopes Test  
## Df RSS SS Rsq F Z Pr(>F)  
## Common Allometry 74 0.77288   
## Group Allometries 72 0.74360 0.029283 0.035364 1.4177 0.85666 0.205  
##   
## The null hypothesis of parallel slopes is supported  
## based on a significance criterion of alpha = 0.1   
##   
## Based on the results of this test, the following ANOVA table is most appropriate  
##   
## Type I (Sequential) Sums of Squares and Cross-products  
## Randomized Residual Permutation Procedure Used  
## 1000 Permutations  
## ANOVA effect sizes and P-values based on empirical F distributions  
##   
##   
## Df SS MS Rsq F Z Pr(>F)   
## size 1 0.04777 0.047775 0.05770 4.5742 1.72764 0.026 \*  
## site 2 0.00739 0.003695 0.00892 0.3538 -0.70587 0.768   
## Residuals 74 0.77288 0.010444 0.93338   
## Total 77 0.82805   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The core result holds! We can push these boundaries further; removing residual randomization and using Cohen’s f-squared values as the random distribution to estimate effect size.

##   
## Homogeneity of Slopes Test  
##   
## Allometry Model  
##   
## Call:  
## procD.allometry(f1 = shape ~ size, f2 = ~site, logsz = FALSE,   
## seed = 42, alpha = 0.1, RRPP = FALSE, effect.type = "cohen",   
## print.progress = FALSE, data = lances\_geomorph, method = "PredLine")   
##   
##   
##   
## Homogeneity of Slopes Test  
## Df RSS SS Rsq F Z Pr(>F)  
## Common Allometry 74 0.77288   
## Group Allometries 72 0.74360 0.029283 0.035364 1.4177 0.85666 0.205  
##   
## The null hypothesis of parallel slopes is supported  
## based on a significance criterion of alpha = 0.1   
##   
## Based on the results of this test, the following ANOVA table is most appropriate  
##   
## Type I (Sequential) Sums of Squares and Cross-products  
## Randomization of Raw Values used  
## 1000 Permutations  
## ANOVA effect sizes and P-values based on empirical Cohen f-squared distributions  
##   
##   
## Df SS MS Rsq F Z Pr(>Cohen f-sq)   
## size 1 0.04777 0.047775 0.05770 4.5742 1.74380 0.026 \*  
## site 2 0.00739 0.003695 0.00892 0.3538 -0.83073 0.802   
## Residuals 74 0.77288 0.010444 0.93338   
## Total 77 0.82805   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The core result still holds.

## Conclusion

Our attempt at a replication of two parts of the paper *Shape as a measure of weapon standardization: From metric to geometric morphometric analysis of the Iron Age ‘Havor’ lance from Southern Scandinavia* by Birch and Martinon-Torres [2] was an overall success despite some missing documentation and inconsistencies between the paper and provided code.

Specifically, the replication attempt of the figure showing the pair-wise correlation of all lance dimensions across the three underlying datasets (original study Fig 8) was used to confirm that the underlying data provided by the authors was the same as that used in the analysis. This replication attempt was complicated by the lack of the code used to generate the plot in the provided R file. Additionally, the correlation coefficients noted in the R file comments did not match the figure in the original paper. We were able to create a similar figure to the one in the original paper; but without a clear expected outcome, we were unable to confirm that the provided dataset was the same one used in the analysis.

The second part of the original paper we attempted to replicate was the original finding that there was found to be an association between overall centroid size and shape that could not be explained through site difference via an ANOVA analysis. The code used to perform this analysis was included in the R file, and the expected results noted in the comments matched the original paper. This made it much easier to evaluate the accuracy of our replication. We ran the same analysis as noted in the provided R code and obtained the same results.

The combination of these two replication efforts combined to make this a successful replication attempt. Although our initial attempt cast doubt on the underlying data, the fact that the second replication matched the expected result suggests that the problem with the first attempt was a transcription error between code and paper and not an indication of bad underlying data.

## References

1. Li, X.J., Bevan, A.H., Martinón-Torres, M., Rehren, T., Cao, W., Xia, Y., Zhao, K., 2014. Crossbows and imperial craft organisation: the bronze triggers of China’s Terracotta Army. Antiquity 88, 126–140.
2. Birch, T., & Martinon-Torres, M. (2018). Shape as a measure of weapon standardisation: From metric to geometric morphometric analysis of the Iron Age ‘Havor’ lance from Southern Scandinavia. Journal of Archaeological Science, 101, 34-51. doi: <https://doi.org/10.1016/j.jas.2018.11.002>