A statistical study of Dressel 20 amphorae rim morphometry from the Baetica province of the Roman Empire

Contents

ABSTRACT

Ceramics are one of the most long-lasting archaeological materials, and thus their study is a dominant factor in reconstructing the social and economic trends of an archaeological site. In this report, the morphometric dimensional values of rim sherds have been analysed, to study potential levels of standardization of production methodology. For this analysis, Dressel 20 transport amphorae from the Baetica region (Spanish Andulasia today) of the former North Western Roman Empire has been considered. This particular type of amphorae were a pivotal transport vessel used in the olive oil trade of the Baetica province.

Through statistical analysis, this paper shows that there was homogeneity in means of external and internal diameters, and subsequently mean shape, across different sites in the concerned region. This points to a consequent homogeneity in production technology of these transport vessels. The study is a starting point, which can lead to further research into connectivity and transfer of knowledge between production sites of transport amphorae.

INTRODUCTION

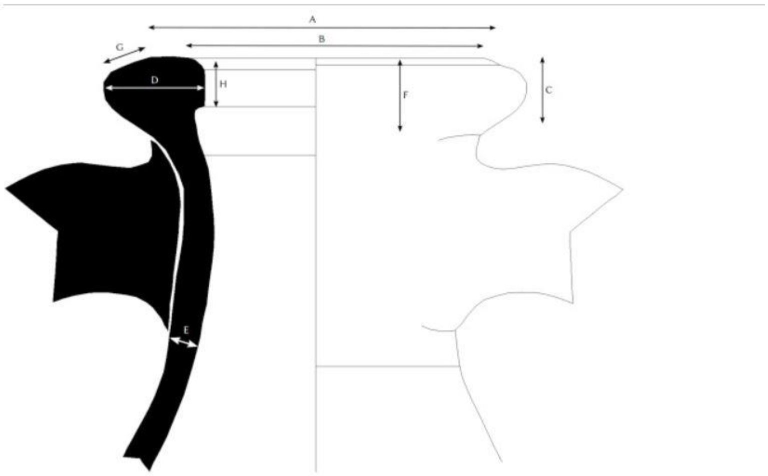
The Baetica region was one of three provinces in the region of 'Hispania' i.e. the Iberian Peninsula region within the former Western Roman Empire. Geographically, region coincides greatly with the modern day Spanish Andalusia. The area has held much interest due to the key role it played in the production and transportation of olive within the regions of the Roman empire and its frontier empires in North Western Europe. The presence of the Dressel 20 type amphorae has been dated to the Julio-Claudian era until around the third quarter of the 3rd c. A.D (Bourgeon 2017). The site of Monta Testaccio, which has been pivotal for the study of the Baetica olive oil trade revealed the presence of almost completely Baetican amphorae within its material culture ((BlÁzquez 1993).

Morphometrics have long been proven useful in the study of ceramics, in terms of typological classification. To put it aptly - "Morphometrics allows the calculation of an average shape, an idealised centroid shape, which takes into account all the individuals in the group, and not merely a single element, supposed to be the best representation, as often used in typology."(Wilczek et al. 2014, 40). Since the morphometrics are directly linked to the mean shape, a study of diversity, or lack thereof, of different morphometric values within a group, would give an indication of the homogeneity of shape. This can lead to determining level of standardization and transfer of production technology within various groups.

The main theme of this project is based on this very concept. A dataset of 413 Dressel 20 amphorae rim sherds, from the sites of Belen, Delicias, Malpica, Parlamento and Villaseca, from the former Baetica province, has been statistically analysed based on some morphometric dimensions. A study of mean behavior of these values can help us trace either morphometric similarity or dissimilarity between the samples. Through this we can make deductions about production technologies of these transport vessels – more specifically, whether they were homogenous across regions.

The dataset analysed consists of three qualitative values & 8 quantitative values, each a different morphometric measurement of the rims found. Of these, only a few have been used for this analysis. This includes three categorical variables – type, consisting of three distinct types of Dressel 20 amphorae (Dressel C, D and E), and two quantitative variables – the interior and exterior diameter of the rim sherds. The morphometric values recorded in the original dataset can be seen from Figure 1.

Through statistical analysis of quantified morphological data of rim sherds, to what extent can we deduce that there was homogenity in production techniques of different varieties of Dressel 20 transport amphorae, used for olive oil trade from the Baetica region, across different sites in the area? Through this hypothesis, can we conclude that there was a level of standardization in the transport amphorae, within the concerned region?



Download : Download high-res image (149KB)   Download : Download full-size image

Fig. 2. The 8 morphometric measurements taken for all amphorae. A: External diameter. B: Inside diameter. C: Rim height. D: Rim width. E: Shape width. F: Rim inside height. G: Rim width 2. H: Protruding rim.

Figure 1: Morphometric dimensions of a Dressel 20 amphorae (Coto-Sarmiento, Rubio-Campillo, and Remesal 2018)

STATISTICAL ANALYSIS

A. Basic Workflow Explanation & Blueprint

Preliminary Data Cleaning: Before the data was analyzed, the data was ‘cleaned’ such that it only contained the values being used for the various analyses. Thus, the levels Dressel 23, Dressel B and Dressel G were dropped from the factor ‘type’. Dressel 23 was dropped since the analysis was solely about the Dressel 20 amphorae, while the other two types were discluded due to their lack of presence across all of the five selected sites and their low abundance in the dataset.

Exploratory Data Analysis: This included some visualizations of counts of the amphorae types by chronology and excavation. The results have been discussed further in Part B of this section.

Hypothesis Testing: To start with, two Chi-square tests were conducted between the categorical variables i.e. between type and chronology, and type and excavation. The first test was a result of the respective visualization done in the previous step between the two variables. The second test was conducted to check whether there was a statistical association between the distribution of amphorae types with each particular site. Since the null hypothesis proved correct, a number of ANOVA tests were conducted in order to check for similarities in means of the exterior and interior diameter dimensions, across the sites, for each type.

The ANOVA tests were done on the basis of each of the three amphorae types, therefore the data was first subset into three smaller datasets on this basis (Dressel C, D and E). The ANOVA test was then conducted within these datasets, for exterior and interior dimensions. In each case, the following steps were followed:

1. **Visualization**: This was done in the form of a boxplot, wherein the quantitative measurements were visualized on one plot for all 5 sites. This was done in order to have a quick preliminary understanding of the data distribution, level of difference in sample sizes for each site and also somewhat identify behavior in terms of normality and outliers.
2. **Normality check**: The assumption of normality is an important one to conduct the ANOVA F-Test. This assumption has been tested on the basis of a combination of methods and reasoning, rather than a direct test result of normality. To start with a, Q-Q plot is generated, using a linear model of the residuals and a Shapiro Wilk test conducted.
   - If the null hypothesis of normality in the Shapiro-Wilk test could not be rejected: the data is considered normal. The distribution of the Q-Q plot was also considered. The homogeneity of variance was checked (second assumption of ANOVA), using Levene's test. For a non-significant difference between variances, the parametric ANOVA F-Test was conducted, and for a significant difference between variances, the non-parametric Kruskal-Willis test was conducted.
   - If the null hypothesis of normality in the Shapiro-Wilk test was rejected: the skewness level was calculated, and the sample size and Q-Q plot was considered. If based on these factors, normality could be 'assumed' without causing too many Type I errors, the parametric test was conducted to test the hypothesis, else the Kruskal Willis.

     Note: Data transformations were not applied to induce normality, since there were different sample sizes of each of the sites, and a hypothesis testing of transformed data could give skewed results.

3. **Parametric Test:** If the results of the parametric test shows a failure to reject the null hypothesis, a similarity in the means is presumed. On the other hand, if the null hypothesis is rejected, a further analysis is done to determine the points of difference.
4. **Outliers handling**: When the hypothesis of data normality is rejected, the outliers have also been observed, with a separate focus on extreme values. The decision to leave in the outliers or removing them for the dataset was based on individual factors and behaviour in each case. This has been explained further in the individual analysis of each of the amphorae types in subsection C.2.

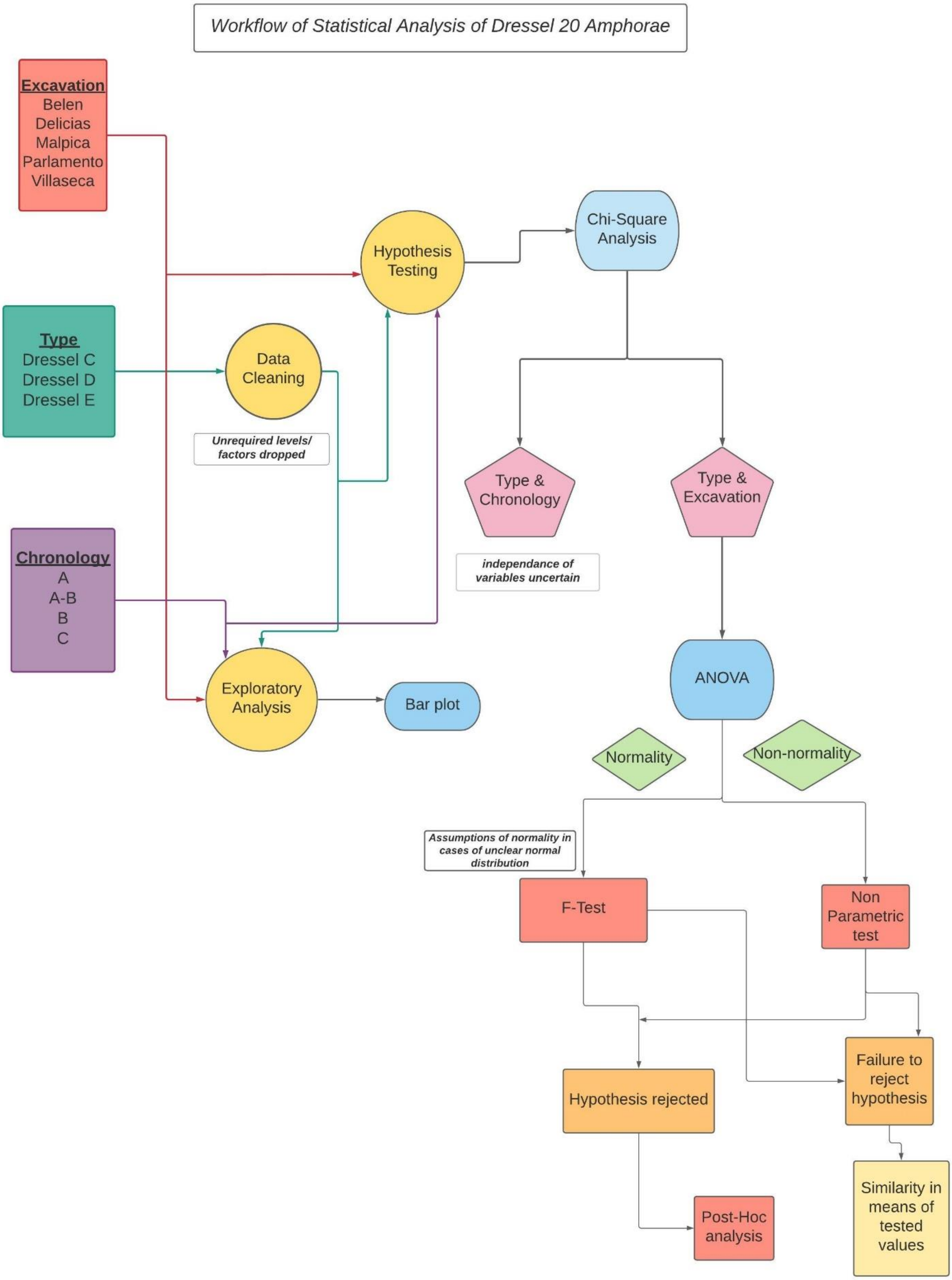


Figure 2: Workflow of statistical analysis

B. Exploratory Data Analysis

A preliminary look at the data consisted mainly of basic visualizations using bar graphs.

Bar graphs: The main research theme of this project is based on production technology of the different amphorae types within each site, and subsequently how they may be related. Therefore, a graph was generated for count of each amphorae type per site (seen in Figure 2). A table was also generated showing the actual numeric frequency (Table 1).

The chronology used in the dataset has not been defined properly in either the metadata or the associated research paper (Coto-Sarmiento, Rubio-Campillo, and Remesal 2018). A bar graph was plotted of the variables ‘type’ and ‘chronology’ to see how if there was a connection (Figure 3). As per the results, there seems to be an almost direct link between the two, in terms of there being a specific type of amphorae used per chronology. This could have an impact on our reasoning, since it would mean that the amphorae types could all be from specific time periods.

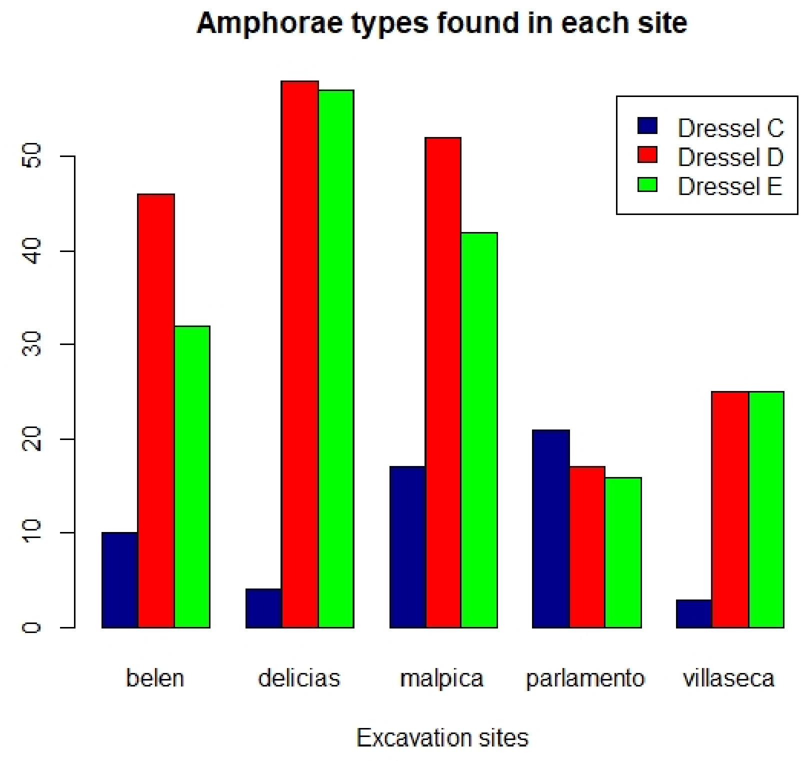


Figure 3: Bar graph showing distribution of each amphorae type per site

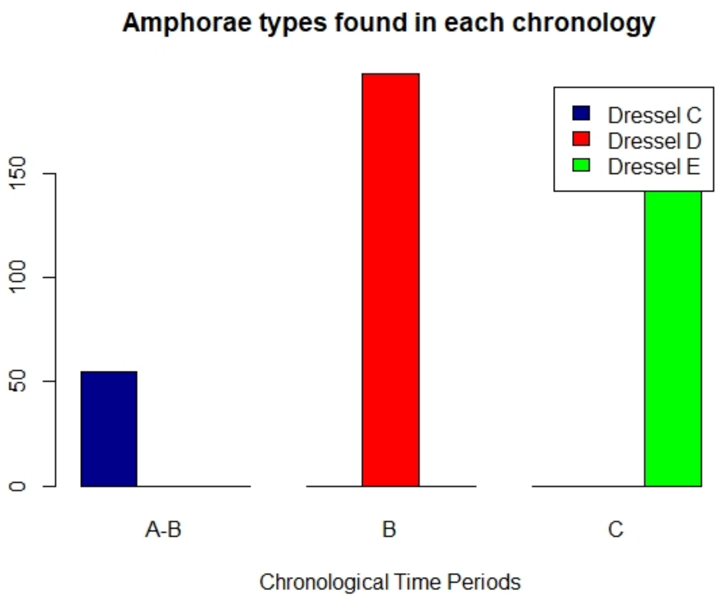


Figure 4: Bar graph showing distribution of each amphorae type per chronology

TABLE 1: Showing count of Dressel C, D and E for the five sites

	belen	delicias	malpica	parlamento	villaseca
Dressel C	10	4	17	21	3
Dressel D	46	58	52	17	25
Dressel E	32	57	42	16	25

C. Hypothesis Testing

1. Chi-Square Analysis test of Independence

A Chi-square test was conducted to check for associations between the categorical variables – amphorae types and excavation sites. The null hypothesis was rejected i.e. there was a statistical association between them. Therefore, we can presume that there was a relation between the type of amphorae found with the site in which they were present. The Chi-Square analysis does not give an indication of the nature of association, but we make can make a preliminary assumption that there could have been a preference for certain type of amphorae within the region. This would however require much further analysis through statistical analysis.

There was also another test conducted to check for associations between amphorae types and chronology. The null hypothesis was rejected in this case as well.

2. ANOVA

The main hypothesis of this paper is the homogeneity of production techniques across the sites. For this the Analysis of Variance/ANOVA method has been used. The test checks whether there is a similarity in means of a) external rim diameters b) internal rim diameters for each of the five sites. It was conducted on the smaller datasets of Dressel C, Dressel D and Dressel E types. The hypothesis is that if these two morphometric dimensions have a similar mean across sites, they probably have a similar mean shape, and therefore followed a stanndardised production method.

The box plot visualisations and Q-Q plots generated can be found in Appendix A and B.

- ❖ Dressel C

There was a total of 55 observations of Dressel C in the dataset.

```
       excavation         type      chronology exterior_diam   inside_diam
 belen      :10   Dressel C:55   A-B:55     Min.   :130   Min.   : 80
 delicias   : 4                             1st Qu.:150   1st Qu.: 88
 malpica    :17                             Median :160   Median : 90
 parlamento:21                              Mean   :161   Mean   : 93
 villaseca : 3                              3rd Qu.:170   3rd Qu.:100
                                            Max.   :190   Max.   :130
```

Figure 5: Statistics summary for Dressel C

Exterior Diameter: The boxplot for visualization of the data can be seen in Appendix Figure. A1.1. The range amongst the sites seems to be quite uneven. The Shapiro-Wilk test for normality failed to reject the null hypothesis. The Q-Q plot showed a near normal distribution (Appendix Figure A1.2). Subsequently, the Levene's test showed that there was no significant difference between variances. The parametric F-Test was conducted and failed to reject the null hypothesis.

Interior Diameter: The Shapiro-Wilk test rejected the null hypothesis of normality. The outliers revealed three values within Malpica and Parlamento. Since the sample size of these two areas already not very high, and there was no significant change in normality of data without them, the outliers were not removed. The total sample size being relatively smaller, the Kruskal-Wallis test was used for analysis, and failed to reject the null hypothesis.

Dimension used	Test used	P value	Association
Exterior_Dimension	F-Test	0.10	T
Interior_Dimension	Kruskal -Wallis	0.45	T

- ❖ Dressel D

There was a total of 198 observations of Dressel D in the dataset.

```
     number          excavation          type       chronology exterior_diam   inside_diam
 Min.   :  1   belen      :46   Dressel D:198   B:198      Min.   :145   Min.   : 70
 1st Qu.: 25   delicias   :58                              1st Qu.:160   1st Qu.: 90
 Median : 48   malpica    :52                              Median :170   Median : 91
 Mean   : 53   parlamento:17                               Mean   :168   Mean   : 94
 3rd Qu.: 80   villaseca :25                               3rd Qu.:175   3rd Qu.:100
 Max.   :128                                               Max.   :190   Max.   :120
```

Figure 6: Statistics summary of Dressel D

Exterior Diameter: The boxplot variation showed a very low range in values for Parlamento, and the maximum at Villaseca. The Shapiro-Wilk test rejected the null hypothesis of normality. The Q-Q plot showed little deviation from the fitted line though, aside from a heavy tail. The step type graph is an indication of several discrete values, possibly caused due to the

Parlamento site. The skewness value was checked and was seen to be only moderately skewed. The outliers were observed to mostly belong to the Parlamento site, and were thus removed, since quite a significant amount of skewness could be observed in the box plot. Since the homogeneity of variance condition was also rejected, the non- parametric test was applied. It failed to reject the null hypothesis.

Interior Diameter: Interestingly, the boxplots of the sites were almost equal to one another. The Q-Q plot as will shows a heavy concentration of discrete values around the center. The Shapiro-Wilk test rejected the null hypothesis, however on the basis of the Q-Q plot showing little deviation and level of skewness revealing a moderate skewness, the parametric F-Test was conducted. The outliers were kept since they were not extreme values and removing had no significant change on the data. It failed to reject the null hypothesis.

Dimension used	**Test used**	**P value**	**Association**
Exterior_Dimension	Kruskal-Wallis	0.15	T
Interior_Dimension	F-Test	0.72	T

- ❖ Dressel E

There was a total of 172 observations of Dressel E in the dataset.

```
     number          excavation          type       chronology exterior_diam  inside_diam
 Min.    :  2   belen      :32   Dressel E:172   C:172       Min.   :130   Min.   : 70
 1st Qu.: 23   delicias   :57                                1st Qu.:165   1st Qu.: 90
 Median : 44   malpica    :42                                Median :170   Median : 90
 Mean   : 50   parlamento:16                                 Mean   :170   Mean   : 94
 3rd Qu.: 71   villaseca :25                                 3rd Qu.:180   3rd Qu.:100
 Max.   :124                                                 Max.   :200   Max.   :140
```

Figure 7: Statistics summary for Dressel E

Exterior Diameter: The boxplots show a not very large difference in range of data. The Q-Q plot shows a ‘heavy tail’ possibly due to a number of extreme values. The outliers were not removed since they would skew the original sample size more than it already was. Since normality could nor be assumed, and there was a significant difference between variances, the Kruskal-Wallis test was applied. The test rejected the null hypothesis.

Interior Diameter: The boxplots for the sites seemed to be very uniform and of the same range. Aside from a rejection of the hypothesis of normality, he skewness level in this case was also highly skewed. Therefore, the Kruskal-Wallis test was applied. It failed to reject the null hypothesis.

Dimension used	**Test used**	**P value**	**Association**
Exterior_Dimension	Kruskal-Willis	1.68e-10	F
Interior_Dimension	Kruskal-Willis	0.05	T

DISCUSSION

The Chi-square analysis on the type and chronology variables was done to see if conclusions could be made about change in production technology over time. However, while there seems to have been an association as per the test, the bar graph shows the occurrence of only one amphorae sherd type per period. Therefore, it was not further incorporated into the analysis.

From the results of the ANOVA analysis, we see that there is a similarity in means of interior and exterior diameters. From this we can deduce that there would be similarity in the mean shape of the amphorae as well. In the case of Dressel D, where the sample sizes are the most similar to each other, we see the box plots show a near equal range of diametric dimensions both inner and outer. This adds weight to the hypothesis, that a similar production technique was followed, leading to a standardization of dimensions. There is some unevenness in the case of Dressel C, however we also see that the sample sizes have quite a large amount of difference, with sites like Delicia and Malpeca consisting of only 3 and 4 samples, respectively.

With respect to Dressel E, we come upon a bit of an anomaly. While the inside diameters show a similarity in means, the null hypothesis was rejected for exterior diameters. This seems to be unlikely with consideration to the fact that the inner diameters have a similar mean, as well as a lot of similarity in range of values, as seen in the boxplot visualizations. This can be explored further through a correlation analysis of the inner and outer diameters of Dressel E sherds. The hypothesis was rejected based on the p-value being less than 0.05. This could have been affected by variability of data between the sample sizes, which is a significant amount in the case of Dressel E, with 57 sherds from Delicia and 16 from Parlamento. Analysis of the hypothesis based on a combination of more criterias would be useful. One possibility is the use of Bayes factor. This gives an analysis of the likelihood of the null hypothesis, relative to the likelihood of the alternative hypothesis. By observing this factor, a deduction can be made as to how significant the likelihood of the alternative hypothesis being correct is in this case. In addition, a calculation of effect size along with 95% confidence interval can give an indication as to how strong of an effect the alternative hypothesis actually has on the sample (Halsey 2019). This consideration is true of the other cases as well, well the null hypothesis was not rejected.

The dataset used in the analysis had a lack of proper associated metadata, which made it difficult to understand some of the variable entries such as the ‘type’ and ‘chronology’, without prior deeper knowledge of amphorae typology & characterisations. The chronology specifically was simple given as A, B and C, without the proper associated time period. Due to lack of adequate prior experience into the study of different Roman amphorae types, and an absence of proper information to the same in the research paper, the analysis is probably limited in terms of what socio-economic or even more specific production techniques associated with these amphorae types could have had an impact. The dataset would have been more ideal if there were other diagnostic pieces in the dataset, apart from rims, in the context of answering the research questions.

The dataset contained more morphometric values which were not used in the analysis. The analysis of similarity or dissimilarity in terms of those dimensional values would give a better understanding and help better identify complementary tests to the ANOVA method, to ascertain the accuracy of the hypothesis. Since the means of the diameters were (mostly) homogenous, correlation analysis as well as linear regression analysis, would better help determine the influence of the two values on each other, and subsequently the impact of this homogeneity on the hypothesis of similar production techniques. This could be done in combination with the other morphometric values as well.

CRITICAL ASSESSMENT

Assumption of chronology: The Chi-square analysis works on the assumption that there is an independence in variables, in the sense that one is not directly a result of the other. In this study, it was difficult to determine from the literature whether this was the case with the type of amphorae produced and chronology. The analysis was conducted on the presumption that the 'chronology' variable referred to the chronological time of each individual sherd. However, if it so happens that it is merely a record of the chronology in which each amphorae type existed, it will render the Chi-square analysis potentially false. This has a high possibility, since it is less than likely that only a certain type of amphorae sherd was found as per a certain chronological period.

The bulk of the hypothesis testing in this analysis was done through the ANOVA method. A fair number of assumptions were taken, which determined whether a parametric or non-parametric test was conducted. In quite a few cases, the Shapiro-Wilk's case showed a lack of normality, however they have been considered near normal for the analysis in cases where a) sample size was larger 2) skewness was moderately skewed 3) the q-q plot did not seem to deviate too much from the fitness line. The others were analyzed through the non-parametric Kruskal -Wallis test. Although studies have shown that the ANOVA method is relatively less sensitive to the violations of assumption of normality (in case of moderate deviations) (Lix et al. 1996), there is still a risk of obtaining false positive values of non-significance. Outliers in the data have also been largely individually kept or left out of the data based on individual analysis.

Websites

https://edav.info/qqplot.html

https://www.statisticshowto.com/skewness/

http://www.sthda.com/english/wiki/compare-multiple-sample-variances-in-r

http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r

https://rdrr.io/cran/rstatix/man/outliers.html

https://www.datanovia.com/en/lessons/anova-in-r/

https://statisticsbyjim.com/basics/remove-outliers/

https://www.guru99.com/r-anova-tutorial.html

https://psychometroscar.com/2018/07/11/normality-residuals-or-dependent-variable/

http://r-statistics.co/Outlier-Treatment-With-R.html

BIBLIOGRAPHY

BlÁzquez, J.M. , 1993. The Latest Work on the Export of Baetican Olive Oil to Rome and the Army. *Greece and Rome* 39 (2), 173–88. https://doi.org/10.1017/S0017383500024153.

Bourgeon, O. , 2017. Baetican Olive-Oil Trade under the Late Empire: New Data on the Production of Late Roman Amphorae (Dressel 23) in the Lower Genii Valley. *Journal of Roman Archaeology*. Journal of Roman Archaeology L.L.C. https://doi.org/10.1017/S1047759400074249.

Coto-Sarmiento, M., X. Rubio-Campillo, and J. Remesal. , 2018. Identifying Social Learning between Roman Amphorae Workshops through Morphometric Similarity. *Journal of Archaeological Science* 96 (August), 117–23. https://doi.org/10.1016/j.jas.2018.06.002.

Halsey, L.G. , 2019. The Reign of the *p* -Value Is over: What Alternative Analyses Could We Employ to Fill the Power Vacuum? *Biology Letters* 15 (5). https://doi.org/10.1098/rsbl.2019.0174.

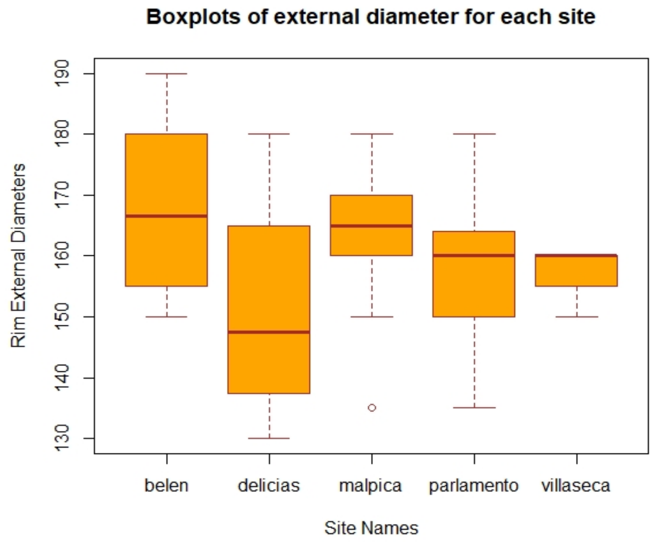
Wilczek, J., F. Monna, P. Barral, L. Burlet, C. Chateau, and N. Navarro. , 2014. Morphometrics of Second Iron Age Ceramics - Strengths, Weaknesses, and Comparison with Traditional Typology. *Journal of Archaeological Science* 50 (1), 39–50. https://doi.org/10.1016/j.jas.2014.05.033.

Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Rev. Educ. Res*. 66: 579-619.

APPENDIX

A. Quantitative Hypothesis Testing results for exterior diameter

A.1 Dressel C



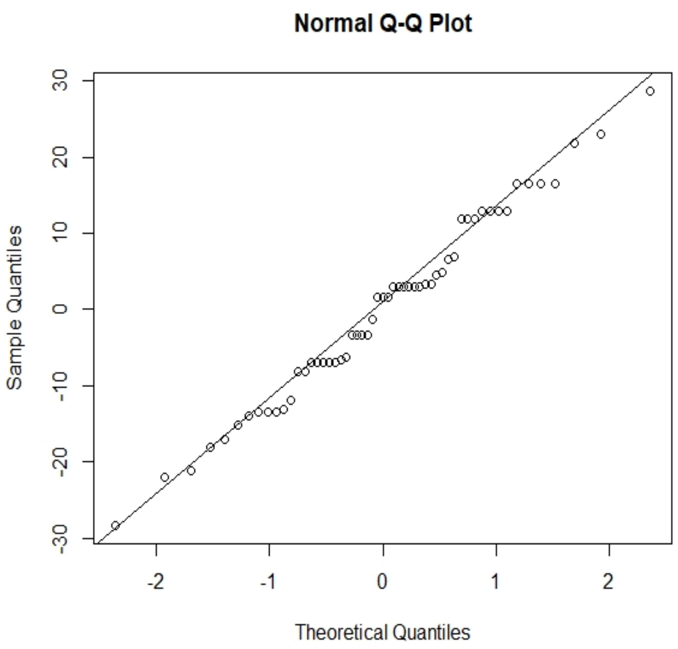
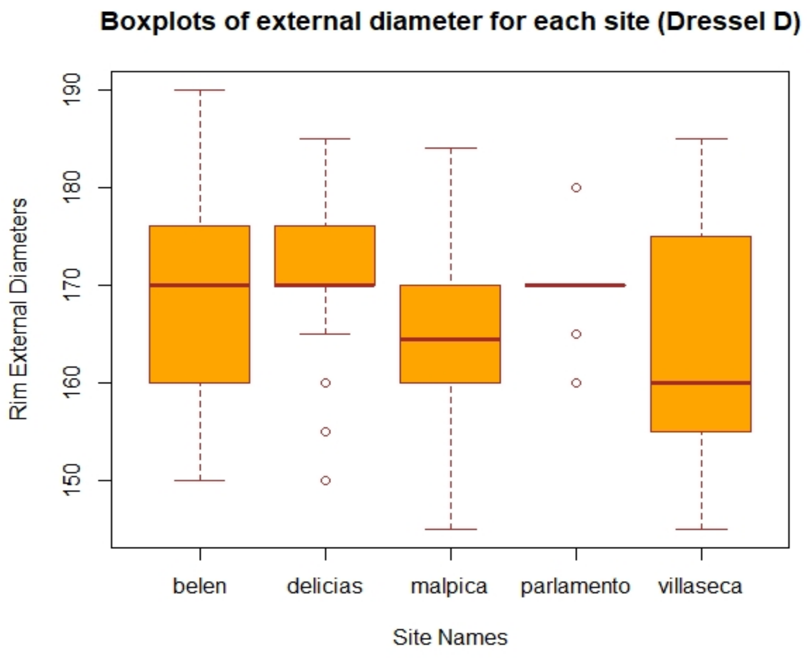


Figure A.1: 1. Box plot of rim external diameters for each site 2. Q-Q Plot to check normality

A.2 Dressel D



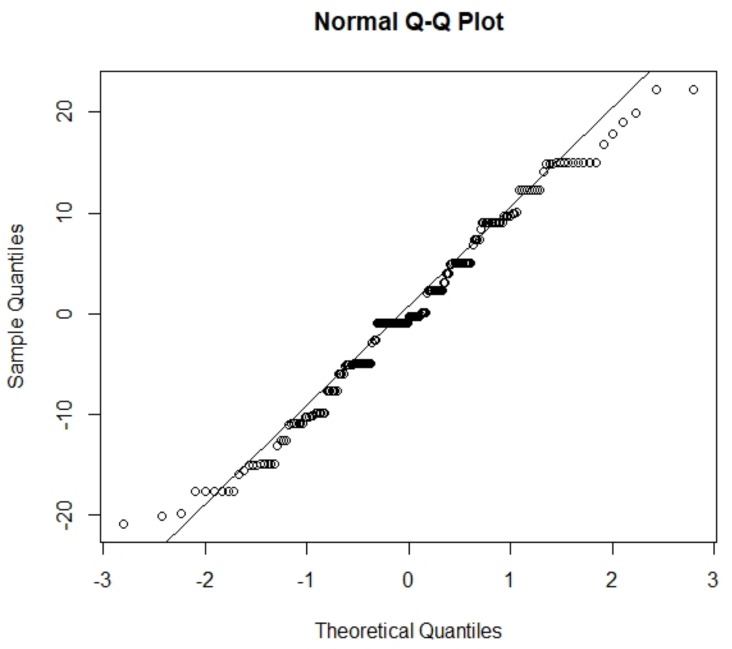
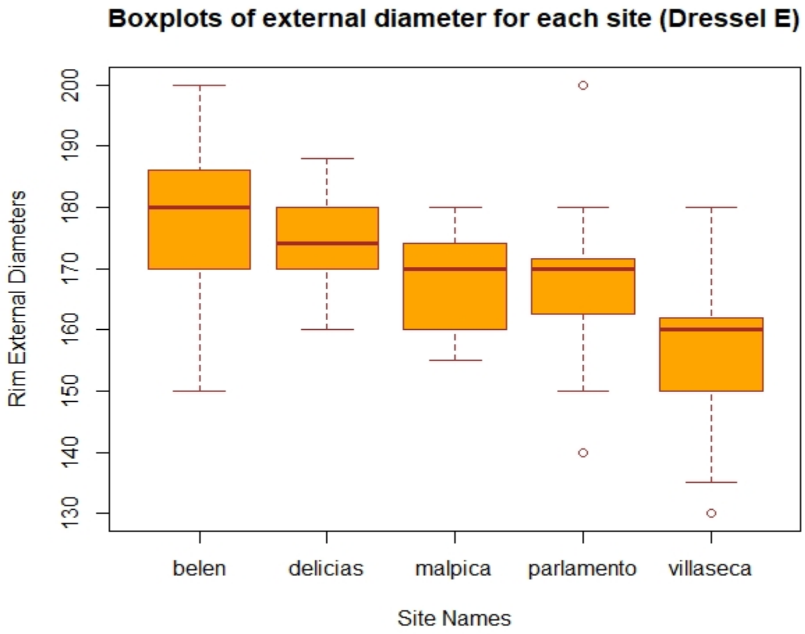


Figure A.2 1 Box graphs for Dressel D (original, without outliers) 2. QQ Plot to test normality (original, without outliers)

A3. Dressel E



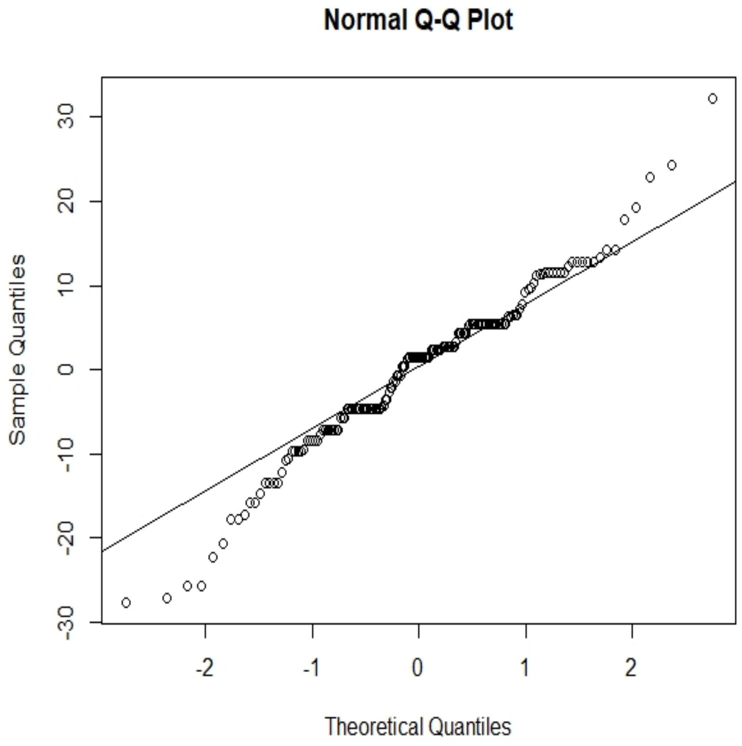
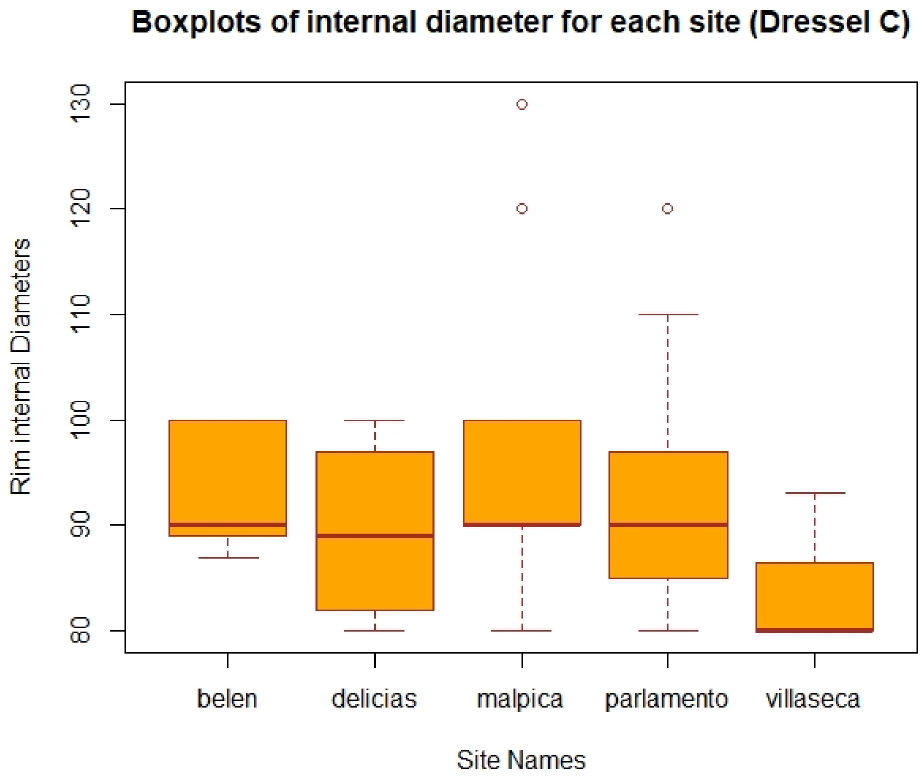


Figure A3.1 1 Box graph 2. Q-Q plot of exterior diameters of Dressel E

B. Quantitative Hypothesis Testing results for inside diameter

B.1 Dressel C



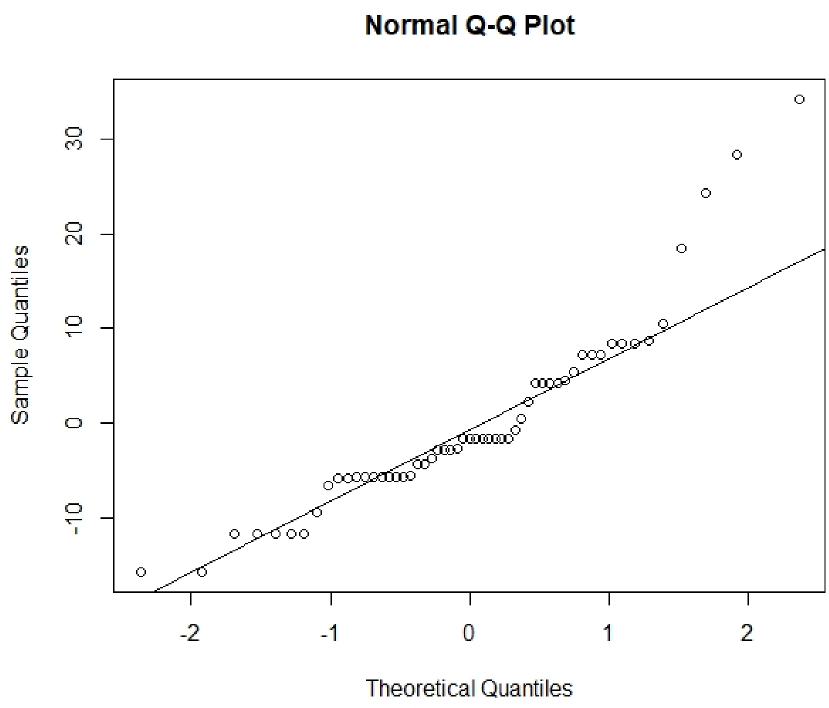
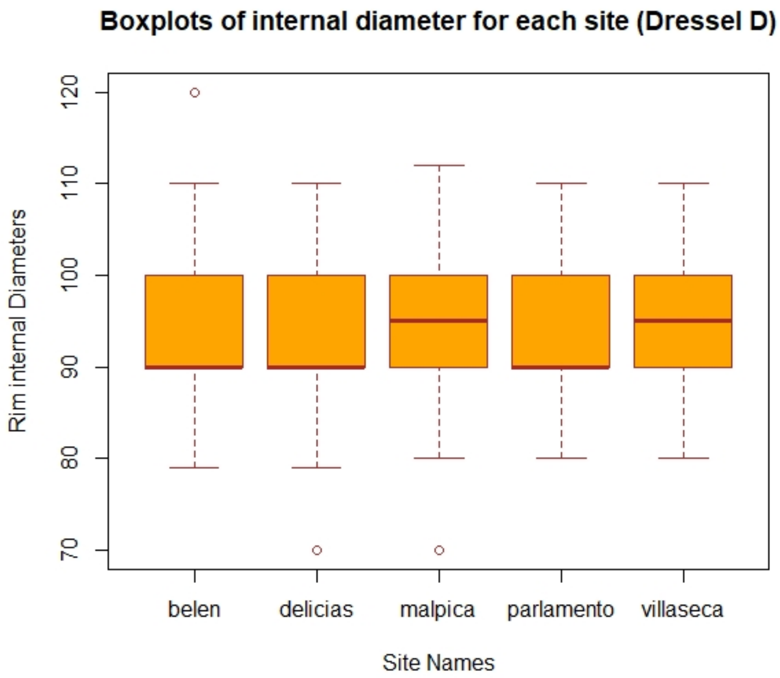


Figure B.1 1:Box plot of inside diameters for each site 2: Q-Q Plot to test normality

B.2 Dressel D



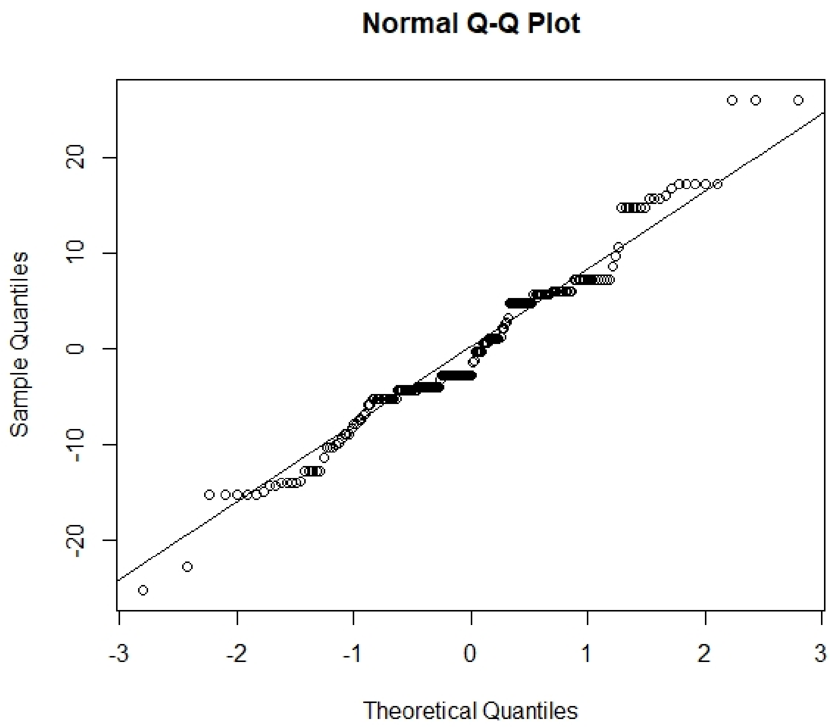
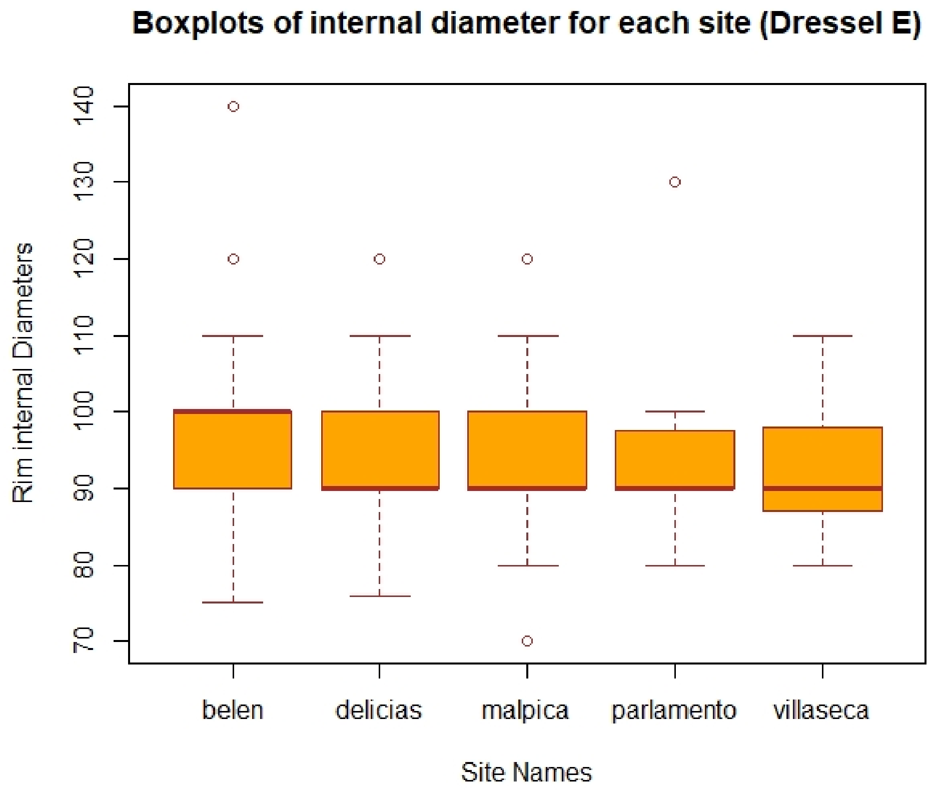


Figure B.2 1:Box graph for inside diameter in each site 2. Q-Q plot for test of normality



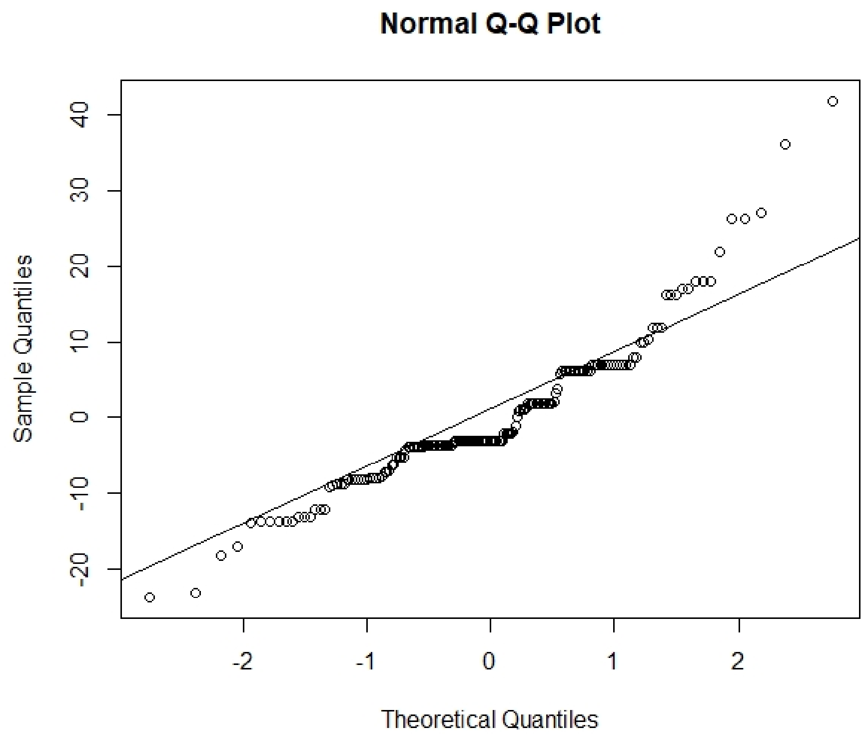


Figure B3 1 Box graph for inside diameter in each site 2. Q-Q plot for test of normality