Using Statistical Methods in Archaeology

A Repository of Useful Resources for Researchers

Created by

Aditi Majoe

under the supervision of Dr. Parth R. Chauhan, Associate Professor and Head of the Department, Humanities & Social Science, IISER-Mohali

Contents

1 About this Resource

This resource was created to provide students of Archaeology with a brief overview of some of the common statistical methods used in the field, guide them through the process of data collection and analysis, and to act as a repository for useful reference resources as jumping-off points. This resource is **not intended to teach statistical theory, or be a manual on applied statistics**, but rather to help students navigate basic methods of data collection, representation, and analysis.

Since most archaeology programmes in India don't require High School-level Maths as a pre-requisite, Archaeology students may have dropped maths as as a subject in High School. With their most recent statistics lesson having been years prior, these students might struggle to catch-up and make sense of any statistics courses in their curricula at the college level. This leaves students with many questions about the statistical methods best suited to their research. Fortunately for the modern student, thorough, comprehensive manuals on using statistics in archaeology already exist (see: Drennan, 2010; Joglekar, 2014; Mishra, 2012). Furthermore, the internet is replete with blogposts, video-tutorials, and lectures on a whole host of statistical questions one might find themselves asking. This document aims not to regurgitate the same information that can be found in any number of statistics manuals, but rather to draw out the most salient points a researcher must consider, and direct them to places where they can learn more. Please refer to the list of resources at the end of every section, or the master-list at the end of the document for further readings.

There are certain problems inherent to the field of Archaeology that make decisions about the application of statistics relatively challenging. For one, archaeologists are, often despite their best efforts, working with small sample sizes and limited datasets. Poor preservation, bioturbation, and the palimpsestic nature of sites are some of the many problems Archaeologists encounter on a regular basis. Archaeologists are also often faced with non-reproducible datasets. Surface artefacts, once collected, or a pit once excavated, cannot be replicated exactly at another location for someone else to study. This is a major hurdle when it comes to *reproducibility and replication*-two key concepts in the process of scientific investigation. This provides even more reason for archaeologists to be trained in the application of rigorous, statistically-sound data collection and analysis methods suitable to their research settings.

Section Resources

- Drennan, Robert D. Statistics for Archaeologists. New York: Springer, 2010.

- Joglekar, Pramod P. Research Methodology for Archaeology Students. Pune: Gayathri Sahitya, 2014.

- Mishra, Ashok K. Statistics and Archaeology. Delhi: Agam Kala Prakashan, 2012.

2 The Role of Statistics in Research

Statistical analysis of data is one of the steps of the Scientific Method. Simply put, the scientific method is used to **systematically** make observations, and frame and test hypotheses.

A researcher typically starts with identifying a gap in the current knowledge from reading the literature of their particular discipline, or by making first-hand observations that they want to investigate. You will notice in the diagram below that Data Analysis falls in the latter half of the process. This should serve as a reminder that the quality of data analysis can only be as good as the quality of the steps that precede it :)
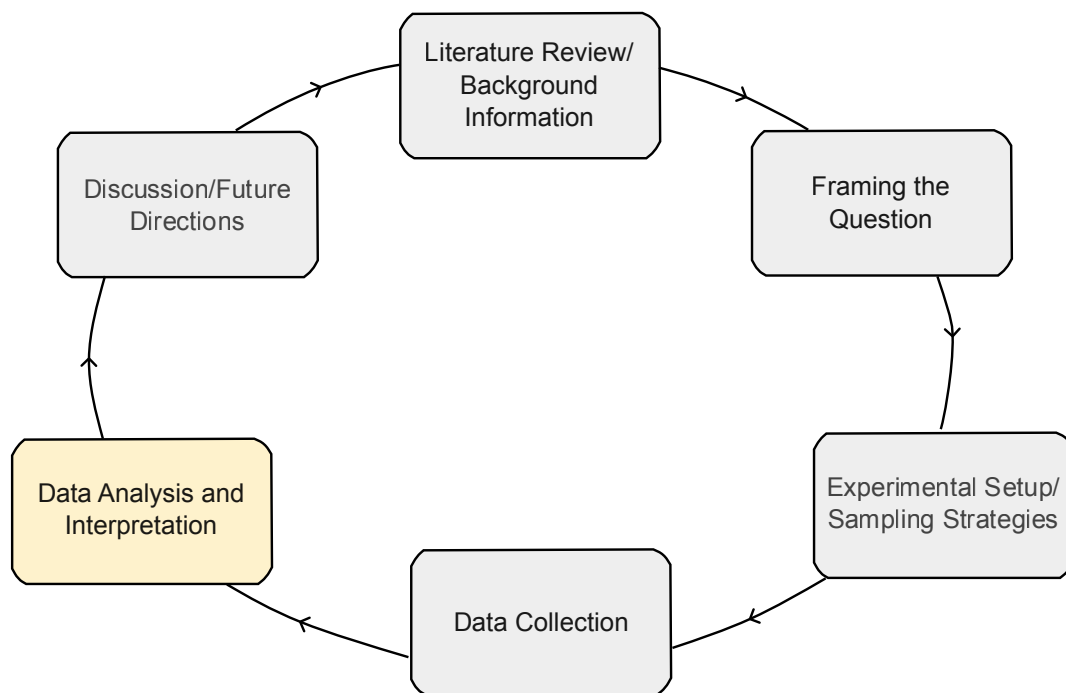

Figure 1: The Scientific Method

Section Resources

- The Scientific Process- American Museum of Natural History

- Carey, Stephen Sayers, and S. S. Carey. A beginner's guide to scientific method. Belmont, CA: Wadsworth Publishing Company, 1994.

- Gower, Barry. Scientific method: A historical and philosophical introduction. Routledge, 2012.

3 Framing the Question

Identifying the avenues of questioning based on previous research is the first step a researcher takes towards building a project. Framing a research question is the next step. It is often deceptively easy. This is because it often seems intuitively obvious in our heads what we intend to study. However, sometimes the question we initially frame is a nebulous, 'umbrella' question. In such cases, we have to break our question down into 'digestible pieces'. This means we have to identify the variables needed to answer the questions of interest, and how you will go about collecting that data.

Framing a good question and carefully considering the variables of interest also has the advantage of helping you determine the data you want to focus your attention, both in the field, and during the data analysis process.

The following resource, though not targeted towards archaeology, provides a thorough picture of the steps to formulating an effective research question: Formulation of Research Question – Stepwise Approach

4 Sampling Methods

In an ideal world, everything you encounter during an excavation or survey would be recorded in the most complete and accurate way possible. However, in the non-ideal, real world, we are constrained by time, manpower, transportation, funding, storage space, and many other factors. We cannot collect, store, or analyse every morsel of data we encounter. It is up to the researcher to determine what kind of data needs to be collected to answer the research question, and how to form representative samples appropriate for their particular field site.

In inferential statistics, we endeavour to draw conclusions about a population based on a subset, or sample of that population. Because of this, we want to make sure that the sampling strategy we apply helps us form samples that are representative of the population of interest.

There has been much debate, and discussion about the best practices when it comes to sampling strategies over the past decades (Woodman, 1981; Schiffer et al., 1978; Banning, 2021)

There are many online resources explaining the different kinds of sampling methods (some of which are included at the end of this section). We need not recapitulate that information here, so as to avoid being repetitive. Let us instead consider different scenarios where one might choose to use different sampling strategies:

<u>Simple Random Sampling</u>

You want to investigate the type of cobbles naturally occurring on the banks of a river. One strategy you may choose to use, is to 'randomly' select 1 m^2 boxes along the bank, and select all cobbles that happen to fall in that box

Figure 2: The above shows a possible sampling strategy, wherein cobbles are collected from randomly selected 1 m^2 squares on a grid

Tip: People are often not very good at being "random," our pattern-seeking brains subconsciously making biased selections despite best intentions. Consider using tools such as an online random number generator to help you make random decisions!

Cluster Sampling

You are interested in studying the prevalence of a certain pigment in the rock paintings of a series of rock shelters with twenty shelters, each with 5 paintings. In this case, you might randomly select ten of the twenty shelters, the underlying assumption being that each of the twenty shelters, i.e. 'clusters,' is similar to the others. Thus, randomly selecting a few of the clusters will still result in a good representative sample. You can use this same logic to further select two or three of the paintings from each of the ten shelters.

Figure 3: Ten out of twenty clusters are randomly selected. Two or three paintings are selected from each of the selected clusters using further random sampling.

<u>Stratified Sampling</u>

You are interested in analysing the salient architectural features of 17th century forts in India. Some of these forts are water forts, while others are hill forts. You recognise that location of the fort might majorly affect the kinds of architectural features you find in them. Thus you have two 'strata,' which need to be sampled from equally. If you chose forts to sample from "randomly," there is a chance of one type of fort being over-represented in your data. Considering the two fort types as separate groups, or strata is a way to control for a known variable which affects the results.
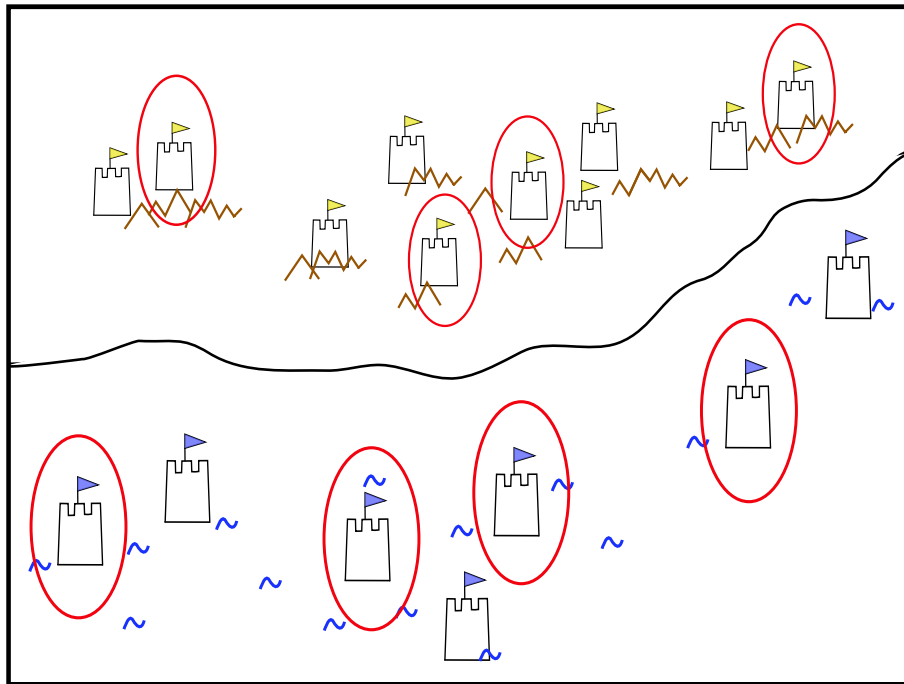

Figure 4: Example of stratified sampling, where water forts and mountain forts comprise two different strata, within which we can sample randomly

<u>Section Resources</u>

- Sampling Strategies

- Archaeological Sampling Techniques-Dig It With Raven

- Green, L. (2007). Analysis of archaeological sampling methods using the complete surface data from the Pirque Alto site in Cochabamba, Bolivia (Doctoral dissertation, Archaeological Studies Program, University of Wisconsin-La Crosse)

- Orton, C. (2000). Sampling in archaeology. Cambridge University Press.

- Woodman, P. C. (1981). Sampling Strategies and Problems of Archaeological Visibility. Ulster Journal of Archaeology, 44/45, 179–184. http://www.jstor.org/stable/20567872

- Schiffer, M. B., Sullivan, A. P., & Klinger, T. C. (1978).The design of archaeological surveys. World Archaeology, 10(1), 1–28. https://doi.org/10.1080/00438243.1978.9979712

- Banning, E. B. (2021). Sampled to death? The rise and fall of probability sampling in archaeology. American Antiquity, 86(1), 43-60.

5 Data Collection, Storage and Management

Once you know what variables you will need to collect in order to answer your research questions, you can make a data sheet with all the variables of interest. A systematic, well-laid-out data sheet is worth spending some time thinking over in the beginning of your project, because it will save you a lot of time and heartache later down the line! You could build your data sheet from scratch, or start with a basic template used in your field and customise it to suit your needs. You will most likely do this using some kind of spreadsheet software, such as MS Excel, Google Sheets, Numbers, LibreOffice Calc, etc.. Whatever software you choose to use, make sure that you set your files to automatically back-up to cloud storage (Google Drive, Dropbox etc.), or get in the habit of manually backing-up your files to a hard-disc or the cloud at regular intervals during the data collection process. Another advantage of computer spreadsheets is that you can leave a digital footprint showing the time and/or location of data collection, in case that information is needed in the future. While digital spreadsheets are incredibly useful, you might choose to print your empty sheets and fill them in physically (e.g. if you cannot carry your devices to the field, if you have no access to power or internet connection on the field, personal preference etc..). If so, you might want to consider taking pictures of filled-out pages on your phone so you have a digital version of your data, as a back-up for if something happens to the physical sheets. There is some room for human error when transcribing handwritten notes to digital datasets, which one should be vigilant of.

Section Resources

- Top 5 Data Management Tips for Graduate Students- University of Wisconsin-Madison

- Research Data Management Support- Utrecht University

6 Data Analysis

6.1 Types of Data

Before we get into data analysis, we need to understand the kind of data we are working with. There are two types of data: numerical (aka quantitative), and categorical (aka qualitative). As the names suggest, numerical variables are those which can be measured and quantified, whereas categorical variables cannot. These are further subdivided into discrete and continuous (in case of numerical data), and nominal and ordinal (in case of categorical data).

Examples of quantitative variables include weight of an artifact (measured in kg), dimensions of an artifact (measured in cm). Examples of qualitative variables include colour of soil, shape of artefact, etc.. If the categorical data have an innate hierarchy to them, they are called nominal. For example, one might have a variable called 'Soil Colour,' with the categories 'yellow-brown,' 'red-brown,' and 'red.' Soil Colour, in this case, would be a nominal categorical variable. On the other hand, one might have a variable called 'Artefact Fragmentation Level,' with categories 'mostly fragmented,' 'moderately fragmented,' 'minimally fragmented,' 'mostly intact'. Clearly, there is an innate hierarchy to these categories, making the variable an ordinal categorical variable.

It is tempting to think that a recorded variable that has numbers in it must automatically be numerical, but this is false. A very common occurrence of this is when a researcher uses numbers as placeholder for categories of a qualitative variable for ease and speed of input into data tables (e.g. 1= mostly fragmented, 2= 'moderately fragmented, 3 = minimally fragmented, and 4 = mostly intact). When analysing this dataset, one should remember that despite the Artefact Fragmentation Level column being populated by numbers, this variable is still categorical.

Subsection Resources

- Qualitative versus Quantitative Variables- Michelle Oja, Taft College

- Data Types in Statistics- GeeksforGeeks

6.2 Hypothesis Testing

Hypothesis testing, as the name implies, helps one determine if the differences they observe between variables are different enough to be considered separate. As we will briefly discuss later on, sometimes we compare our sample to an established population parameter, and sometimes we compare samples with each other.

There are an abundance of resources on the internet explaining the theory of hypothesis testing, as well as its steps. Please refer to any of the following as starting points:

- Understanding Hypothesis Testing- GeeksforGeeks

- Hypothesis Testing- La Trobe University

- Hypothesis testing (ALL YOU NEED TO KNOW!)- zedstatistics Youtube

- Intro to Hypothesis Testing in Statistics - Hypothesis Testing Statistics Problems & Examples- Math and Science Youtube

It can be challenging to determine that kind of test is appropriate for the kind of data we have. The flowchart on the following page might help you decide what kind of test you should conduct:

Is your data quantitative (means) or categorical (proportions)?

Quantitative

Categorical

Are the two variables independent of one another?

Relationship

Are you studying the relationship between two variables, or comparing means?

Are you studying the relationship between two variables, or comparing proportions?

Compare proportions

Z-test

Yes

No

Comparing means

Relationship

Correlation

Regression

How many samples do you have?

Do you have a 2x2 contingency table?

No

Yes

One Sample to be compared to a population mean

Two Samples to be compared to each other

Three or more samples to be compared to each other

Fisher test/ Chi-square

Chi-square

One sample t-test

Two-sample t-test

ANOVA

Are the assumptions of the test met?

No

Appropriate alternative/ nonparametric test
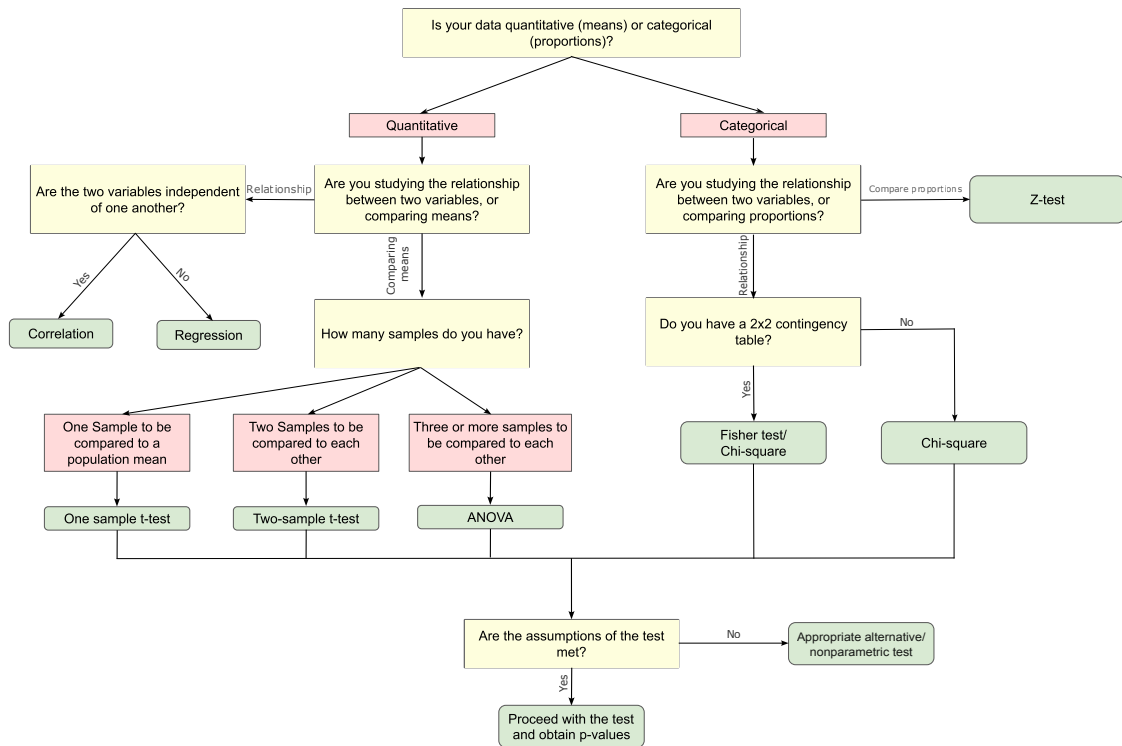
Yes

Proceed with the test and obtain p-values

Figure 5: Decision Tree for Hypothesis Testing

The following table provides the most common parametric tests performed on different types of data when working with two variables:

	Quantitative	Categorical
Quantitative	Correlation/Regression	T-tests/ANOVA
Categorical	-	Chi-square/Fisher Test

It is important to note that every hypothesis test has some underlying assumptions. If you should find that your data do not satisfy the assumptions of a certain parametric test, consider using alternative tests.

Archaeologists often find themselves in a position where their sample sizes are simply not large enough to justify parametric tests. In some such cases, this can be overcome by 'bootstrapping' or simulating samples. Other cases might call upon the use of non-parametric tests. The approach you choose depends on the nature of your data.

The table below lists some of the most common non-parametric alternatives to common hypothesis tests:

Variable Types	Parametric Test	Non-Parametric Test
Quantitative; Categorical	t-test (one-sample, or two-sample unpaired)	Mann-Whitney test (aka Wilcoxon rank-sum test)
Quantitative; Categorical (2 paired groups)	Paired t-test	Wilcoxon Signed-Rank test
Quantitative; Categorical (3 groups or more)	ANOVA	Kruskal-Wallis test

You can also use online tools such as this to help you determine what hypothesis test you should run on your data.

6.2.1 Post Hoc Testing

There will be instances of hypothesis testing where the p-value will not tell you everything you want to know. Most commonly in ANOVA or Chi-sqaure tests, the p-value only indicates whether all groups are equal to one another (null hypothesis), or if atleast one of the groups is significantly different (alternative hypothesis). Post hoc testing is done to determine *which* group is different, by performing multiple pairwise comparisons between groups while adjusting for error-rates. There are many types of post-hoc tests, most common being Tukey's Test, and Holm-Bonferonni Test. As with hypothesis testing, every post-hoc test has certain conditions under which they are most appropriately used. Thus, the method you choose must depend on the kind of data you have.

Section Resources

- Which Statistics Test Should I Use?- Social Science Statistics

- Post Hoc Definition and Types of Tests- Statistics How To

- Using Post Hoc Tests with ANOVA- Statistics by Jim

- Understanding Hypothesis Testing

- Hypothesis Testing- La Trobe University

- Hypothesis testing (ALL YOU NEED TO KNOW!)- zedstatistics Youtube

- Intro to Hypothesis Testing in Statistics - Hypothesis Testing Statistics Problems & Examples- Math and Science Youtube

6.2.2 Other Multivariate Analyses

We do multivariate analyses when performing tests that compare more than two variables at a time. Oftentimes two or more of these variables may be dependent.

MANOVA, MANCOVA, ANCOVA, Two-way ANOVA

We are already familiar with ANOVA (i.e. one-way ANOVA) where we compare the means of multiple groups at the same time. We can also perform two-way ANOVAs, ANCOVAs (**AN**alysis of **COVA**riance), MANOVAs (**M**ultivariate **AN**alysis **O**f **VA**riance), and MANCOVAs(**M**ultivariate **AN**alysis of **COVA**riance). Each of the tests have different types of response and explanatory variables. Refer to the chart below for a quick guide-through:

Test	Dependent Variable/(s)	Independent Variable/(s)
One-way ANOVA	1 numerical variable	1 categorical variable
Two-way ANOVA	1 numerical variable	2 categorical variables
ANCOVA	1 numerical variable	> 1 categorical and numerical variables
One-way MANOVA	> 1 numerical variables	1 categorical variable
Two-way MANOVA	> 1 numerical variables	2 categorical variables
MANCOVA	> 1 numerical variables	> 1 categorical and numerical variables

Subsection Resources

- ANOVA, ANCOVA, MANOVA and MANCOVA: Understand the difference- Stat Pharm

- ANOVA, ANCOVA, MANOVA & MANCOVA(example)- Research Methodology Advanced Tools

- Statistical Soup: ANOVA, ANCOVA, MANOVA, & MANCOVA- Stats Make Me Cry Consulting

PCA, MCA, and Clustering Data

PCA (Principal Component Analysis), MCA (Multiple Correspondence Analysis), and Cluster Analysis are very common methods used in social science research. The first two are used to determine the variables that capture the most variability in your dataset, PCA with continuous quantitative variables, and MCA with categorical variables. Finally, cluster analysis 'groups' observations together. All three methods consider multiple variables at once, and are often used in conjuction with each other (see: Ravisankar et al., 2014; Seetha & Velraj, 2016).
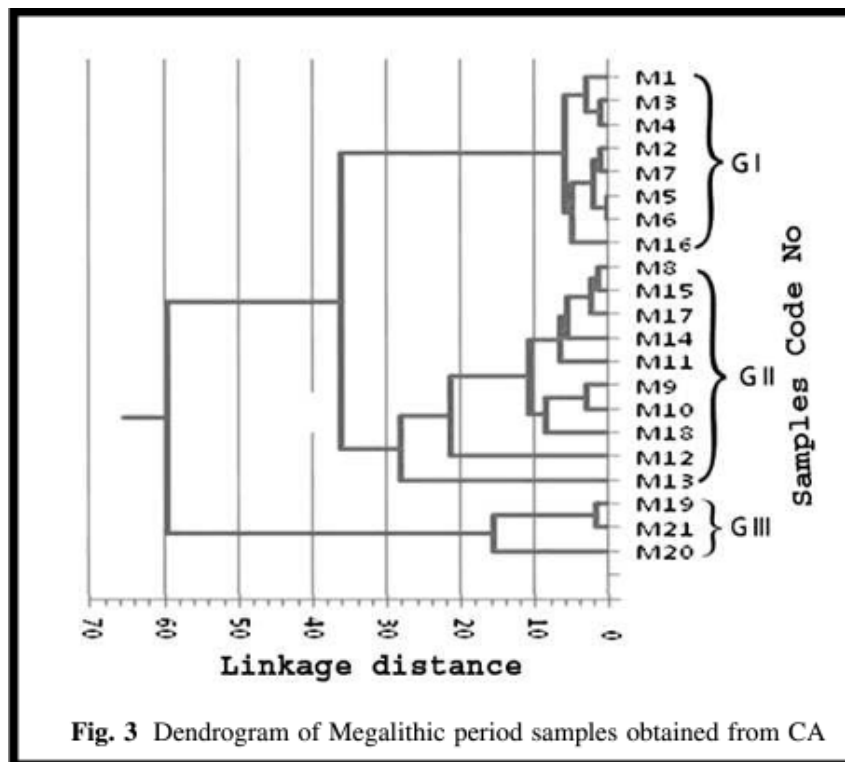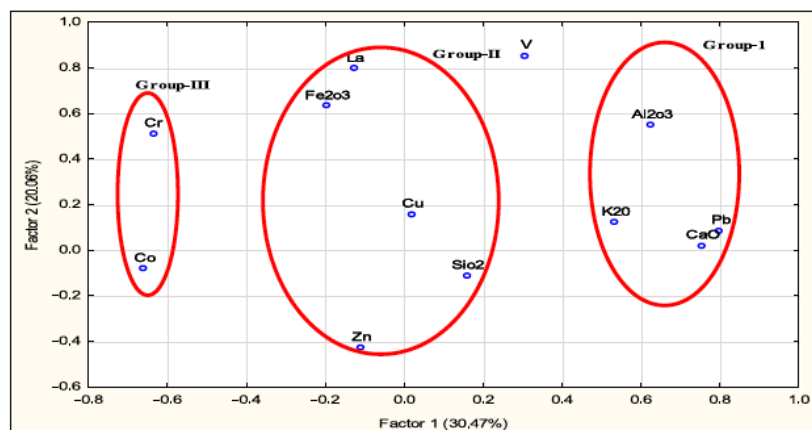
Figure 6: Example of a dendogram from Dasari et al., 2012

Figure 7: Example of a PCA from Ravisankar et al., 2014

<u>Subsection Resources</u>

- Principal Component Analysis (PCA)- Steve Burton Youtube

- Principal Component Analysis (PCA) in Excel

- A Comprehensive Guide to Cluster Analysis: Applications, Best Practices and Resources-Displayr

- Tutorial on MCA - Multiple Correspondence Analysis - with R- François Husson Youtube

- Dasari, K., Acharya, R., Das, N., & Reddy, A. (2012). A standardless approach of INAA for grouping study of ancient potteries. Journal of Radioanalytical and Nuclear Chemistry,

294(3), 429-434.

- Ravisankar, R., Naseerutheen, A., Chandrasekaran, A., Bramha, S. N., Kanagasabapathy, K. V., Prasad, M. V. R., & Satpathy, K. K. (2014). Energy dispersive X-ray fluorescence analysis of ancient potteries from Vellore District Tamilnadu, India with statistical approach. Journal of Radiation Research and Applied Sciences, 7(1), 44-54.

- Seetha, D., & Velraj, G. (2016). Characterization and chemometric analysis of ancient pot shards trenched from Arpakkam, Tamil Nadu, India. Journal of applied research and technology, 14(5), 345-353.

6.2.3 Reporting Results

It is important that you report your results in a format that conveys all the required information to your audience. There are some established conventions for how to report different tests across citation formats. For example, you can find the APA convention here. At minimum, you will need to state sample size (n), and p-value for the test you conducted, though you might also wish to state your alpha level (α) and confidence intervals (CI) in some cases. Different journals might follow their own conventions on reporting results. It's a good idea to familiarise yourself with them ahead of time.

6.2.4 What if you don't have Significant Results?

Many a researcher has had their hopes crushed by a less-than-stellar p-value (you can tell by the hopelessness in their eyes). One day, you too, might have to come face-to-face with an obstinately-larger-than-0.05 p-value! Or maybe that has already happened (my sympathies). An archaeologist might find themselves in this situation more often than not, because frequentist statistics shines in the presence of large sample sizes, and large sample sizes. . . we have not.

As you may already know, in hypothesis testing, we compare our p-value to the probability of making a Type I Error, which is also known as the 'alpha level.' So when the 'alpha value' is 0.05, there is a 5% chance of making a Type I Error (i.e. rejecting the null hypothesis when it is true). A Type II Error, on the other hand, is the probability of failing to reject an incorrect null, which is also called a False Negative result. In other words, it is the failure to identify an effect that really exists. Meanwhile, Type I Error is concluding there is an effect when there really isn't one. Broadly speaking, the two errors are inversely proportional. (To learn more about Errors: https://www.scribbr.com/statistics/type-i-and-type-ii-errors/)

	Null Hypothesis Correct	Null Hypothesis Incorrect
Failed to reject null	Correct conclusion	Type 2 Error (False Negative)
Rejected null	Type 1 error (False positive' also α-level)	Correct conclusion

You may ask (and rightly so) why we set alpha at 5% instead of 1%, or 10%, or 26%?! The answer is that. . . well. . . because other people have. . . ? You could adjust your alpha level to a higher percentage if you don't mind increasing the chances of getting a false positive (i.e. increasing chance of a Type 1 error). On the other hand, this also means you will be decreasing the chance of a false negative (i.e. decreasing the chance of a Type 2 error). It is up to you to decide if that trade-off makes sense for your particular case.

Much discussion has been had about the way p-values are used (and abused) in research over the past decades (Wasserstein & Lazar, 2016) The case against statistical significance testing). Many social scientists have started to adopt and incorporate Bayesian statistics into their analysis approach. Everything we have discussed so far falls into the realm of Frequentist Statistics, which relies on fixed probabilities, and compares a dataset against a set null hypothesis. Bayesian statistics on the other hand, updates probabilities according to new data. An advantage of such an approach is that it can accommodate for varying sample sizes. That however, to utilise a phrase beloved by researchers worldwide, is beyond the scope of this project :)

To answer the question asked in the title of this section: What should you do if you don't have significant results? Consider the factors that might have affected it. Is your sample size too small? Is there a lot of variability in your sample? Revisit what made you frame your alternative hypothesis (the effect you expected to see). Are there other published studies which speak to your results? How do your results sit in relation to other trends in the dataset? These are all avenues of discussions, and questions to guide future projects. *The absence of a statistically significant result does not necessarily mean the absence of opportunities for intellectual insight or meaningful discourse.*

Section Resources

- Type I & Type II Errors — Differences, Examples, Visualizations- Scribbr

- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133.

- Efron, B. (2005). Bayesians, frequentists, and scientists. Journal of the American Statistical Association, 100(469), 1-5.

- Albers, C. J., Kiers, H. A., & van Ravenzwaaij, D. (2018). Credible confidence: A pragmatic view on the frequentist vs Bayesian debate. Collabra: Psychology, 4(1), 31.

- Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis.

- Hackenberger, B. K. (2019). Bayes or not Bayes, is this the question?. Croatian medical journal, 60(1), 50. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6406060/)

- Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O'Connor, J. P., & Price,

G. J. (2022). Understanding the differences between Bayesian and frequentist statistics. International journal of radiation oncology, biology, physics, 112(5), 1076-1082.

- Otárola-Castillo, E., & Torquato, M. G. (2018). Bayesian statistics in archaeology. Annual Review of Anthropology, 47(1), 435-453.

- Otárola-Castillo, E., Torquato, M. G., & Buck, C. E. (2023). The bayesian inferential paradigm in archaeology. Handbook of Archaeological Sciences, 2, 1193-1209.

7 Graphical representation

Graphs provide an efficient way to give your reader (and yourself!) an idea of the nature and spread of your data, and how different variables look in relation to each other. A well made graph is easy to understand at first glance, and gives the reader all the information they need. Crucial components that your graph should include are: a title, labelled axes with the units of the variables in parentheses, clearly labelled legend for additional information, and a caption briefly describing the figure. The font style and size of the graph's text must be legible.

In addition to these, you must also think about the overall look of your graph. Because it is a visual medium, you have to keep in mind a certain sense of aesthetic quality while making your graphs. You can use colours strategically to help draw the reader's attention to key findings. For example: you could use contrasting colours to highlight dis-similarities, or gradient colours to represent gradation in data. Keep in mind that not all printers/projectors will be able to produce the exact shades as your computer screen, so your posters/presentations might look different in physical form than in digital form. Furthermore, some journals might have rules regarding use of colour, or an extra charge associated with colour prints. In such cases, you might choose to use different types of lines or point types in your graphs instead of colour.

To determine what graph you should use to represent your data, you should consider the following questions:

1. How many variables do you want the graph to represent?

2. Are the variables quantitative or qualitative?

If you want to represent just one variable, then you will most likely be using histograms, or density plots for quantitative data, and bar charts (or less commonly pie charts) for categorical data.
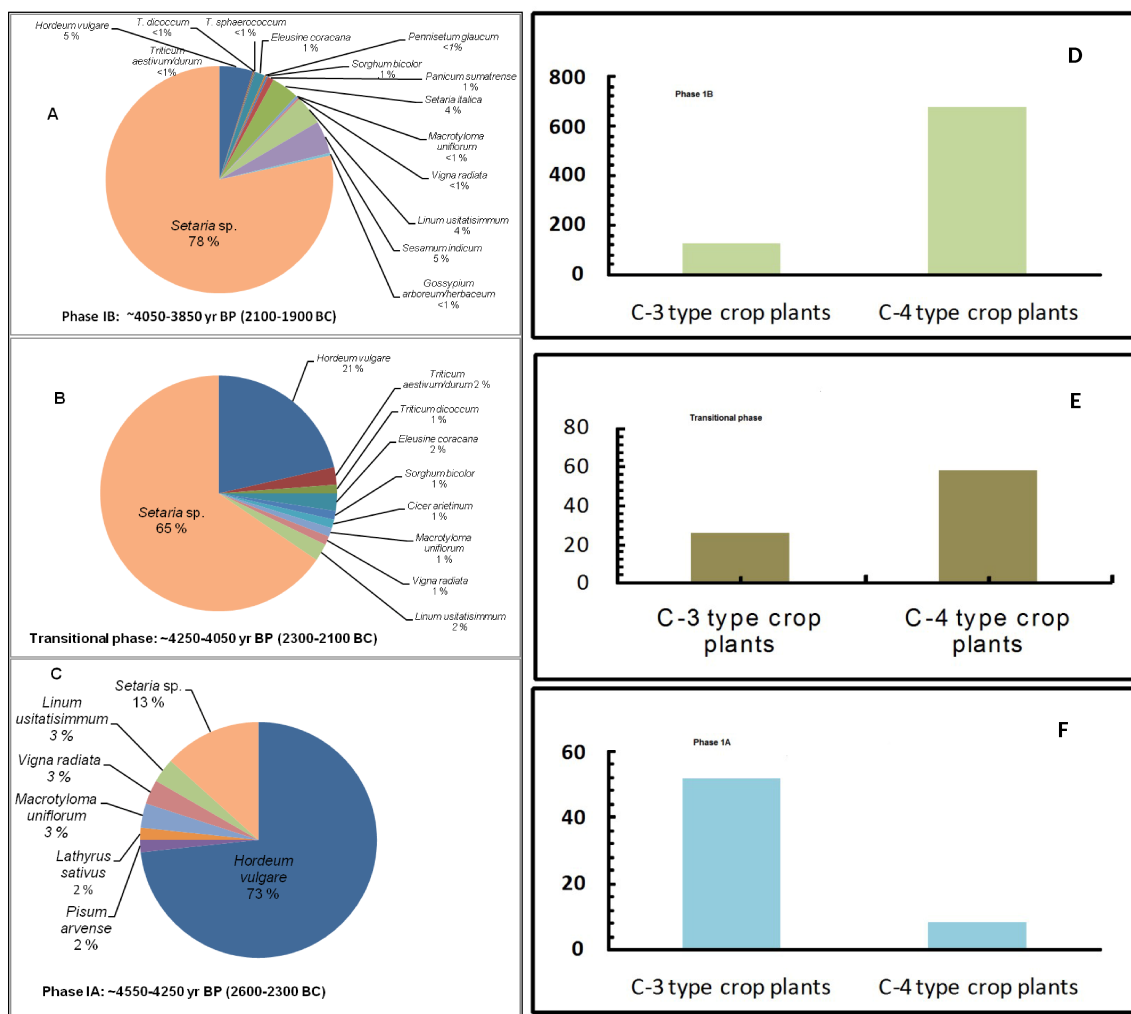
Figure 8: Examples of univariate plots (pie charts and histograms) from Pokharia et al., 2017

For <u>bivariate</u> plots, you can refer to the table below, which summarises the most common graph types for different types of data combinations"

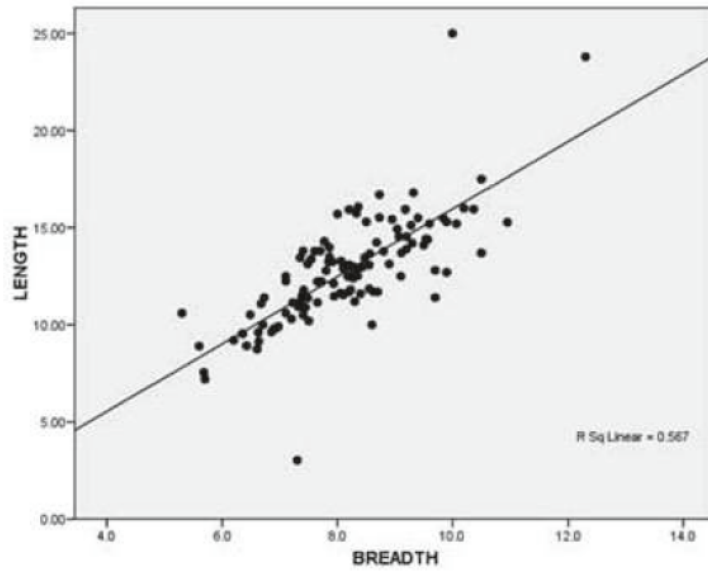|	Quantitative	Categorical
Quantitative	Scatter plot with regression line (aka trendline)	Box plot
Categorical	-	Contingency table,Bar plot (stacked or side-by-side)

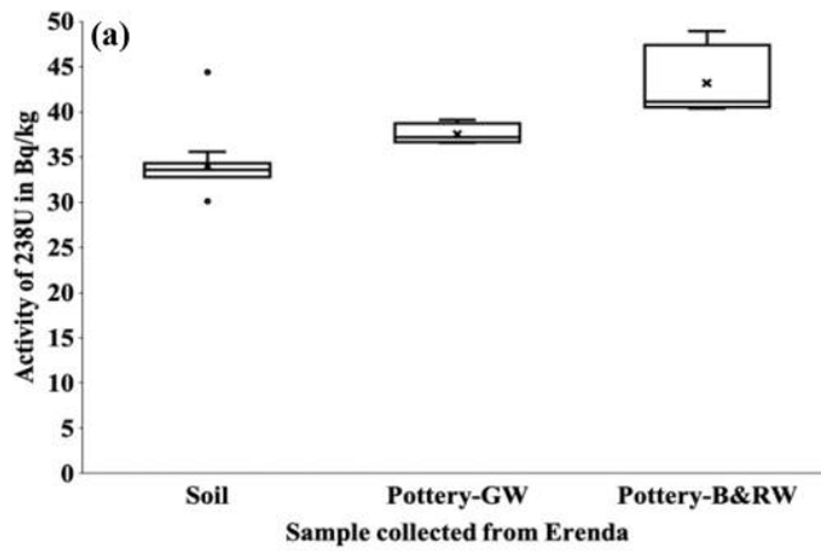Figure 9: A bivariate scatterplot of length vs. breadth. (source: Chauhan, 2010)

Figure 10: A bivariate boxplot showing levels of 283U in different samples (source: Naskar et al., 2024)

If you need to graph more than two variables, you can incorporate colour or shapes into your 2D graph to represent additional variables.
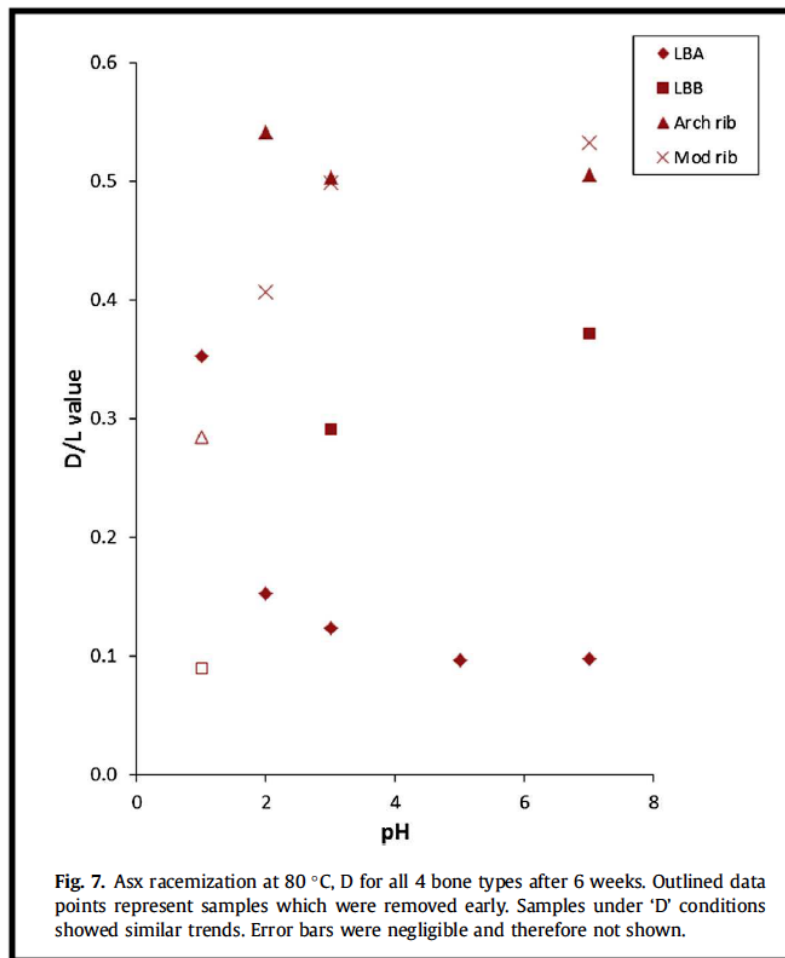

Fig. 7. Asx racemization at 80 °C, D for all 4 bone types after 6 weeks. Outlined data points represent samples which were removed early. Samples under 'D' conditions showed similar trends. Error bars were negligible and therefore not shown.

Figure 11: Figure from High et al., 2015. Note the figure represents three variables: one along the x-axis (pH), second along the y-axis (D/L value), and the third variable (bone type) represented with different "point shapes"

7.1 Reading and Interpreting Graphs

While it might seem intuitive to some people most of the time, it is worthwhile to spend a little time learning how to read graphs (and tables, and diagrams, for that matter).

The first piece of information you want to look at is what variables are represented on the x and y axes. Pay attention to the units of these variables.

Consider this set of graphs from Akhilesh, K., & Pappu, S. (2015).

Fig. 7. Artefact dimensions: A. Box plot showing lengths (mm) of all complete artefacts (n = 1147); B. Boxplot showing lengths (mm) of all complete flakes (and those with step terminations) through time (Layers 6, 7, 8) (n = 1388); C. Scatter plot of Length and Breadth for all complete flakes as per layer (n = 1388); D. Scatter plot for dimensions of flakes as compared with biface reduction flakes.
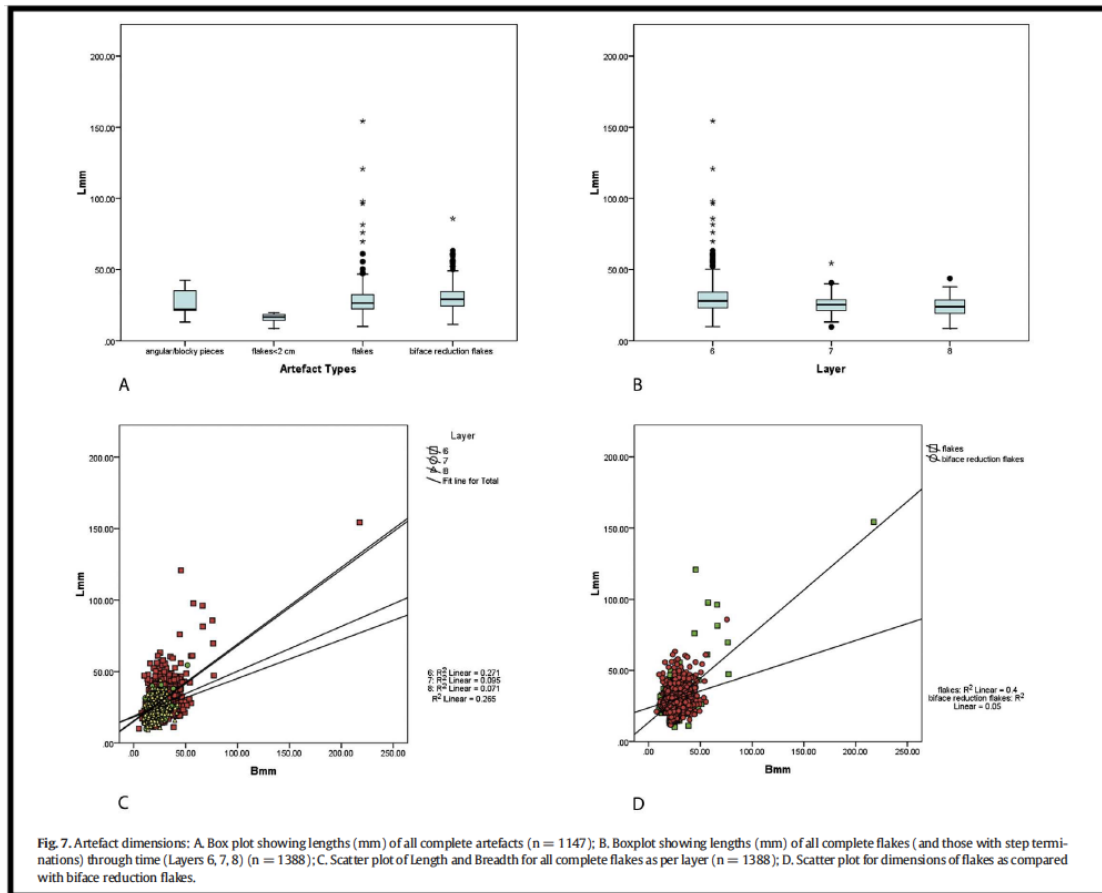
Figure 12: Graphs and figure caption from Akhilesh & Pappu, 2015

You should notice here:

1. The type of graph: We have two boxplots in the top row, and two scatterplots in the bottom row

2. The axes: Graph A has Artefact types on the x-axis, while Graph B has Layer on the x-axis. Both graphs C and D have Breadth in millimeters on their x-axes. All four graphs have Length in millimeters on their Y-axes.

3. The bounds of the axes: The y-axis ranges from 0.00-200.00 mm, while the x-axes for the bottom graphs range from 0.00-250.00mm

Next, you might notice the spread of the data in the boxplots. For example, in graph A, we can see that the categories of 'flakes' and 'biface reduction flakes' have quite a few outliers above Q3. The category of 'flake < 2cm,' on the other hand, has the least variability of the four categories. This makes intuitive sense. Afterall, the bounds of the category by definition are contained to less than 2cm. Analysing graphs in this manner gives you an idea of the nature of the data. Also remember that the more overlap there is between the boxplot spreads, the less likely they are to

be significantly different statistically. So there is a chance that the lengths of the different artefact types might not be significantly different from one another.

Similarly, in graph D, we can see two variables represented in the x and y axes respectively, and a third variable (Artefact Type) being represented via colour and shape of the points. You can see a regression line for each of the two artefact types, one of which is steeper than the other.

By comparing the two scatter clouds, you can see that flakes have more spread in the y-direction than the bifacial flakes. This is not surprising to us given what we learnt from Graph A. In Graph D, however, we also see that flakes have more spread along the x-axis as well. Moreover, there is one outlier flake in particular that seems to be both longer and wider than the rest of the flakes (this is the right-most point in the graph).

Finally, look at the regression coefficients provided at the foot of the graph. R-squared for flakes is 0.4 (or 40%), while R-squared for bifacial flakes is 0.05 (i.e. 5%). Does this match what you see in the graph? This means that 40% of the variability in Length of flakes can be explained by the Breadth of the flakes, whereas only 5% of the variability in the Length of bifacial flakes can be explained by their Breadth. Another way to think of this is that length and breadth of a flake have a stronger correlation than length and breadth of a bifacial flake.

That concludes the analysis of this graph. However, you might have some follow-up questions. For example, is the stronger correlation in flakes caused by the outliers? If so, would this effect disappear if the outliers were excluded from the analysis? What are the pros and cons of excluding the outliers versus retaining them in the dataset? You can now proceed with reading the paper with these new questions in mind.

<u>Section Resources</u>

- [Misleading Graphs](#)- Statistics How To

- [How to interpret scientific & statistical graphs](#)- Vanderbilt University

- [Biology 101: How to Understand Graphs](#)-Nucleus Biology Youtube

- [How to Interpret Boxplots](#)-LabXchange

- [How to Interpret Histograms](#)-LabXchange

- [Scatterplots: Using, Examples, and Interpreting](#)- Statistics by Jim

- [How to Make Good Figures for Scientific Papers](#)- Simplified Science Publishing

- [Tables and Graphs](#)- The Science of Scientific Writing, MIT

- [Best Practices of Graphs and Charts in Research Papers](#)- Mind the Graph

- Akhilesh, K., & Pappu, S. (2015). Bits and pieces: Lithic waste products as indicators of Acheulean behaviour at Attirampakkam, India. Journal of Archaeological Science: Reports, 4, 226-241.

- Chauhan, P. R. (2010). Metrical variability between South Asian handaxe assemblages: preliminary observations. New perspectives on Old Stones: Analytical approaches to Paleolithic technologies, 119-166.

- High, K., Milner, N., Panter, I., & Penkman, K. E. H. (2015). Apatite for destruction: investigating bone degradation due to high acidity at Star Carr. Journal of Archaeological Science, 59, 159-168.

- Naskar, N., Shaha, C., Ghosh, A., & Gangopadhyay, K. (2024). Study on distribution of radionuclides in soil and pottery samples of archaeological sites of eastern India. Journal of Radioanalytical and Nuclear Chemistry, 333(3), 1585-1595.

- Pokharia, A. K., Agnihotri, R., Sharma, S., Bajpai, S., Nath, J., Kumaran, R. N., & Negi, B. C. (2017). Altered cropping pattern and cultural continuation with declined prosperity following abrupt and extreme arid event at 4,200 yrs BP: Evidence from an Indus archaeological site Khirsara, Gujarat, western India. PloS one, 12(10), e0185684.

8 What software should you use?

We have already discussed software options for data collection and management in Data Collection, Storage, and Management section. This section will focus on software for Data Analysis and Graphical Presentation. You can see below the comparison between three of the most common Data Analysis softwares in use in the Social Sciences today:

Microsoft Excel	SPSS	R
User- friendly interface	Interface very similar to Excel	User interface can be intimidating
Doesn't require any coding	Doesn't require any coding	Requires coding (but replicating analyses is quicker as a result)
Can handle large datasets with analyses and graphs being made directly into the datasheet	Can handle large datasets with analyses and graphs being made directly into the datasheet	Handles large datasets especially well without the risk of accidentally losing or changing data since analysis done is separately from datasheet
Can handle basic statistical analyses and graphs well	Can handle basic statistical analyses and graphs well	More advanced statistical methods + more customisation options!
Many online resources available	Few online resources available	Many online resources available
Many options for customisation of graphs	Some options for customisation of graphs	Most options for customisation of graphs out of the three
Not free to download and use	Not free to download and use	Free and open-source

When thinking about your options, consider:

- Your needs, e.g., what kinds of tests do you need to perform?;

- Your skill level, e.g. can you code? If not, will it be worth your time to learn? (coding is time-intensive up-front, but it might be an investment into your skill development and long-term professional goals);

- Your resources, e.g., do you have access to a device with the software? Does this device have the processing power to run the kind of tests you need? Etc..

While all three softwares provide options to create and customise graphs, researchers sometimes find that they need additional tools to have graphs and figures that do justice to the data they are trying to present. Two popular softwares used for this purpose are Adobe Illustrator (https://www.adobe.com/in/products/illustrator.html), and Inkscape (https://inkscape.org/), both of which are powerful vector graphics editors. You can use this Inkscape beginners guide to help you get started: https://inkscape-manuals.readthedocs.io/en/latest/.

Realistically speaking, you will most likely have to use different softwares in combination to suit your needs specifically. At the end of the day, knowing which methods to apply, and having a robust understanding of the statistical logic behind the method, is far more important than the softwares you use.

While we are on the topic of helpful academic software, you might want to (if you aren't already) use a citation manager. Citation/reference managers help you keep track of all the reading materials you refer to over the course of your research, making it much easier for you to revisit relevant papers later on. The biggest advantage of using a citation manager, however, is the huge amount of time it saves you, by automatically creating citations and bibliographies in the format of your choosing. If you are still on the fence, I should mention that unlike the other softwares discussed in this section, most of the commonly used reference softwares have a very small learning curve! (yes, even the free ones!) Check out this University of Pennsylvania resource for a thorough breakdown of each of the most popular options: https://guides.library.upenn.edu/citationmgmt.

In conclusion... please use a citation manager! I am begging you! You don't have to waste hours of your life manually inputting and editing references. You deserve better!

9   Other Useful Resources

- Learning Statistics with R- Navarro

- Statistics How To

- Working with Archaeological Data- American Schools of Overseas Research (ASOR)

- Excel Statistical functions (reference)

- Discovering Statistics by Professor Andy Field

- Carlson, D. L. (2017). Quantitative methods in archaeology using R. Cambridge University Press.

- SAA QUANTARCH: Quantitative Methods & Statistical Computing in Archaeology Facebook group for community help

Master-List of Document Resources

- The Scientific Process- American Museum of Natural History
  https://www.amnh.org/explore/videos/the-scientific-process

- Sampling Strategies https://la.utexas.edu/users/denbow/labs/survey.htm

- Archaeological Sampling Techniques-Dig It With Raven Youtube
  https://www.youtube.com/watch?v=ZxrQeqPi-Nw

- Top 5 Data Management Tips for Graduate Students- University of Wisconsin-Madison
  https://researchdata.wisc.edu/news/top-5-data-management-tips-for-graduate-students/

- Research Data Management Support- Utrecht University
  https://www.uu.nl/en/research/research-data-management/guides/
  storing-and-preserving-data

- Which Statistics Test Should I Use?- Social Science Statistics
  https://www.socscistatistics.com/tests/what_stats_test_wizard.aspx

- Post Hoc Definition and Types of Tests- Statistics How To https://www.statisticshowto.
  com/probability-and-statistics/statistics-definitions/post-hoc/

- Using Post Hoc Tests with ANOVA- Statistics by Jim
  https://statisticsbyjim.com/anova/post-hoc-tests-anova/

- Understanding Hypothesis Testing- Geeksforgeeks
  https://www.geeksforgeeks.org/understanding-hypothesis-testing/

- Hypothesis Testing- La Trobe University
  https://latrobe.libguides.com/maths/hypothesis-testing

- Hypothesis testing- zedstatistics Youtube
  https://www.youtube.com/watch?v=8JIe_cz6qGA

- Intro to Hypothesis Testing in Statistics - Hypothesis Testing Statistics Problems & Examples- Math and Science Youtube https://www.youtube.com/watch?v=VK-rnA3-41c

- ANOVA, ANCOVA, MANOVA and MANCOVA: Understand the difference- Stat Pharm Youtube https://www.youtube.com/watch?v=Q116ZnLy5uA

- ANOVA, ANCOVA, MANOVA & MANCOVA(example)- Research Methodology Advanced Tools Youtube https://www.youtube.com/watch?v=S43Vwr6ofMY

- Statistical Soup: ANOVA, ANCOVA, MANOVA, & MANCOVA- Stats Make Me Cry Consulting
  https://www.statsmakemecry.com/smmctheblog/stats-soup-anova-ancova-manova-mancova

- Principal Component Analysis (PCA)- Steve Burton Youtube
  https://www.youtube.com/watch?v=fkf4IBRSeEc

- Principal Component Analysis (PCA) in Excel
  https://help.xlstat.com/6776-principal-component-analysis-pca-excel

- A Comprehensive Guide to Cluster Analysis: Applications, Best Practices and Resources- Displayr
  https://www.displayr.com/understanding-cluster-analysis-a-comprehensive-guide/

- Tutorial on MCA - Multiple Correspondence Analysis - with R- François Husson Youtube
  https://www.youtube.com/watch?v=0z30RDikYaw

- Type I & Type II Errors — Differences, Examples, Visualizations- Scribbr
  https://www.scribbr.com/statistics/type-i-and-type-ii-errors/

- Misleading Graphs- Statistics How To
  https://www.statisticshowto.com/probability-and-statistics/
  descriptive-statistics/misleading-graphs/

- How to interpret scientific & statistical graphs- Vanderbilt University
  https://biostat.app.vumc.org/wiki/pub/Main/TheresaScott/Interpret.Graphs.
  TAScott.handout.pdf

- Biology 101: How to Understand Graphs-Nucleus Biology Youtube
  https://www.youtube.com/watch?v=JXejcX1nQs4

- How to Interpret Boxplots-LabXchange
  https://www.labxchange.org/library/items/lb:LabXchange:d8863c77:html:1

- How to Interpret Histograms-LabXchange
  https://www.labxchange.org/library/items/lb:LabXchange:10d3270e:html:1

- Scatterplots: Using, Examples, and Interpreting- Statistics by Jim
  https://statisticsbyjim.com/graphs/scatterplots/

- How to Make Good Figures for Scientific Papers- Simplified Science Publishing

  https://www.simplifiedsciencepublishing.com/resources/
  how-to-make-good-figures-for-scientific-papers

- Tables and Graphs- The Science of Scientific Writing, MIT

  https://web.mit.edu/7.021/www/lectures/TablesGraphs_MO.pdf

- Best Practices of Graphs and Charts in Research Papers- Mind the Graph

  https://mindthegraph.com/blog/graphs-charts-tips-research-paper/

- Reporting Results of Common Statistical Tests in APA Format- Psychology Writing Center, University of Washington

  https://psych.uw.edu/storage/writing_center/stats.pdf

- Inkscape Beginner's Guide

  https://inkscape-manuals.readthedocs.io/en/latest/

- Citation Management Tools: Overview- University of Pennsylvania

  https://guides.library.upenn.edu/citationmgmt

- Akhilesh, K., & Pappu, S. (2015). Bits and pieces: Lithic waste products as indicators of Acheulean behaviour at Attirampakkam, India. Journal of Archaeological Science: Reports, 4, 226–241.

- Albers, C. J., Kiers, H. A., & van Ravenzwaaij, D. (2018). Credible confidence: A pragmatic view on the frequentist vs Bayesian debate. Collabra: Psychology, 4(1), 31.

- Banning, E. B. (2021). Sampled to death? The rise and fall of probability sampling in archaeology. American Antiquity, 86(1), 43–60.

- Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis.

- Carey, Stephen Sayers, and S. S. Carey. A beginner's guide to scientific method. Belmont, CA: Wadsworth Publishing Company, 1994.

- Chauhan, P. R. (2010). Metrical Variability Between South Asian Handaxe Assemblages: Preliminary Observations. In S. Lycett & P. Chauhan (Eds.), New Perspectives on Old Stones: Analytical Approaches to Paleolithic Technologies (pp. 119–166). Springer. https://doi.org/10.1007/978-1-4419-6861-6_6

- Dasari, K., Acharya, R., Das, N., & Reddy, A. (2012). A standardless approach of INAA for grouping study of ancient potteries. Journal of Radioanalytical and Nuclear Chemistry, 294(3), 429–434.

- Drennan, R. D. (2010). Statistics for archaeologists. Springer.

- Efron, B. (2005). Bayesians, frequentists, and scientists. Journal of the American Statistical Association, 100(469), 1–5.

- Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O'Connor, J. P. B., & Price, G. J. (2022). Understanding the Differences Between Bayesian and Frequentist Statistics. International Journal of Radiation Oncology, Biology, Physics, 112(5), 1076–1082. https://doi.org/10.1016/j.ijrobp.2021.12.011

- Green, L. (2007). Analysis of archaeological sampling methods using the complete surface data from the Pirque Alto site in Cochabamba, Bolivia. https://www.uwlax.edu/globalassets/offices-services/urc/jur-online/pdf/2007/green.pdf

- Gower, Barry. Scientific method: A historical and philosophical introduction. Routledge, 2012.

- Hackenberger, B. K. (2019). Bayes or not Bayes, is this the question? Croatian Medical Journal, 60(1), 50.

- High, K., Milner, N., Panter, I., & Penkman, K. (2015). Apatite for destruction: Investigating bone degradation due to high acidity at Star Carr. Journal of Archaeological Science, 59, 159–168.

- Joglekar, P. P. (2014). Research Methodology for Archaeology Students. Gayatri Sahitya.

- Naskar, N., Shaha, C., Ghosh, A., & Gangopadhyay, K. (2024). Study on distribution of radionuclides in soil and pottery samples of archaeological sites of eastern India. Journal of Radioanalytical and Nuclear Chemistry, 333(3), 1585–1595. https://doi.org/10.1007/s10967-023-09214-7

- Orton, C. (2000). Sampling in archaeology. Cambridge University Press.

- Otárola-Castillo, E., & Torquato, M. G. (2018). Bayesian Statistics in Archaeology. Annual Review of Anthropology, 47(Volume 47, 2018), 435–453. https://doi.org/10.1146/annurev-anthro-102317-045834

- Otárola-Castillo, E., Torquato, M. G., & Buck, C. E. (2023). The bayesian inferential paradigm in archaeology. Handbook of Archaeological Sciences, 2, 1193–1209.

- Pokharia, A. K., Agnihotri, R., Sharma, S., Bajpai, S., Nath, J., Kumaran, R. N., & Negi, B. C. (2017). Altered cropping pattern and cultural continuation with declined prosperity following abrupt and extreme arid event at 4,200 yrs BP: Evidence from an Indus archaeological site Khirsara, Gujarat, western India. PLOS ONE, 12(10), e0185684. https://doi.org/10.1371/journal.pone.0185684

- Ratan, S. K., Anand, T., & Ratan, J. (2019). Formulation of Research Question - Stepwise Approach. Journal of Indian Association of Pediatric Surgeons, 24(1), 15–20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6322175/

- Ravisankar, R., Naseerutheen, A., Chandrasekaran, A., Bramha, S., Kanagasabapathy, K., Prasad, M., & Satpathy, K. (2014). Energy dispersive X-ray fluorescence analysis of ancient potteries from Vellore District Tamilnadu, India with statistical approach. Journal of Radiation Research and Applied Sciences, 7(1), 44–54.

- Schiffer, M. B., Sullivan, A. P., & Klinger, T. C. (1978). The design of archaeological surveys. World Archaeology, 10(1), 1–28.

- Seetha, D., & Velraj, G. (2016). Characterization and chemometric analysis of ancient pot shards trenched from Arpakkam, Tamil Nadu, India. Journal of Applied Research and Technology, 14(5), 345–353. https://doi.org/10.1016/j.jart.2016.08.002

- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

- Woodman, P. C. (1981). Sampling Strategies and Problems of Archaeological Visibility. Ulster Journal of Archaeology, 44/45, 179–184.