Analysis of Lab Standards for Stable Isotopes of Tooth Enamel Research

Ryan Smith

Ryne Bevan

Client: Dr. Suzanne Birch

UGA Anthropology Professor

UGA Paleoecology Lab Director

Stat 5020W

Introduction

Dr. Suzanne Birch as an anthropology professor and paleoecology lab director at UGA. As a lab director, she has conducted her own study on the tooth enamel of various animal fossils since 2016. The purpose of her study is to get a better understanding of these animals through analysis of delta carbon 13 ($\delta^{13}C$) and delta oxygen 18 ($\delta^{18}O$) values of tooth enamel samples, which are measurements of the ratio of the stable isotopes of each element, reported in parts per thousand. The tooth enamel samples have been gathered from various archeological dig sites across the world.

The process for analyzing these tooth enamel samples involves treating each sample with bleach to remove impurities, then treating them with an acid solution to offset the bleach. After the samples have been treated, they are run through one of two different machines: the gas bench and the automated trace gas pre-concentrator (PreCon). These machines measure the $\delta^{13}C$ and $\delta^{18}O$ values of the samples. The $\delta^{13}C$ and $\delta^{18}O$ values are then recorded, and the analysis of these values can form a basis for inference on animal diet.

To ensure the machines are running correctly and lab protocols are being properly followed, lab standards are run in regular intervals. These lab standards can be thought of as control samples. They are run through the machines, with their $\delta^{13}C$ and $\delta^{18}O$ values being measured and checked to ensure they maintain a consistent value.

There are two standards run by default when using the gas bench or PreCon, called ‘machine’, or ‘external’, standards. Along with those, Dr. Birch also enforces her own lab standards. These standards are tooth enamel samples themselves, from archeological cow (AC), modern cow (MC), and archaeological elephant (AE)) fossils. The lab standards go through the same acid/bleach treatment process as the study samples.

Dr. Birch has assigned us the task of analyzing her lab standards, with the purpose of answering three questions:

1. Are the $\delta^{13}C$ and $\delta^{18}O$ values of the lab standards different from year to year? If they are, it could change how Dr. Birch interprets the rest of the data in her study.
2. Dr. Birch also analyzed a batch of untreated standards. Are the $\delta^{13}C$ and $\delta^{18}O$ values of the untreated standards significantly different from the treated standards? If they aren't, she would no longer have to waste time with the treatment process.
3. Is there a relationship between the type of machine used with $\delta^{13}C$ and $\delta^{18}O$ values? If there is a significant relationship, that could also change how her sample data is interpreted.

The available data includes the actual dig site samples, Dr. Birch's lab standards, and external/machine standards compiled together in separate spreadsheets corresponding to the dig site the samples came from and the date the tests were run. Each spreadsheet contains the $\delta^{13}C$ values and $\delta^{18}O$ values of the samples and standards. Before analysis, we filtered the lab standard data out of each spreadsheet and sorted them into groups by standard type (AC, MC, AE) and testing date. There are 151 AC standard observations, 88 MC standard observations, 61 AE standard observations.

EDA

Our exploratory data analysis is focused on the distributions of $\delta^{13}C$ and $\delta^{18}O$ values for each lab standard type. The distributions will act as a precursor to whether the $\delta^{13}C$ and $\delta^{18}O$ values are affected based on year or machine type, due to the fact that if the values were not affected, we would expect to see approximately symmetrical distributions. We analyzed the distribution for each combination of lab standard type and $\delta^{13}C/\delta^{18}O$ values using violin plots. We also included some basic summary statistics to give further insight on the data.

The first set of violin plots, displayed in Figure 1, show the distribution of $\delta^{13}C$ values for the Archeological Cow, Modern Cow, and Archeological Elephant. The important thing to note in these graphs is that each of them are not as symmetrical as one would expect if the $\delta^{13}C$ values were consistent. If the $\delta^{13}C$ were relatively equal across the entire period of Dr. Birch's study, the distributions would be approximately symmetrical. The AC and AE plots appear to be skewed to the left, while the MC plot is centered, though the tails span a distance longer than we would expect. We can see that there are several observations that stray relatively far beyond the IQR in each graph, as well as several outlying observations. The skewness and density of the upper and lower quartiles for each distribution is potentially telling of a difference in $\delta^{13}C$ values throughout the study.

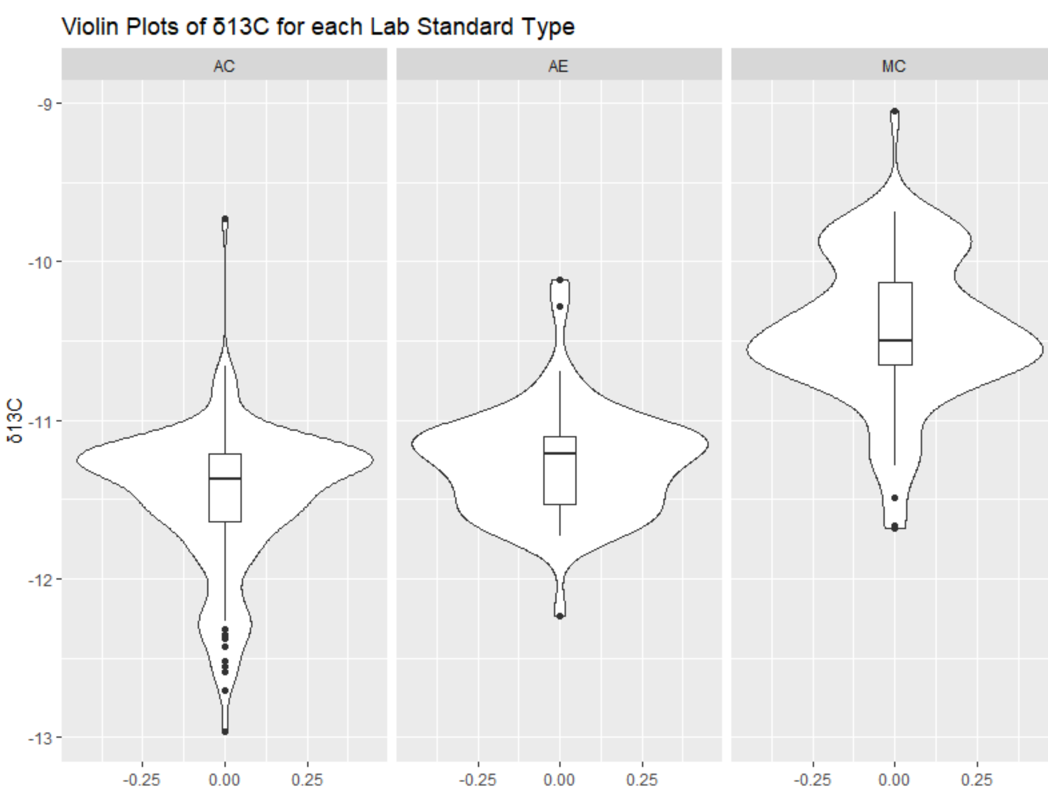


Figure 1

Summary Statistics for AC, MC, and AE $\delta^{13}C$ values

	Min	1st quartile	median	mean	3rd quartile	Max	n
AC	-12.96	-11.64	-11.38	-11.50	-11.22	-9.73	151
MC	-11.68	-10.66	-10.52	-10.46	-10.24	-9.05	88
AE	-12.23	-11.53	-11.21	-11.27	-11.10	-10.11	61

The second set of violin plots, displayed in Figure 2, are the $\delta^{18}O$ values for the AC, MC, and AE lab standards. Like the $\delta^{13}C$ violin plots, these don't look as symmetrical as one would expect if the $\delta^{18}O$ values had remained consistent throughout the study. The AC and AE plots are again skewed to the left, while the MC plot's tails are wider and span further than a symmetrical distribution. We also again see a large density of values spanning outside of the IQR of each distribution, along with several outlying values, especially in the AC distribution. Again, this is potentially telling of a difference in $\delta^{18}O$ throughout the study.

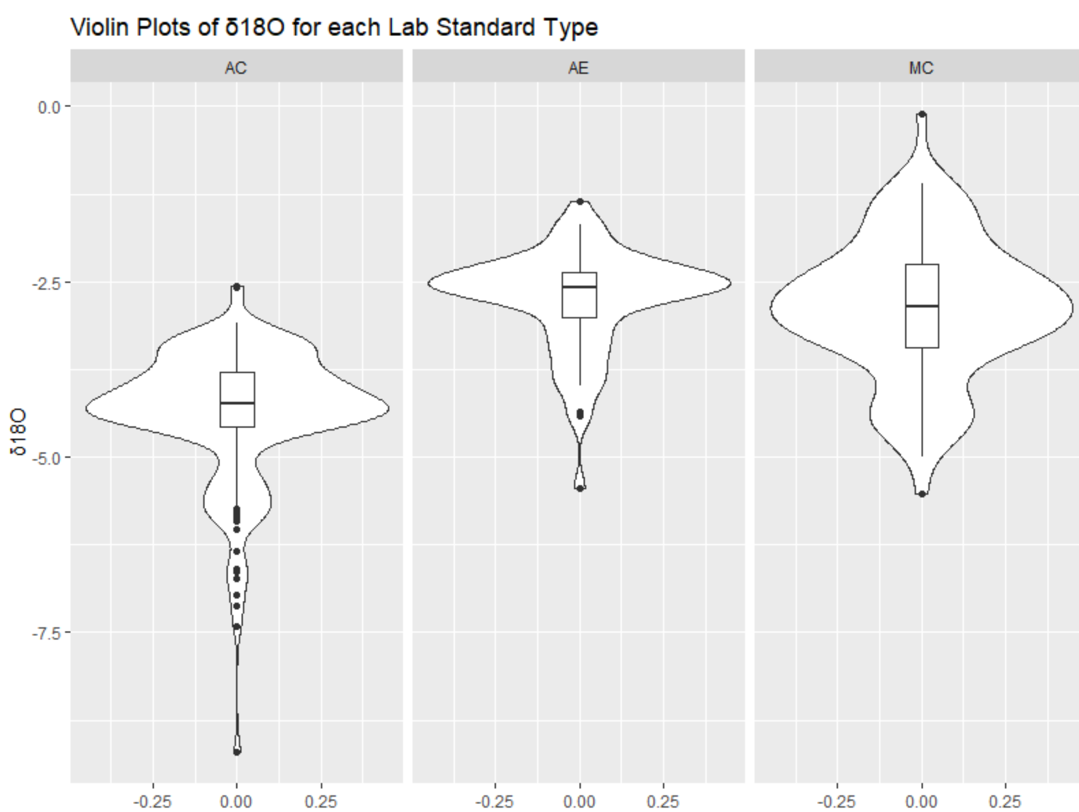


Figure 2

Summary Statistics for AC, MC, and AE δ^{18}O values

	Min	1st Quartile	Median	Mean	3rd quartile	Max	n
AC	-7.420	-4.565	-4.220	-4.346	-3.775	-2.560	151
MC	-5.530	-3.453	-2.780	-2.856	-2.230	-.110	88
AE	-5.450	-3.020	-2.590	-2.748	-2.360	-1.360	61

Lastly, in Figure 3 are the scatterplots of δ^{13}C and δ^{18}O for Archeological Cow, Modern Cow, and Archeological Elephant, respectively. These plots are interesting because if the δ^{13}C and δ^{18}O values remained consistent throughout the study, we would expect to see the values distributed in a small clump in the middle of each plot. We do not see that here, and in fact we can observe possible trends in each plot, implying there may be a correlation between the δ^{13}C values and δ^{18}O values. This is unexpected, as Dr. Birch did not inform us of a possible correlation between the two. If the δ^{13}C values and δ^{18}O values have changed throughout the study, it may be worth investing if an unusual δ^{13}C value corresponds with an unusual δ^{18}O value.

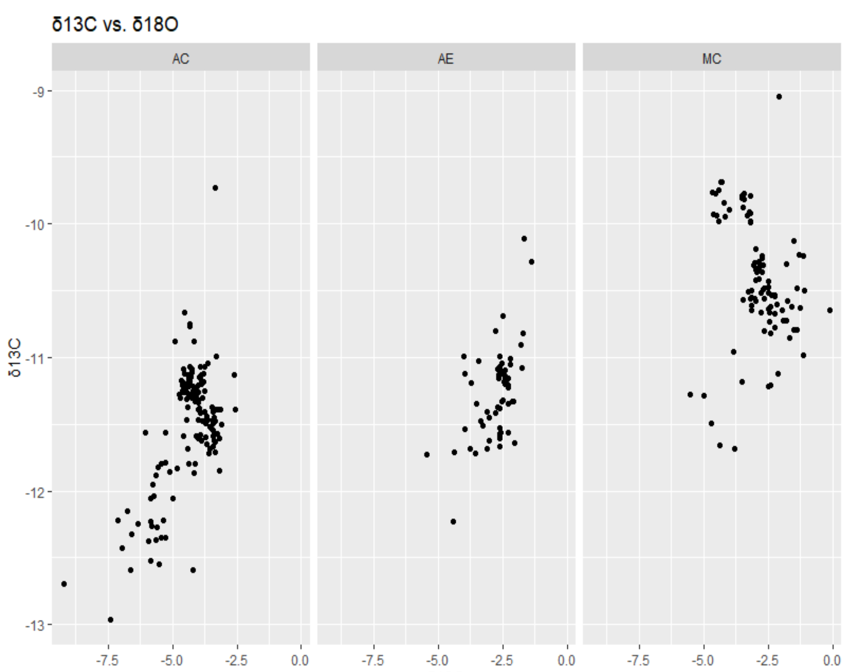


Figure 3

Analysis: $\delta^{13}C$ and $\delta^{18}O$ Values versus Year

Our EDA has given us reason to further investigate the possibility that the lab standard values have not been consistent throughout the course of Dr. Birch's study. Foremost, Dr. Birch was interested in knowing if the $\delta^{13}C$ and $\delta^{18}O$ values were significantly different from a year to year basis. To analyze this, we would essentially be comparing the difference in means between multiple levels of a factor. In this case the factor is 'year', and the levels are the different years in which Dr. Birch has conducted her study and recorded lab standard $\delta^{13}C$ and $\delta^{18}O$ values.

There are a few different statistical methods to test the difference in means between levels of a factor. We first decided to fit regression models with $\delta^{13}C$ and $\delta^{18}O$ values as the response variable and levels of year as the explanatory variable for each type of lab standard. Regression models can be fit with categorical variables, or in this case a single categorical factor with multiple levels, where the coefficient for each level is multiplied by a binary 'dummy' variable in the linear function to calculate a predicted value. After fitting a regression model, t-tests are ran on the coefficients corresponding to level of year to test the significance of the effect of each year on the predicted $\delta^{13}C$ or $\delta^{18}O$ values.

Before we began our test, we created dot plots for levels of year versus $\delta^{13}C$ and $\delta^{18}O$ values for each lab standard type to visualize the data we would be testing. Figure 4 displays the plots for $\delta^{18}O$, and Figures 5 displays the plots for $\delta^{13}C$.

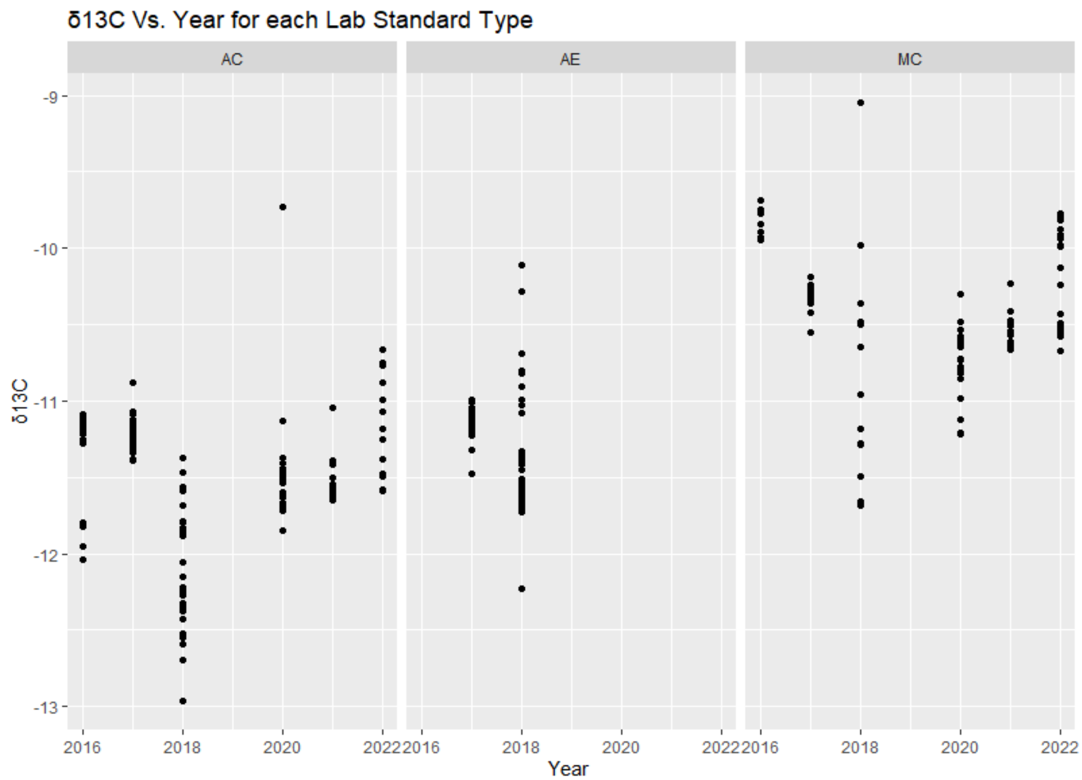


Figure 4

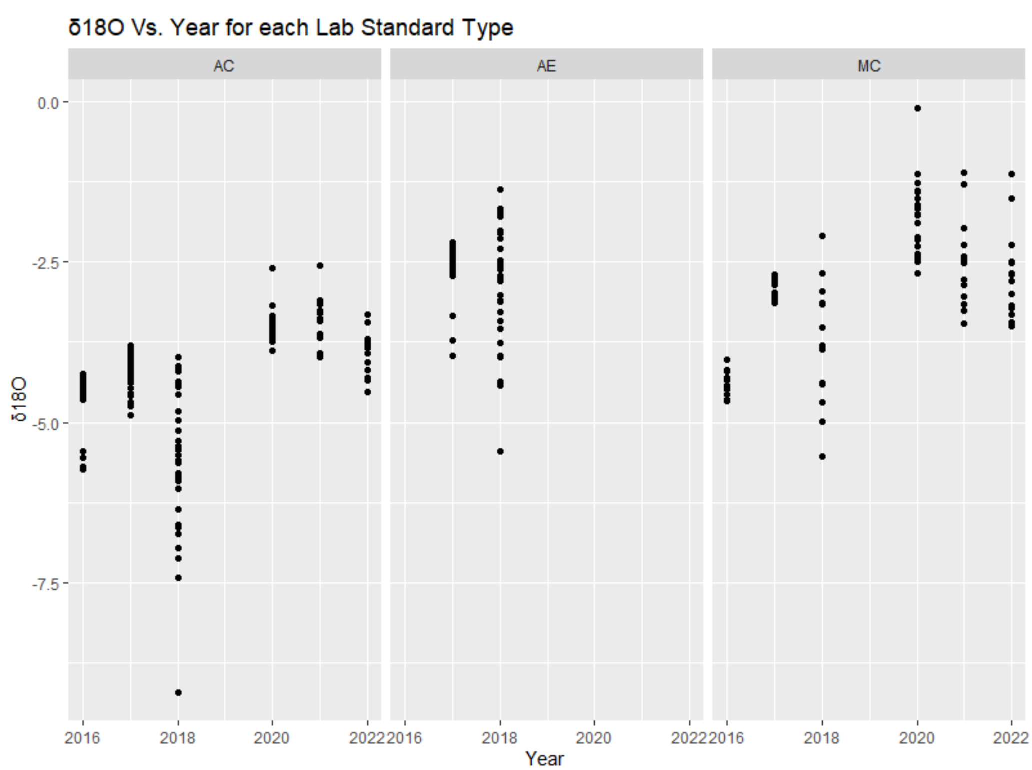


Figure 5

Assumptions when fitting a regression model to categorical variables include the assumption of homoscedasticity. This means that the variances of δ^{13}C and δ^{18}O values for each level of year must be approximately equal. As seen in the plots above, 2018 observations of δ^{13}C and δ^{18}O for each lab standard type appear to have significantly different variances. To be certain, we ran the Levene Test on levels of year for δ^{13}C and δ^{18}O across each lab standard type, which tests if the variances of different groups are equal. The resulting p-values from each Levene Test were low enough to reject the null hypothesis for each test, and conclude that the variance of δ^{13}C and δ^{18}O values between levels of year were significantly different for each lab standard type. This means that we would fail to meet the assumption of homoscedasticity for the regression model to test the difference in mean values of δ^{13}C and δ^{18}O for levels of year.

There are other methods of testing difference in means between groups, one specifically being the one-way ANOVA test. Though the one-way ANOVA test has the same assumption of homoscedasticity, another variation, called Welch's ANOVA, can be used on groups with unequal or unknown variances. The null hypothesis for Welch's ANOVA, similar to the regression model test before, is that mean δ^{13}C or δ^{18}O values across all levels of year are equal. The alternative hypothesis is that at least one is not equal. We ran Welch's ANOVA test for each of the lab standard types, and the results are listed below (Since AE only had two levels of year, Welch's two-sample t-test is ran by default):

	AC	MC	AE
δ^{13}C F-value	33.027	94.763	7.5607
δ^{13}C p-value	<.0000001	<.00000001	.008868
δ^{18}O F-value	62.364	94.763	1.984
δ^{18}O p-value	<.00000001	<.0000001	.1644

Five of our resulting p-values calculated from the Welch's ANOVA test are small enough to present significant evidence that we should reject the null hypothesis and conclude that the mean value of δ^{13}C or δ^{18}O is not equal for at least one level of year. Only the p-value corresponding to δ^{18}O for AE was large enough to fail to reject the null hypothesis. Based on these results, we can say that the level of year likely does have an effect on δ^{13}C/ δ^{18}O values. However, while the ANOVA test tells us that at least one of the levels of year has a different mean value of δ^{13}C or δ^{18}O than the other levels, it does not tell us which specific level. Therefore, a post-hoc test is required.

For our post-hoc test, we used the Tukey HSD test, which essentially functions as a simultaneous t-test between every level of year. The results from this tell us which specific levels of Year the δ^{13}C or δ^{18}O mean values differ from one another.

After running the Tukey HSD test for every combination of δ^{13}C, δ^{18}O and lab standard type, we observed some interesting results. As we suspected from looking at the dot plots in Figure 5, 2018 stuck out among the other levels of year, as it had significantly low corresponding p-values for a large majority of tests between other levels of year for each combination of lab standard type and δ^{13}C/δ^{18}O value. There were also a few other significantly low p-values corresponding to various tests between the other levels of year for different combinations of lab standard type and $\delta^{13}C/\delta^{18}O$ value, though no patterns were observed for any other specific year.

These results were presented to Dr. Birch, and she suggested we investigate the 'external', or 'machine' standards by year to see if they followed the same pattern. If the $\delta^{13}C$ and $\delta^{18}O$ values for the machine standards did not follow the same pattern, and there were no significant differences in values between years, then her own lab standards may be at fault. Otherwise, there may be an underlying factor causing the difference in $\delta^{13}C$ and $\delta^{18}O$ values.

There are two types of machine standards, labeled as 'Fisher' and 'A1296'. We again fit Welch's ANOVA to each combination of machine standard type and $\delta^{13}C/\delta^{18}O$ to test a difference in mean $\delta^{13}C$ or $\delta^{18}O$ values between levels of year. The ANOVA results are listed in the table below, as well as dot plots in Figure 6 and Figure 7.

Welch's ANOVA p-values for Difference in Mean Between Years (Machine Standards)

	$\delta^{13}C$	$\delta^{18}O$
Fisher	0.007237	0.9847
A1296	0.003793	0.9899

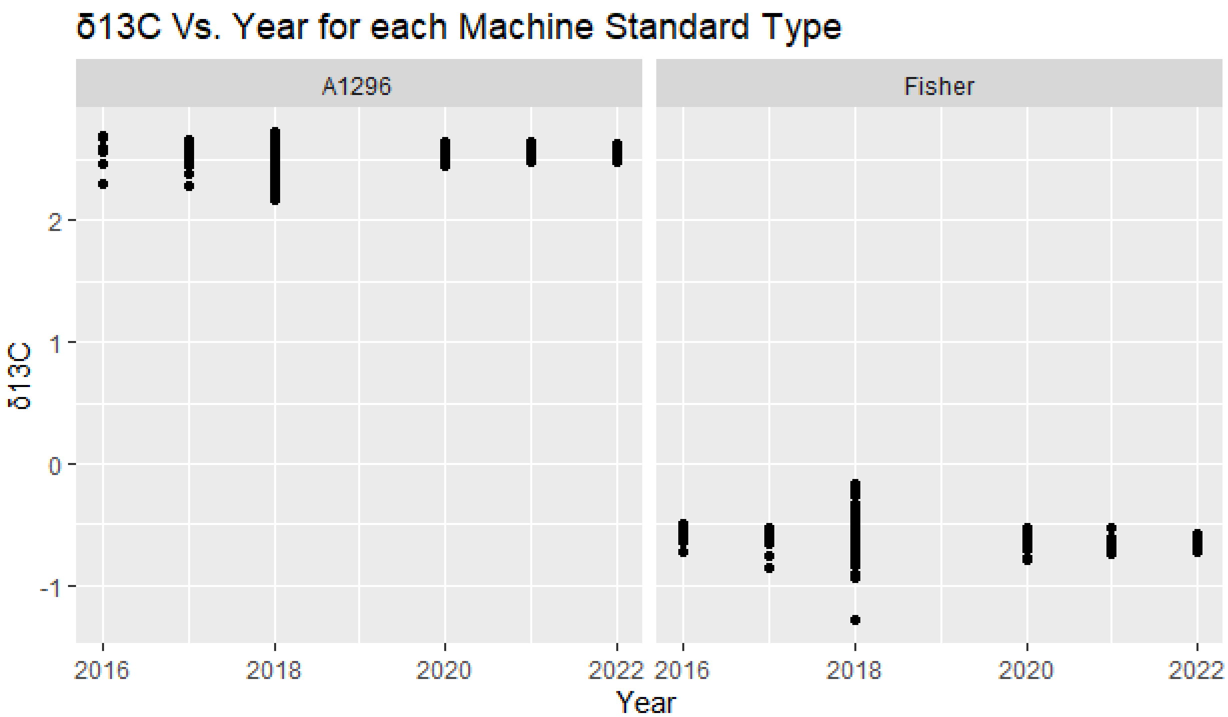


Figure 6

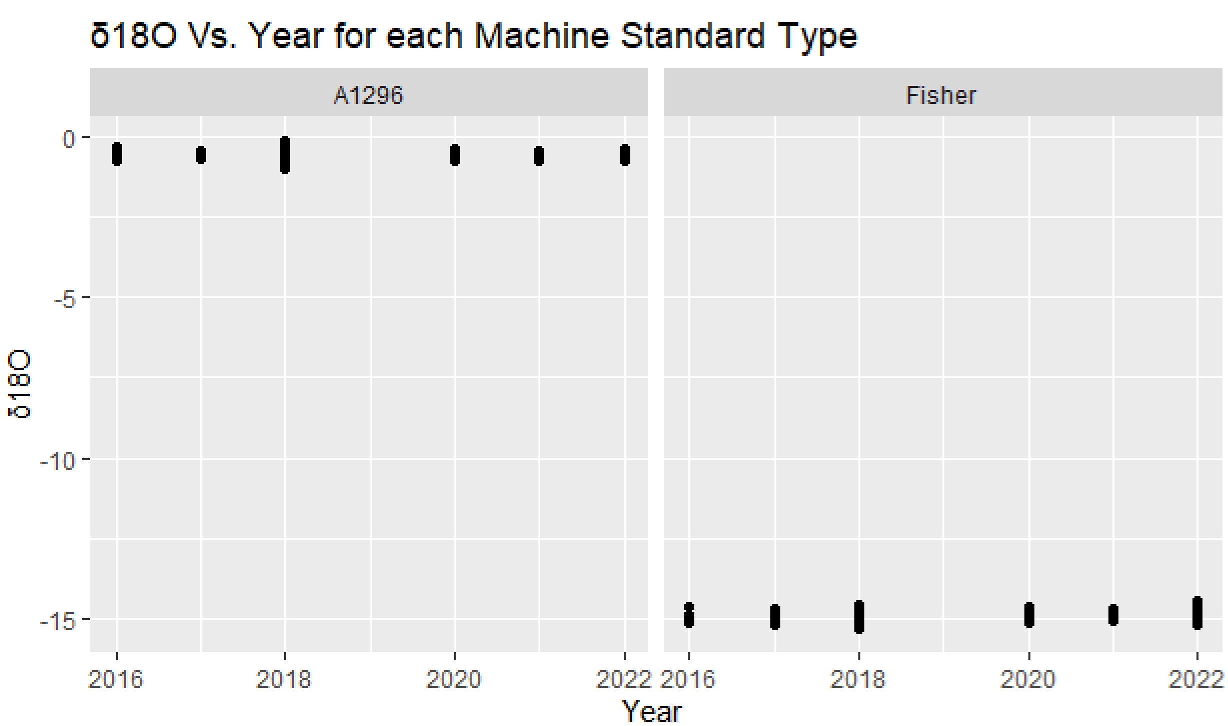


Figure 7

The resulting p-values from the Welch's ANOVA tests suggest rejecting the null hypothesis that there is no difference in mean between levels of year for $\delta^{13}C$ and either machine standard type. We would fail to reject the null hypothesis for $\delta^{18}O$ and either machine type. These results are interesting as they imply an interaction between level of year and $\delta^{13}C$ for each machine standard type, but no interaction between level of year and $\delta^{18}O$ value for each machine standard type. From the $\delta^{13}C$ dot plots, we again see that the year 2018 appears to have higher variance of values compared to the other levels of year.

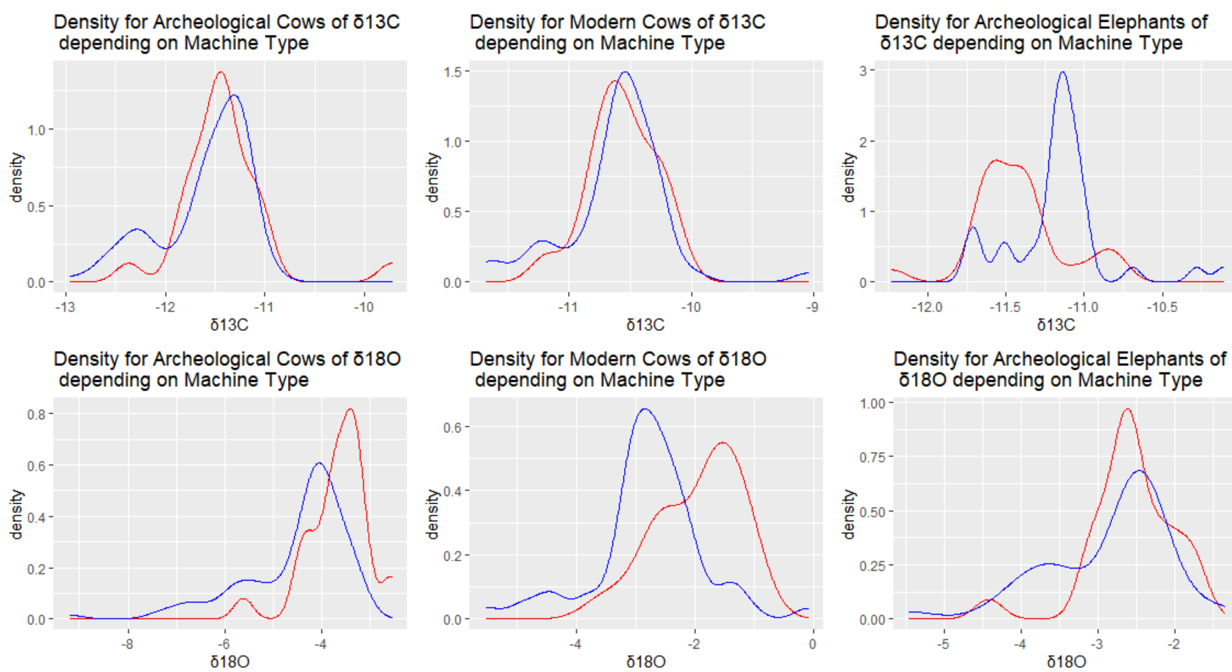
The fact that there is a significant difference in $\delta^{13}C$ values between different levels of year for both machine standard types, as well as a similar pattern in the dot plots compared to the dot plots of Dr. Birch's lab standards, implies that it likely isn't faulty lab standards causing the difference in $\delta^{13}C$ or $\delta^{18}O$ values. There is perhaps another factor at play here, one which we may not have the data to analyze, such as user error when conducting the research.

Analysis: PreCon versus Gas Bench

The next question Dr. Birch posed for us was whether the machine type in which the $\delta^{13}C$ and $\delta^{18}O$ values are produced had an effect on the values. Testing this would be fairly simple, running Welch's t-test on $\delta^{13}C$ and $\delta^{18}O$ values from the PreCon and gas bench machines across each lab standard type. The null hypothesis being that the means of $\delta^{13}C$ and $\delta^{18}O$ values are equal for each machine type, and the alternative being that they are different. Welch's t-tests were run, and the results are given below, as well as density plots comparing $\delta^{13}C$ and $\delta^{18}O$ values over machine type (blue is gas bench, red is PreCon).

Two-Sample t-Test Results for Testing $\delta^{13}C$ and $\delta^{18}O$ means from Different Machine Type

	AC $\delta^{13}C$ Vs. Machine Type	MC $\delta^{13}C$ Vs. Machine Type	AE $\delta^{13}C$ Vs. Machine Type	AC $\delta^{18}O$ Vs. Machine Type	MC $\delta^{18}O$ Vs. Machine Type	AE $\delta^{18}O$ Vs. Machine Type
T- statistic	1.79	.56009	3.1817	4.8903	3.858	1.6701
P-value	.08175	.5786	.002476	.000008561	.00522	.1002
95% Confidence Interval	(-0.0244, 0.39288)	(-0.130,0 .2299)	(-0.4298184, -0.0973211)	(0.4695517, 1.1206866)	(0.3931754, 1.2728430)1	(-0.058477, 0.6482520)



From our Welch's t-Test results, we have three p-values that are significant at the alpha = .05, corresponding to: AE $\delta^{13}C$, AC $\delta^{18}O$, MC $\delta^{18}O$. Therefore, we could reject the null hypothesis that the $\delta^{13}C/\delta^{18}O$ mean values are equal for those specific lab standard types, and fail to reject the null for the remaining three. Based on these results and the fact we only have significant evidence to conclude machine type may have an effect on only a certain combination of lab standard type and $\delta^{13}C/\delta^{18}O$ values, we cannot come to a certain conclusion on whether machine type affects $\delta^{13}C$ and $\delta^{18}O$ values. Further investigation is necessary.

Analysis: δ^{13}C and δ^{18}O Values for Treated vs Untreated Lab Standards

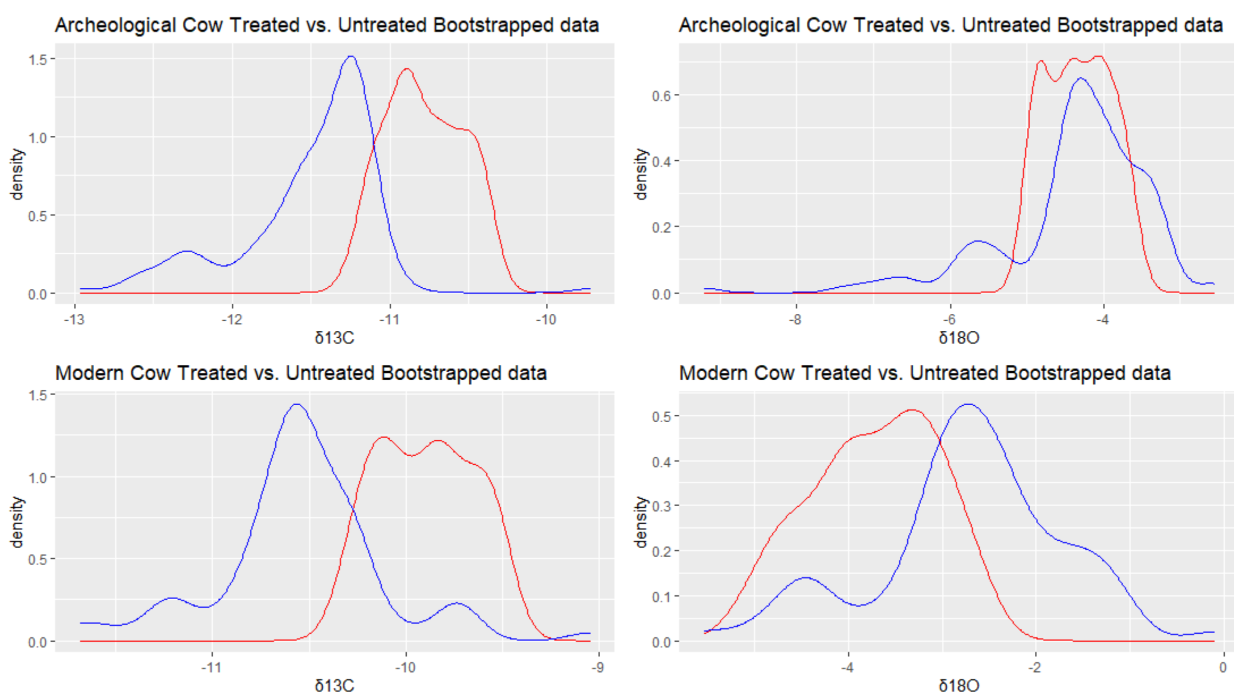
The last question Dr. Birch posed for us was whether the treatment method she applies to her lab standards has an effect on the $\delta^{13}C$ or $\delta^{18}O$ values. As stated in the introduction, the treatment process involves mixing the lab standards with a bleach solution to remove impurities, then applying an acid solution to offset the bleach. Dr. Birch assumes the treatment method has an effect on the $\delta^{13}C$ and $\delta^{18}O$ values, though she wanted us to test if that assumption was true.

Since almost every lab standard analyzed throughout her study was treated, she analyzed a new batch of untreated standards for the purpose of this test. The batch was relatively small though, with only 5 untreated MC and AC standards, along with 10 untreated MC standards from previous runs from years past. Therefore, we would need to apply bootstrapping to the untreated standard observations before we could do meaningful analysis. Bootstrapping simulates the process of additional data collection by resampling the original observations to create a sampling distribution. We also added 'perturbation' to our bootstrapping method, which applies pseudo random noise to each resampled observation in order to create a more realistic distribution.

After our bootstrapped sampling distribution for the untreated lab standards were created, we used Welch's t-test to compare the means between the treated lab standards and untreated lab standards for each combination of lab standard, $\delta^{13}C$, and $\delta^{18}O$. The null hypothesis for the t-tests being the means are equal, and the alternative hypothesis being the means are unequal. The resulting p-values are given in a table below, as well as density plots comparing the distributions (blue is treated, red is untreated).

Welch's t-test Results for Treated Lab Standards vs. Bootstrapped Untreated Lab Standards

	δ^{13}C p-value	δ^{18}O p-value
Archeological Cow	<.00001	.5769
Modern Cow	<.00001	<.00001



From the given p-values, there is significant evidence to reject the null hypothesis for AC δ^{13}C, MC δ^{13}C, and MC δ^{18}O, suggesting a difference in means between treated and untreated for those combinations of lab standard type and δ^{13}C/δ^{18}O values. Only for AC δ^{18}O would we fail to reject the null hypothesis and conclude there is no difference in means between treated and untreated lab standard.

Conclusions

Based on the analysis of Dr. Suzanne Birch's lab standards, we were able to answer three key questions that have implications for her study on stable isotopes of animal fossils.

First, we found that there is a significant difference in the mean $\delta^{13}C$ and $\delta^{18}O$ values of the lab standards from year to year. The year 2018 stands out as having significantly different $\delta^{13}C$ and $\delta^{18}O$ values than other years for almost every lab standard type. These results challenge the assumed consistency and reliability of the lab protocols, and potentially undermine the validity of her research fossil sample data. With that being said, we cannot say for certain what caused the inconsistency in the lab standards as we lack the means to provide further insight with the given data.

Second, we found there may be a potential relationship between the type of machine used and the $\delta^{13}C$ and $\delta^{18}O$ values, though the results varied based on the combination of lab standard type and $\delta^{13}C$ or $\delta^{18}O$. These results could suggest that she should only use one type of machine for her study to control for the potential interaction.

Finally, we determined that there is a significant difference between the mean $\delta^{13}C$ and $\delta^{18}O$ values of treated and untreated standards, suggesting that the treatment process is likely validated in her study. It may be worth replicating this test with a much larger sample of untreated lab standards, as our analysis was conducted on a bootstrapped sample distribution which may not have captured the parameters of the 'true' population well.