MODELING THE PERCENTAGE OF CHEMICAL COMPONENTS IN ROMAN POTTERY

HENGRUI LUO

Abstract. This paper discusses a method of constructing a statistical model (MLR) to model the MnO percentage in the Roman pottery. The author uses methods of linear models and model building in order to make his model valid in predicting the percentage of MnO based on percentages of FeO, MgO, CaO, KO in the same piece of pottery. The author also discussed the diagnostic of the model and further questions to be studied.

1. Introduction

1.1. **Question.** The question I want to investigate is *whether the percentage of MnO contained in the British Roman pottery can be predicted by the percentages of other kinds of chemical components contained in the same piece of pottery.* This problem attracts my interest for various reasons:

(1) Classification of pottery.
It has been verified that the containment of Mn in the pigment of pottery is a special characteristic of a family of Roman pottery.[1]An instant question is to ask whether we can predict the percentage of MnO in Roman Pottery using the percentage of other chemical components in the same piece of pottery.
For example, if a new piece of pottery artifact is obtained from an archaeological excavation, we can detect the percentage of MnO and determine its relation with other chemical components.
Will such a relationship be of a useful indicator of the source of this piece of pottery? If the answer is confirmative, then this kind of evidence can serve as a complement of archaeologists ' professional knowledge.

(2) Inference for composite components.
With some knowledge from chemistry, we know that in nature certain kinds of chemical elements will compose a certain form of composite easily.
For example, it is proposed by archaeologists that in paintings of pottery the Fe and Mn combined to make red-brown color on the surface of the pottery. [2] I want to find out the correlation between certain kinds of elements in order to support this hypothesis. If a linear relationship is founded, then its coefficient of determination can implicate such a relationship.

(3) Inference for climate conditions.
Combined with an expert's knowledge from archaeology, the relation between chemical components in Roman pottery can tell us information about the situation under which the Roman pottery is produced.
For example, the small percentage of certain kind of chemicals may indicate the weather effect. [3] [4]

1.2. **Statistical formulation.** Formally, the problem stated above can be considered as following statistical problems:
The first problem is to choose predictors/regressors in order to predict the response variable MnO from these variables. According to the principal component analysis which is done in [Tubb et.al] and [Mirti et.al], as well as the variable selection study conducted by [Baxter&Jackson]. I chose Fe_2O_3, MgO, CaO, K_2O as predictors to obtain optimal results.
The second problem is to fit a regression model to these predictor variables chosen.

2. Data

2.1. **Data source.** The raw data to be analyzed is from the R dataset pottery in the HSAUR package, which is a faithful record of the data presented in the original paper [Tubb et.al]. According to the archaeologists, the data

[1][Sabbatini et.al] p.121. "The different color of this family of pigments, brown more than red, supports this hypothesis and, more interestingly, XPS showed the presence of Mn on all the samples belonging to this group except one."

[2][Sabbatini et.al] pp.120-121

[3][Sanders]

[4][Tubb et.al]:"Weathering is the breakdown of an existing chemical or mechanical structure, by the processes of erosion and corrosion. In chemical analysis it is important to know how extensive are the effects of corrosion in order to decide which parts of the sample are most representative of the unweathered product."

is obtained from a typical analytical chemical method of examining a burnt piece of pottery in a special kind of solution[5].

This dataset consists of *no observational data* since each indicator number was obtained from analytical method and experiment, thus we can draw a relatively solid conclusion about the association from the data from our regression analysis.

However, the samples are obtained from five kiln sites, so there might be *block-effect* in this dataset which should be brought to our attention.

Also, in order to verify the validity of my model, I collect another data set of the same format from another similar research [Mirti et.al]. I will use my model to check and achieve the goals I mentioned above.

2.2. **Data preprocess.** The raw dataset from R package missed 3 entries in [Tubb et.al], Table 2. Namely, the package failed to record:

```
GA4 17.4,7.48,1.71,1.01,0.40,3.16,0.03,0.084,0.017
C13 12.7,6.69,4.45,0.20,0.22,4.70,0.73,0.394,0.024
C14 12.5,6.44,3.94,0.22,0.23,0.81,0.75,0.177,0.019
```

According to my analysis above, we want to examine whether the block-effect is significant. So I also include the kiln categories obtained from [Tubb et.al]. And we should also notice that the component is expressed in terms of percentage which has been normalized using a linear scale.

(1) From both the scatterplot matrix and the boxplots against the kiln factor, we can see obvious block effect and hence we should take the kiln factor into consideration when modeling.
(2) From the scatterplot matrix, there are rather obvious inter-relationship between each variables, some are rather significant and others are not.
    Among these relationships, we can see some of these relationships are obviously linear while others may have a curved relationship which requires our further considerations of transformations.
    (a) The factor of component BaO seems not related to any of the other factor, and its blocking effect is not significant, hence I want to exclude this variable in our further analysis since it seems like a random noise instead of a contributor.[6]
    (b) The factor of component $Fe_2O_3, MgO, K_2O, Al_2O_3$ seems highly correlated with each other pretty highly. They also showed a strong linear relation with the rest of variables individually. So I want to pool them into one group. Also their percentages are relatively high, so it might be a good idea to regard this group of variables as candidates of regressors since they can be easily detected in pottery materials.
    Their percentages in the pottery are relatively large so it might be easier to measure them as predictors.
    (c) The factor of component MnO, TiO_2, Na_2O, CaO have moderate linear relationship with $Fe_2O_3, MgO, K_2O, Al_2O_3$, and they seem uncorrelated with each other too strongly. Therefore I am willing to put them as response variables and build each of them a model.
    Another reason for choosing them as response, as mentioned in the introduction, is that they are relatively scarce in percentage in the pottery. So predicting these variables will achieve our goals mentioned above.
(3) Based on the summary of the data, in terms of strictly positive percentage scale, we can apply the range-rule [7] to conclude that no further transformation is needed for linearity assumption.

3. Methods

I use the scatterplot matrix and boxplots[8] as major tools for exploratory data analysis, in this procedure I determine which variables should be predictors and how they are related to the response variable MnO.

Although a cluster method seems more favorable to archaeologists [Mirti et.al], I think a linear regression might not only offer a better explanation but also present a better prediction.

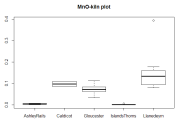
[5][Tubb et.al]pp.153-155: "The method used is a lithium metaborate fusion technique...The solution obtained was examined by atomic absorption spectrophotometry. Lanthanum oxide acts as a releasing agent and helps, as lanthanum ions, to prevent chemical interferences."

[6]Coincidentally, in later researches, the researchers did not measure BaO either. See [Mirti et.al].

[7][Weisberg] p.188

[8][Weisberg] Chap.1

FIGURE 4.1. MnO-kiln plot



I use multiple linear regression[9] as my basic regression tool, and use the transformation tools[10] to make the model satisfy the linear model assumptions.

In all, I derive my model in form of a group indicator model[11] in order to express the blocking effect caused by different kilns mentioned in the data source part.

4. Results

4.1. **Modeling.**

4.1.1. *Overview.* The model I chose is the standard MLR model

$$Y_{MnO,i} = \beta_{Fe_2O_3,i} \cdot X_{Fe_2O_3,i} + \beta_{MgO} X_{MgO,i} + \beta_{CaO} X_{CaO,i} + \beta_{K_2O} X_{K_2O} + e_i$$

$$e_i \overset{i.i.d}{\sim} N(o, \sigma^2)$$

$Y_{MnO,i}$ is the percentage of MnO in the i-th sample observation; $X_{Fe_2O_3,i}$ is the percentage of Fe_2O_3 in the i-th sample observation; $X_{MgO,i}$ is the percentage of MgO in the i-th sample observation;$X_{CaO,i}$ is the percentage of CaO in the i-th sample observation; $X_{K_2O,i}$ is the percentage of K_2O in the i-th sample observation; e_i is the error term. The $\beta_{Fe_2O_3}, \beta_{MgO}, \beta_{CaO}, \beta_{K_2O}$ are the coefficients of this model. In later analysis, we will introduce more coefficients if there is a significant block-effect.

First I looked into the scatterplot matrix of MnO against all these predictors and observed that there is some obvious relationship, some are linear while others are curved. So it is possible to model these relations via multiple linear models.

Secondly, I looked into the boxplots of MnO against the blocking factor kiln. The boxplots showed that we might want to divide the observations into at least 3 groups due to their differences in the percentage containment of MnO.

Thus I divided them into three groups and decided to fit them into sub-MLR models individually. And eventually I will derive a group indicator model by joining them together.

4.1.2. *Group 1: IslandsThorns and AshleyRails kilns.* These two kilns have the lowest percentage of MnO in average. So I fit them into the model including all four predictors first.
The diagnostic of this model seems moderate.

(1) (Linearity)From the residual plot we can see the linearity assumption seems good;
(2) (Homoscedasticity)From the residual plot we can see the homoscedasticity seems good with no obvious pattern; From the output of Brown-Forsythe test we can be sure that the homoscedasticity seems a valid assumption.
(3) (Outliers/Leverage)From the Cook's distance plot, it does not indicate any outliers;
(4) (Normality)From the the normality probability plot, the empirical distribution quantiles seems quite linear;
(5) (Randomness)From the Durbin-Watson test, it does not show significant evidence of non-randomness for such a small sample; From the inverse residual plot, it does not show any transformation is required.
(6) (Lurking)No more lurking variable since there is no special pattern in residual plots and we have already included all the predictors since we already considered the blocking effect caused by the kilns.

The summary is not very satisfying since the t-test of coefficients show that none of the coefficients are significant. One reason is the small sample size of this group; the other possibility is that there are redundant predictors in this model. However, due to its small size, our model is not strongly supported by the dataset we used for modelling. Therefore we have to validate this model on a new dataset.

[9][Weisberg] Chap.3
[10][Weisberg] Chap.8,9
[11][Weisberg] Chap.5

FIGURE 4.2. The Scatter Plot Matrix against Predictors

4.1.3. *Group 2: Gloucester and Caldicot kilns.* These two kilns have the second highest percentage of MnO in average. So I fit them into the model including all four predictors first.
The diagnostic plots seems good considering the sample size.

The diagnostic of this model seems moderate.

(1) (Linearity)From the residual plot we can see the linearity assumption holds and therefore linear model can be applied;
(2) (Homoscedasticity)From the residual plot we can see the homoscedasticity seems good with no obvious pattern; From the output of Brown-Forsythe test we can be sure that the homoscedasticity seems a valid assumption.
(3) (Outliers/Leverage)From the Cook's distance plot, it indicate observation #1 to be a possible outlier;
(4) (Normality)From the the normality probability plot, the empirical distribution quantiles seems quite linear except for the observation #21, which is acceptable due to relatively moderate sample size;
(5) (Randomness)From the Durbin-Watson test, it does not show significant evidence of non-randomness for such a sample; From the inverse residual plot, it does not show any transformation is required.
(6) (Lurking)There could be some lurking variables since there is a slightly cluster pattern in residual plots and we have already include all the predictors.
(7) (Dropping) From the summary of t-tests of each predictor variable, I decide to drop the predictor CaO, K_2O from this model since their coefficients are not significant.

The summary, however, indicates that all regression coefficients are significant while the coefficient of determination is relatively low (around 45%).

One cause could be due to the nature of pottery sample since Fe_2O_3, MgO, CaO, K_2O is not an exhausting list of chemical components contained in the pottery, there could be other kinds of chemical components explaining the MnO percentage in this group. This point can also be deduced from the slightly clustered pattern in residual plot. If we look back to the scatter plot matrix, we can see that there is a slight pattern which indicates possible lurking variables not recorded by the dataset.

Another plausible explanaton of this fact can be detected from the scatterplot matrix. These four predictors are approximately independent but not strictly so, therefore dropping these two predictors makes the rest two predictors Fe_2O_3, MgO highly significant with relatively low coefficient of determination since they are explaining in the same way.

In all, we have a relatively satisfying realistic model with this relatively small sample.

4.1.4. *Group 3: Llanedeyrn kiln.* This group consists of only one kiln, and from the scatterplot matrix and the summary of an exploratory linear model, we can see there is no evident linear relationship between the percentage of MnO and other chemical elements Fe_2O_3, MgO, CaO, K_2O in group 3. So no linear model can be constructed from this group of pottery.

4.1.5. *Conclusion.* We can only fit a linear model to the first two groups of kilns using a group indicator model. From the output of R, the explicit model is:

$$\begin{cases} MnO = 0.009089 - 0.000113 \cdot Fe_2O_3 + 0.020145 \cdot MgO & IslandsThorns, AshleyRails \\ +0.032156 \cdot CaO - 0.009824 \cdot K_2O & \\ MnO = -0.10162 + 0.01704 \cdot Fe_2O_3 + 0.02657 \cdot MgO & Gloucester, Caldicot \end{cases}$$

4.2. **Validation on new data set.** I used new records from [Mirti et.al] Table 2. The 21 observations are from Augusta Praetoria Roman pottery.

From the residual plot obtained from predicting MnO using the group indicator model derived above, we can see that $MnO = -0.10162 + 0.01704 \cdot Fe_2O_3 + 0.02657 \cdot MgO$ is the model which predicts the MnO percentage in the new dataset best.

Thus I am concluding that Augusta Praetoria Roman pottery has similar composition of MnO as those samples from kilns Gloucester,Caldicot.

FIGURE 6.1. Diagnostic Plots for Group 1

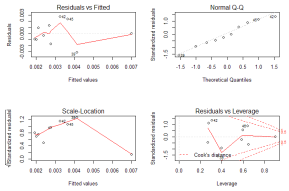


FIGURE 6.2. Inverse Residual Plot for Group 1

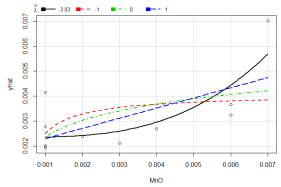
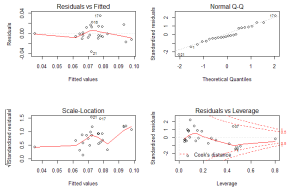


FIGURE 6.3. Diagnostic Plots for Group 2 (full model)



5. DISCUSSION

5.1. **Answer to the archaeological question.** The original method taken by [Tubb et.al] is the principal analysis method for clustering. The conclusion drawn by them is to use iron, magnesium, calcium and potassium as classifying predictors i.e. Fe_2O_3, MgO, CaO, K_2O as a classifiers to build their classifying tree.

However, their method cannot be used to predict the percentage of chemical component in the pottery. Also, although their method successfully tell the pottery from different kilns apart, they did not establish distinct models for each kiln site. Therefore their result cannot be used for future prediction as I did in this paper.

5.2. **Limitation.** One limitation of my study is that I do not include the variable selection as part of my study but borrow from existing results about the choice of predictors. As indicated in [Tubb et.al] , [Mirti et.al] and [Baxter&Jackson], I can make my own choice of variable using principal component analysis or professional knowledge from archaeology. This limtation can be remedied if I include a variable selection using AIC, BIC or other variable selection indicators. Yet in case of archaeologic data, the professional knowledge from experts are also a valuable source. Therefore I will say such a lack of variable selection will not reduce the validity of my model.

5.3. **Further questions.** For further analysis, I will probably look into principle component analysis and principal factor analysis to determine which predictors are most appropriate to be included in our model.

Secondly, I will introduce a formal method of variable selection into my scheme. Instead of merely looking at t-test output, I will introduce step-wise selection or information criteria like AIC, BIC to determine how to drop predictors.

An interesting question is that if we were given a piece of pottery, what kinds of chemical components will serve as an appropriate set of predictors. And if such a group of indicators are specified, an alike model will be very helpful in archaeologists' future work in determining the source of a certain piece of Roman pottery.

My sincere thanks to Ms. Q.Lin, a Ph.D. candidate from Psychology department of *Yale University*, provided generous suggestions and polishments on this paper.

6. APPENDIX

6.1. **Diagnostic for Group 1.**

FIGURE 6.4. Diagnostic Plots for Group 2 (reduced model)

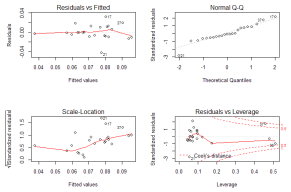


FIGURE 6.5. Inverse Residual Plot for Group 2

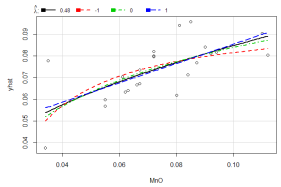
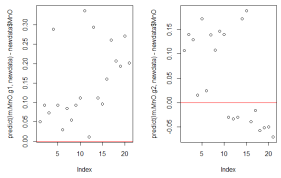


FIGURE 6.6. Residual Plots for Sub-models Group 1 and Group 2



6.2. **Diagnostic for Group 2.**

6.3. **Prediction on new Dataset.** We use the following R-code to predict:

```
newdata=roman2
par(mfrow=c(1,2))
plot(predict(lm.MnO.g1,newdata)-newdata$MnO)
abline(h=0,col="red")
plot(predict(lm.MnO.g2,newdata)-newdata$MnO)
abline(h=0,col="red")
#Good fit!
summary(roman)
summary(roman2)
```

7. DATASET AND CODE

7.1. **Dataset from** [Tubb et.al] **Table 2.** :

	Al2O3	Fe2O3	MgO	CaO	Na2O	K2O	TiO2	MnO	BaO	kiln
1	18.8	9.52	2.00	0.79	0.40	3.20	1.01	0.077	0.015	Gloucester
2	16.9	7.33	1.65	0.84	0.40	3.05	0.99	0.067	0.018	Gloucester
3	18.2	7.64	1.82	0.77	0.40	3.07	0.98	0.087	0.014	Gloucester
4	17.4	7.48	1.71	1.01	0.40	3.16	0.03	0.084	0.017	Gloucester
5	16.9	7.29	1.56	0.76	0.40	3.05	1.00	0.063	0.019	Gloucester
6	17.8	7.24	1.83	0.92	0.43	3.12	0.93	0.061	0.019	Gloucester
7	18.8	7.45	2.06	0.87	0.25	3.26	0.98	0.072	0.017	Gloucester
8	16.5	7.05	1.81	1.73	0.33	3.20	0.95	0.066	0.019	Gloucester
9	18.0	7.42	2.06	1.00	0.28	3.37	0.96	0.072	0.017	Gloucester
10	15.8	7.15	1.62	0.71	0.38	3.25	0.93	0.062	0.017	Gloucester

11	14.6	6.87	1.67	0.76	0.33	3.06	0.91	0.055	0.012	Gloucester
12	13.7	5.83	1.50	0.66	0.13	2.25	0.75	0.034	0.012	Gloucester
13	14.6	6.76	1.63	1.48	0.20	3.02	0.87	0.055	0.016	Gloucester
14	14.8	7.07	1.62	1.44	0.24	3.03	0.86	0.080	0.016	Gloucester
15	17.1	7.79	1.99	0.83	0.46	3.13	0.93	0.090	0.020	Gloucester
16	16.8	7.86	1.86	0.84	0.46	2.93	0.94	0.094	0.020	Gloucester
17	15.8	7.65	1.94	0.81	0.83	3.33	0.96	0.112	0.019	Gloucester
18	18.6	7.85	2.33	0.87	0.38	3.17	0.98	0.081	0.018	Gloucester
19	16.9	7.87	1.83	1.31	0.53	3.09	0.95	0.092	0.023	Gloucester
20	18.9	7.58	2.05	0.83	0.13	3.29	0.98	0.072	0.015	Gloucester
21	18.0	7.50	1.94	0.69	0.12	3.14	0.93	0.035	0.017	Gloucester
22	17.8	7.28	1.92	0.81	0.18	3.15	0.90	0.067	0.017	Gloucester
23	14.4	7.00	4.30	0.15	0.51	4.25	0.79	0.160	0.019	Llanedeyrn
24	13.8	7.08	3.43	0.12	0.17	4.14	0.77	0.144	0.020	Llanedeyrn
25	14.6	7.09	3.88	0.13	0.20	4.36	0.81	0.124	0.019	Llanedeyrn
26	11.5	6.37	5.64	0.16	0.14	3.89	0.69	0.087	0.009	Llanedeyrn
27	13.8	7.06	5.34	0.20	0.20	4.31	0.71	0.101	0.021	Llanedeyrn
28	10.9	6.26	3.47	0.17	0.22	3.40	0.66	0.109	0.010	Llanedeyrn
29	10.1	4.26	4.26	0.20	0.18	3.32	0.59	0.149	0.017	Llanedeyrn
30	11.6	5.78	5.91	0.18	0.16	3.70	0.65	0.082	0.015	Llanedeyrn
31	11.1	5.49	4.52	0.29	0.30	4.03	0.63	0.080	0.016	Llanedeyrn
32	13.4	6.92	7.23	0.28	0.20	4.54	0.69	0.163	0.017	Llanedeyrn
33	12.4	6.13	5.69	0.22	0.54	4.65	0.70	0.159	0.015	Llanedeyrn
34	13.1	6.64	5.51	0.31	0.24	4.89	0.72	0.094	0.017	Llanedeyrn
35	12.7	6.69	4.45	0.20	0.22	4.70	0.73	0.394	0.024	Llanedeyrn
36	12.5	6.44	3.94	0.22	0.23	0.81	0.75	0.177	0.019	Llanedeyrn
37	11.6	5.39	3.77	0.29	0.06	4.51	0.56	0.110	0.015	Caldicot
38	11.8	5.44	3.94	0.30	0.04	4.64	0.59	0.085	0.013	Caldicot
39	18.3	1.28	0.67	0.03	0.03	1.96	0.65	0.001	0.014	IslandsThorns
40	15.8	2.39	0.63	0.01	0.04	1.94	1.29	0.001	0.014	IslandsThorns
41	18.0	1.50	0.67	0.01	0.06	2.11	0.92	0.001	0.016	IslandsThorns
42	18.0	1.88	0.68	0.01	0.04	2.00	1.11	0.006	0.022	IslandsThorns
43	20.8	1.51	0.72	0.07	0.10	2.37	1.26	0.002	0.016	IslandsThorns
44	17.7	1.12	0.56	0.06	0.06	2.06	0.79	0.001	0.013	AshleyRails
45	18.3	1.14	0.67	0.06	0.05	2.11	0.89	0.006	0.019	AshleyRails
46	16.7	0.92	0.53	0.01	0.05	1.76	0.91	0.004	0.013	AshleyRails
47	14.8	2.74	0.67	0.03	0.05	2.15	1.34	0.003	0.015	AshleyRails
48	19.1	1.64	0.60	0.10	0.03	1.75	1.04	0.007	0.018	AshleyRails

7.2. **Dataset from** [Mirti et.al] **Table 2.** :

	Al2O3	Fe2O3	MgO	CaO	Na2O	K2O	TiO2	MnO
1	17.25	8.65	7.56	2.06	1.56	3.72	0.99	0.14
2	19.65	8.62	8.08	2.45	1.63	3.79	1.01	0.12
3	18.10	9.14	7.71	2.33	1.70	3.62	0.97	0.13
4	15.91	6.08	3.89	9.99	1.46	2.98	0.81	0.09
5	18.61	9.51	8.72	2.09	1.61	3.86	1.00	0.12
6	20.82	8.00	4.89	2.94	1.52	3.18	0.99	0.14
7	18.23	8.49	8.12	2.18	1.61	3.76	0.99	0.12
8	18.70	9.21	7.25	2.43	1.60	3.82	0.98	0.14
9	17.81	8.97	8.85	2.53	1.74	3.52	0.96	0.14
10	18.07	9.04	7.44	3.14	1.47	3.77	0.94	0.11
11	15.02	5.88	2.68	12.84	0.78	3.94	0.84	0.10
12	16.09	4.38	0.90	1.03	0.60	1.91	1.33	0.03
13	15.30	6.13	3.29	11.58	0.94	3.35	0.87	0.12
14	17.55	9.09	10.12	2.49	1.35	3.14	0.59	0.15

```
15 19.50  9.68 10.36  2.10 1.28 3.93 0.60 0.15
16 17.77  7.44  3.23  8.18 0.89 2.70 0.80 0.15
17 21.68  5.99  1.63 10.07 0.15 4.56 0.92 0.06
18 21.37  5.71  0.65  9.05 0.05 3.62 0.77 0.07
19 17.73  7.52  3.10  9.59 0.96 2.65 0.81 0.16
20 22.15  5.60  0.99 10.82 0.13 3.54 0.70 0.07
21 18.02  7.58  3.09 10.48 0.81 2.67 0.83 0.18
```

7.3. **Diagnostic for Group 1.** :

```
par(mfrow=c(1,1))
pairs(MnO~Fe2O3+MgO+CaO+K2O,data=roman, main="Scatterplot Matrix")
plot(roman$kiln,roman$MnO,data=roman,main="MnO-kiln plot")
#Exploring the potential predictors

par(mfrow=c(2,2))
roman_MnO_1=roman[roman$kiln=="IslandsThorns" | roman$kiln=="AshleyRails",]
roman_MnO_2=roman[roman$kiln=="Gloucester" | roman$kiln=="Caldicot",]
roman_MnO_3=roman[roman$kiln=="Llanedeyrn" ,]
#Divide the group according to our observation from boxplots.

leveneTest(roman$MnO,roman$kiln)
#Ensure the homoscedasticity using Brown-Forsythe test.

par(mfrow=c(2,2))
lm.MnO.g1<-lm(MnO~Fe2O3+MgO+CaO+K2O,data=roman_MnO_1)
plot(lm.MnO.g1);dwtest(lm.MnO.g1,alternative = "t");
par(mfrow=c(1,1));invResPlot(lm.MnO.g1);
summary(lm.MnO.g1)
##Fit and check group1 submodel
##Although the assumptions are quite good, we still do not have a strong belief that this model
     is correct OR not since the sample size is not big. So I tend to believe that there is not
     linear relationships between chemical elements in this group of kilns.
```

```
OUTPUT:
Levene's Test for Homogeneity of Variance (center = median)
          Df F value Pr(>F)
group      4  3.0035 0.0285 *
      43
---  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 0.1 ' ' 1

Durbin-Watson test
data:  lm.MnO.g1
DW = 1.2914, p-value = 0.1072
alternative hypothesis: true autocorrelation is not 0

    lambda       RSS
1   3.03276 7.788633e-06
2  -1.00000 1.753902e-05
3   0.00000 1.538670e-05
4   1.00000 1.244249e-05

Call: lm(formula = MnO ~ Fe2O3 + MgO + CaO + K2O, data = roman_MnO_1)
Residuals:
         39          40          41          42          43          44          45
-3.152e-03  -1.774e-03  -1.010e-03   2.751e-03  -3.916e-04  -9.361e-04   2.341e-03
         46          47          48
```

```
1.306e-03  8.797e-04 -1.474e-05
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.009089   0.009701   0.937    0.392
Fe2O3        -0.000113   0.001590  -0.071    0.946
MgO           0.020145   0.021659   0.930    0.395
CaO           0.032156   0.026806   1.200    0.284
K2O          -0.009824   0.006580  -1.493    0.196
Residual standard error: 0.002476 on 5 degrees of freedom
Multiple R-squared:  0.4058,    Adjusted R-squared:  -0.06949
F-statistic: 0.8538 on 4 and 5 DF,  p-value: 0.5482
```

7.4. **Diagnostic for Group 2.** :

```
par(mfrow=c(2,2))
lm.MnO.g2<-lm(MnO~Fe2O3+MgO+CaO+K2O,data=roman_MnO_2)
plot(lm.MnO.g2);summary(lm.MnO.g2)
##Fit and check group2 submodel
par(mfrow=c(2,2))
roman_MnO_2.g2=roman_MnO_2[-c(1),]
##Drop the outlier #1 from the dataset due to its abnormally high percentage of Fe2O3.
lm.MnO.g2<-lm(MnO~Fe2O3+MgO,data=roman_MnO_2.g2)
plot(lm.MnO.g2);dwtest(lm.MnO.g2,alternative = "t");
par(mfrow=c(1,1));invResPlot(lm.MnO.g2); summary(lm.MnO.g2)
##Drop CaO,K2O from this submodel since its regression coefficients are not significant.
    Whether to include Al2O3 is answered by trial and fails. ##However, this regression depends
     highly on the exclusion of the outlier #1
```

```
OUTPUT:
Call: lm(formula = MnO ~ Fe2O3 + MgO + CaO + K2O, data = roman_MnO_2)
Residuals:
Min        1Q    Median        3Q       Max
-0.037143 -0.007836 -0.002770 0.011988 0.034169
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.083429  0.053223  -1.568   0.1335
Fe2O3        0.009770  0.004716   2.072   0.0522 .
MgO          0.012474  0.017432   0.716   0.4829
CaO          0.009025  0.012944   0.697   0.4941
K2O          0.016520  0.020160   0.819   0.4227
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '. 0.1 ' 1
Residual standard error: 0.01643 on 19 degrees of freedom
Multiple R-squared: 0.4074,    Adjusted R-squared: 0.2826
F-statistic: 3.265 on 4 and 19 DF, p-value: 0.03381

Durbin-Watson test
data:  lm.MnO.g2
DW = 1.7731, p-value = 0.3701
alternative hypothesis: true autocorrelation is not 0

      lambda        RSS
1  0.4793462 0.002111616
2 -1.0000000 0.002296785
3  0.0000000 0.002132812
4  1.0000000 0.002135497
```

```
Call: lm(formula = MnO ~ Fe2O3 + MgO, data = roman_MnO_2.g2)
Residuals:
Min        1Q    Median        3Q       Max
-0.042730 -0.007811 -0.001263 0.010480 0.031714
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.101623  0.051218  -1.984 0.061133 .
Fe2O3        0.017041   0.005839   2.919 0.008494 **
MgO          0.026571   0.006721   3.954 0.000784 ***
--- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’. 0.1 ‘’1
Residual standard error: 0.01548 on 20 degrees of freedom
Multiple R-squared: 0.4456,    Adjusted R-squared: 0.3901
F-statistic: 8.036 on 2 and 20 DF, p-value: 0.002745
```

References

[Tubb et.al] Tubb, A., Parker, A. J., & Nickless, G. (1980). THE ANALYSIS OF ROMANO-BRITISH POTTERY BY ATOMIC ABSORPTION SPECTROPHOTOMETRY. Archaeometry, 22(2), 153-171.

[Mirti et.al] Mirti, P., Aruga, R., Zelano, V., Appolonia, L., & Aceto, M. (1990). Investigation of Roman terra sigillata by atomic absorption and emission spectroscopy and multivariate analysis of data. Fresenius' journal of analytical chemistry, 336(3), 215-221.

[Sanders] Sanders, H. P. (1973). PORE-SIZE DISTRIBUTION DETERMINATIONS IN NEOLITHIC, IRON AGE, ROMAN AND OTHER POTTERY. Archaeometry, 15(1), 159-161.

[Weisberg] Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons.

[Baxter&Jackson] Baxter, M. J., & Jackson, C. M. (2001). Variable selection in artefact compositional studies. Archaeometry, 43(2), 253-268.

[Sabbatini et.al] Sabbatini, L., Tarantino, M. G., Zambonin, P. G., & De Benedetto, G. E. (2000). Analytical characterization of paintings on pre-Roman pottery by means of spectroscopic techniques. Part II: Red, brown and black colored shards. Fresenius' journal of analytical chemistry, 366(1), 116-124.

E-mail address: mailto:luo.619@osu.edu