

MSDS 6372 : Principal Component Analysis

Glass Identification : Refractive Index (RI) Analysis

Jecha, Jean jjecha@mail.smu.edu ; Lisbona, Randall rlisbona@mail.smu.edu ;
Nguyen, Heidi hqnguyen@mail.smu.edu Southern Methodist University

1. Introduction

"Now you see it, now you don't. Glass is a bit of a riddle. It's hard enough to protect us, but it shatters with incredible ease. It's made from opaque sand, yet it's completely transparent. And, perhaps most surprisingly of all, it behaves like a solid material... but it's also a sort of weird liquid in disguise! You can find glass wherever you look: most rooms in your home will have a glass window and, if not that, perhaps a glass mirror... or a glass lightbulb. Glass is one of the world's oldest and most versatile human-created materials." (Woodford 2016)

Soda Lime glass is the most common type of glass used for windows and glass containers. Its primary components are Silica (Quartz sand), Sodium Oxide (Soda Ash), and Calcium Oxide (Lime). (Soda-Lime Glass n.d.) Soda Lime glass dates as far back as 2000 B.C. It is made in batch and continuous casting processes with refining temperatures as high as 3090°F. (IMI-NFG course on Processing Glass 2015) When the mixture cools, it completely transforms into a different structure that is not quite solid. It becomes a material that scientists refer to as an **amorphous solid**, a cross between a solid and a liquid. Usually other minerals are added into the liquid glass to change its working temperature, hardness, chemical properties, or color based on the purpose of the glass.

Refractive index (RI) determines how much light is bent or refracted when entering or exiting a transparent material or in our case glass. The purpose of this observational study is to understand the different chemical compositions of glass and determine if refractive index can be predicted based on composition. This study looks at seven chemical compositions of Soda Lime Silica Glass using a dataset from The University of California, Irvine which contains 214 observations of 7 different glass types composed of 8 minerals by % weight. (UCI Machine Learning Repository n.d.)

2. Exploratory Analysis

The dataset (**Ref: Table A**) includes compositions for Building and Vehicle windows (float and non-float processed) containers, tableware, and headlamps. (Float Glass process n.d.) SAS proc Means was used to summarize the data. Note that Silica, Soda, and Lime make up 95% of the composition by weight

Table A - Variable List

Variable	Label	N	Mean	Cum. %	Std Dev	Minimum	Maximum
RI	RI - Refractive Index	214	1.52		0	1.51	1.53
SI02	SI02 - Silicon Oxide (Silica Sand)	214	72.65	72.65	0.77	69.81	75.41
NA2O	NA2O - Sodium Oxide (Soda Ash)	214	13.41	86.06	0.82	10.73	17.38
CAO	CAO - Calcium Oxide (Limestone)	214	8.96	95.02	1.42	5.43	16.19
MGO	MGO - Magnesium Oxide	214	2.68	97.70	1.44	0	4.49
AL2O3	AL2O3 - Aluminum Oxide	214	1.44	99.14	0.5	0.29	3.5
K2O	K2O - Potassium Oxide (Potash)	214	0.5	99.64	0.65	0	6.21
BAO	BAO - Barium Oxide	214	0.18	99.82	0.5	0	3.15
FE2O3	FE2O3 - Iron Oxide	214	0.06	99.88	0.1	0	0.51

The box and whisker plot (Ref: Figure 1) shows the variability of each mineral in the UCI glass dataset as a percent of weight.

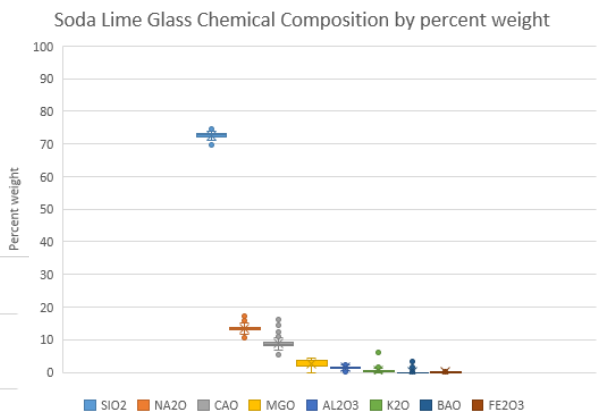


Figure 1 – Chemical Composition

The box and Whisker plot (Ref: Figure2) shows a relationship between the type of glass and the refractive index (RI). Further analysis will help determine if the refractive index can be predicted based on glass composition.

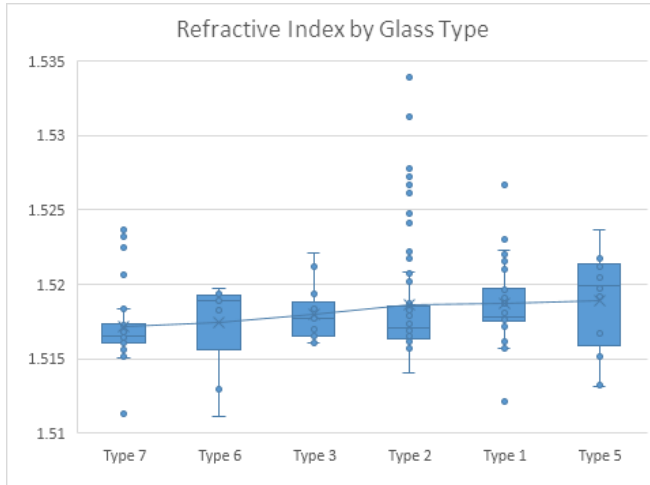


Figure 2 – RI by Glass Type

Scatter plots (Ref: Figure 3) show a strong negative relationship for Silica and a strong positive relationship for Calcium Oxide on Refractive index.

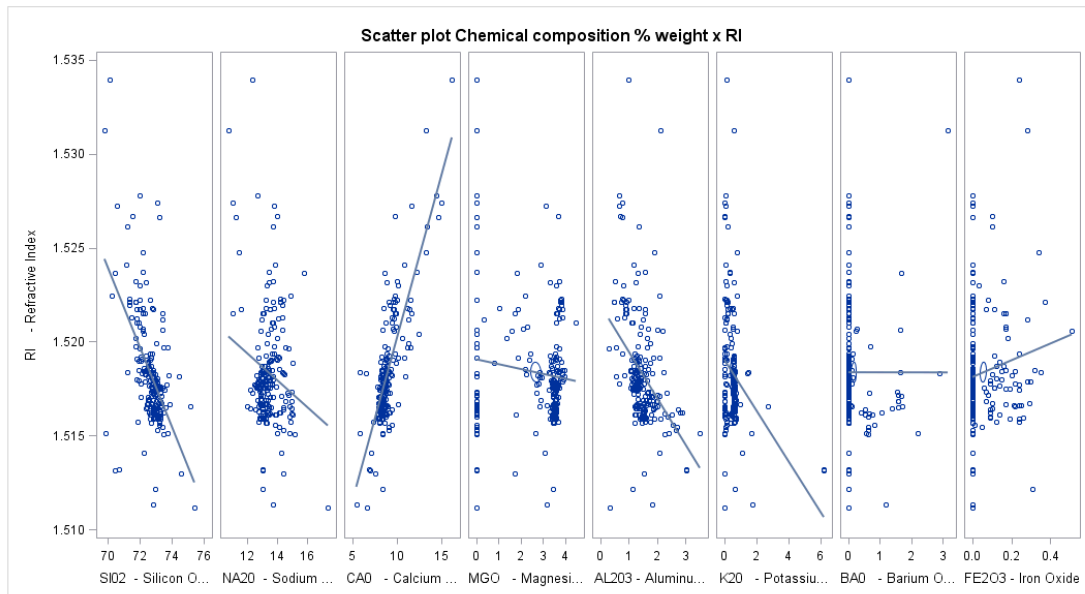


Figure 3 – Scatter Plot Chemical Composition % Weight by RI

3. Validation of assumptions

For purposes of this study, it is assumed that the refractive index of soda lime glass can be predicted using statistical tools such as principal components and regression. Although there are only 8 minerals in the dataset, the objective of this project is to use principal components to reduce the number of independent variables.

4. Principal Components Analysis (PCA)

Correlations:

In our study, we are using the correlation matrix over the covariance matrix because of the variance in the variables of these data. The correlation matrix is sums of squares and cross products from the standardized data. This correlation will tell use which elements have the highest positive and highest negative correlation.

In the correlation matrix (**Ref: Table B**), we see a high positive correlation between variables **Sodium Oxide**(NA20) and **Barium Oxide**(BA0) with **0.33**, **Aluminum Oxide**(AL203) and **Potassium Oxide**(K20) with **0.33** and **Aluminum Oxide** (AL203) and **Barium Oxide**(BA0) with **0.48**. We also see a low negative correlation between variables **Magnesium Oxide**(MG0) and **Aluminum Oxide**(AL203) with **-0.48**, **Magnesium Oxide**(MG0) and **Barium Oxide**(BA0) with **-0.49** and **Magnesium Oxide**(MG0) and **Calcium Oxide**(CA0) with **-0.44**.

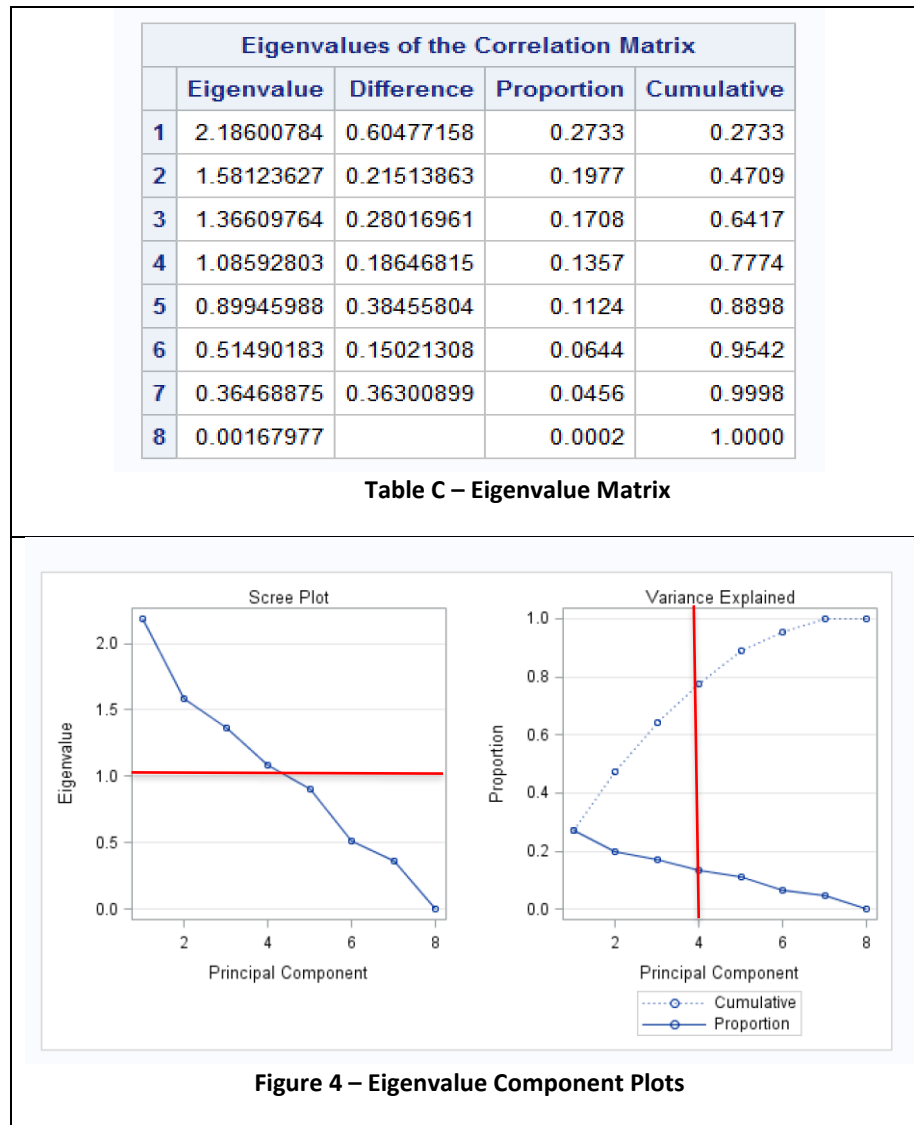
Correlation Matrix								
	NA20	MG0	AL203	SI02	K20	CA0	BA0	FE203
NA20	1.0000	-0.2737	0.1568	-0.0698	-0.2661	-0.2754	0.3266	-0.2413
MG0	-0.2737	1.0000	-0.4818	-0.1659	0.0054	-0.4438	-0.4923	0.0831
AL203	0.1568	-0.4818	1.0000	-0.0055	0.3260	-0.2596	0.4794	-0.0744
SI02	-0.0698	-0.1659	-0.0055	1.0000	-0.1933	-0.2087	-0.1022	-0.0942
K20	-0.2661	0.0054	0.3260	-0.1933	1.0000	-0.3178	-0.0426	-0.0077
CA0	-0.2754	-0.4438	-0.2596	-0.2087	-0.3178	1.0000	-0.1128	0.1250
BA0	0.3266	-0.4923	0.4794	-0.1022	-0.0426	-0.1128	1.0000	-0.0587
FE203	-0.2413	0.0831	-0.0744	-0.0942	-0.0077	0.1250	-0.0587	1.0000

Table B – Correlation Matrix Standardized Data

Eigenvalues:

Looking at the correlation matrix (**Ref: Table C**) the first four (4) principal components variables have Eigenvalues greater than one (1) explaining 77% of the variance. The largest difference is between component one (1) and component two (2) with 0.60.

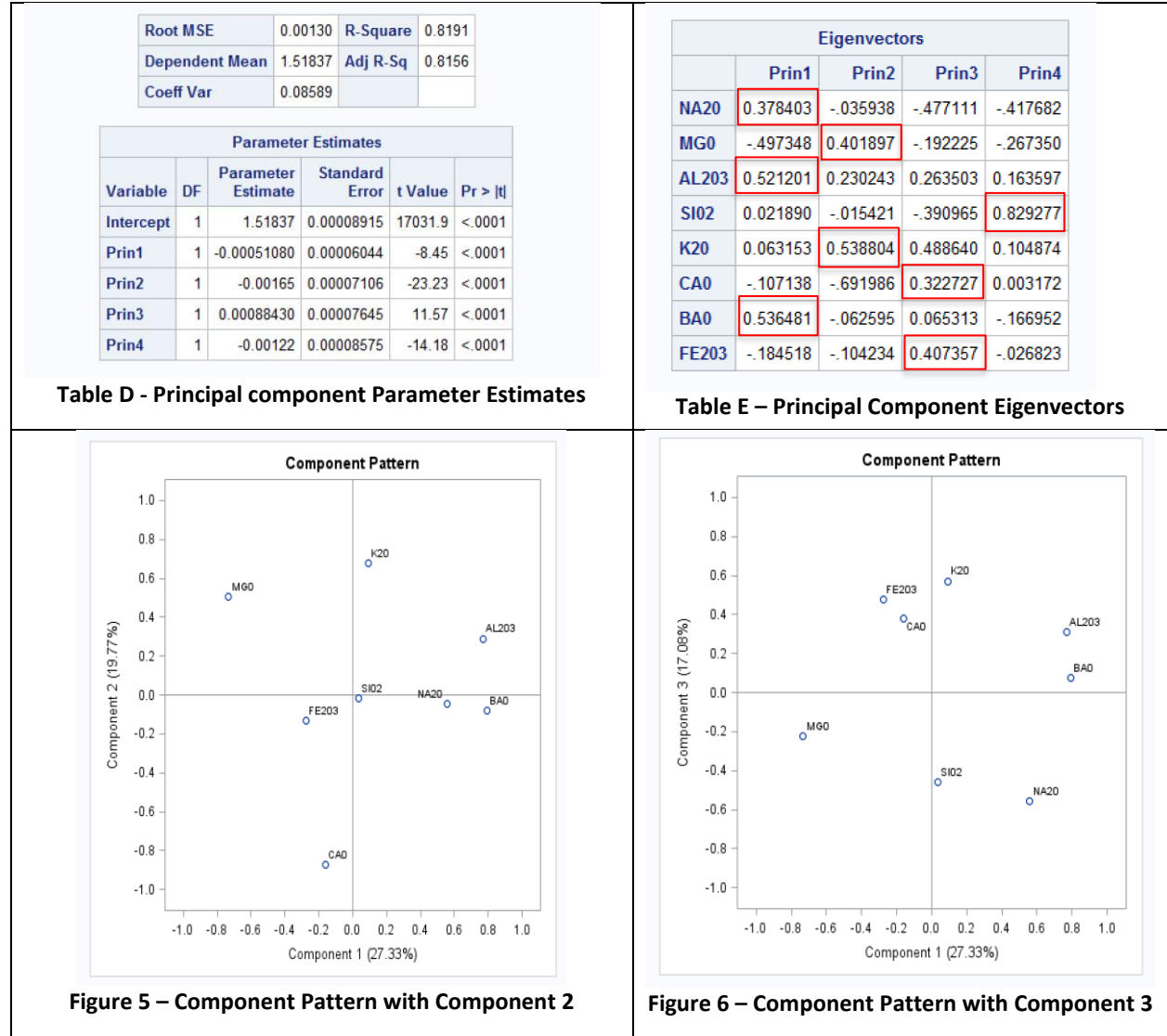
Our Scree Plot (**Ref: Figure 4**) for these data doesn't show a distinct steep curve, that bends to a flat horizontal line as we might expect to see. Instead this Scree Plot seem to descend at an angle almost linear pattern but still shows our four primary components with Eigenvalues above one. Next to the Scree Plot in **Figure 4** is the Variance Explained plot which is a graphical view of the Eigenvalue Matrix in **Table C**. Between the Eigenvalue Matrix and the Variance Explained graph helps us decide the number of components.



Components:

From the Eigenvalues analysis, above we will focus on the first four (4) principal components more closely. Running our model with the four principal components (**Ref: Table D**) we show that all principal components are statistically significant ($p\text{-values} < 0.0001$). Component patterns shown in **Figure 5** is a graphical view of the first component with the second component from the Eigenvectors in table E. From this we see that both Magnesium Oxide(MG0) and Potassium Oxide(K20) correlates highly with the second

component. Component patterns shown in **Figure 6** show a graphical view of the first component with the third component. Here we see a high correlation with the third component for Iron Oxide (FE203) and Potassium Oxide (K2O). The Eigenvector **Table E** shows final variables for each principal component with the highlighted boxes that is used in the regression equation below.



5. Conclusion

Statistical Conclusion:

The regression equation based on the response RI and statistical significance of the first four principal components:

$$RI = 1.5184 - 5.1080 \times 10^{-4}W_1 + 1.6500 \times 10^{-3}W_2 + 8.8430 \times 10^{-4}W_3 - 1.2200 \times 10^{-3}W_4$$

Based on Eigenvalues matrix (**Ref: Table E**), we can see that the listed original variables mostly contribute based on the direction of their maximum variance, to the principal components prin1, prin2, prin3, and prin4:

Prin1: Sodium Oxide (NA2O), Aluminum Oxide (AL2O3) and Barium Oxide (BAO)

Prin2: Magnesium Oxide (MGO) and Potassium Oxide (K2O)

Prin3: Calcium Oxide (CAO) and Iron Oxide (FE2O3)

Prin4: Silicon Oxide (SIO2)

The regression equation of the response RI with the original predictor variables:

$$RI = 1.5184 - 5.1080 * 10^{-4}(0.438403 * NA2O + 0.521201 * AL2O3 + 0.53648 * BAO) + 1.6500 * 10^{-3}(0.401897 * MGO + 0.538804 * K2O) + 8.8430 * 10^{-4}(0.322727 * CAO + 0.407357 * FE2O3) - 1.2200 * 10^{-3}(0.829277 * SIO2)$$

Scope:

As we know, the statistical association from these observational data cannot be used to establish a causal interpretation. However, based on the parameter estimates, we do see that there is a very strong correlation between given attributes.

APPENDIX

SAS CODE

```
data glass;
infile 'glass.csv' dlm=',' firstobs=2 dsd;
input ID RI NA20 MG0 AL203 SI02 K20 CA0 BA0 FE203 Type ;
run;

proc print data=glass ; run;

/* Initial princomp with standardized data and correlation matrix */
proc princomp data=glass;
var NA20 MG0 AL203 SI02 K20 CA0 BA0 FE203 ;
run;

/* Trying the proc factor for principal components analysis validate our
number of components*/
proc factor data=glass method=prin scree ;
var NA20 MG0 AL203 SI02 K20 CA0 BA0 FE203 ;
run;

/* Rerunning the princomp with only the first 4 principal components id
RI;*/
proc princomp data=glass out=glassPC plots(ncomp=3)=all n=4;
var NA20 MG0 AL203 SI02 K20 CA0 BA0 FE203 ;

run;
/* final selection for the principal components using proc reg - all have
a p-value < 0.0001
the adjusted R-square is 0.82*/
proc reg data = glassPC;
model RI = prin1 prin2 prin3 prin4;
run;
```

References

Bibliography

- n.d. *Float Glass process*. <http://www.pilkington.com/pilkington-information/about+pilkington/education/float+process/step+by+step.htm>.
- n.d. *History of Glass*. https://en.wikipedia.org/wiki/History_of_glass.
2015. *IMI-NFG course on Processing Glass*.
http://www.lehigh.edu/imi/teched/GlassProcess/Lectures/Lecture02_Hubert_Melting%20and%20fining%20processes.pdf.
- n.d. "SAS princomp." *OKState.edu*. <https://www.okstate.edu/sas/v8/saspdf/stat/chap52.pdf>.
- n.d. *Soda-Lime Glass*. https://en.wikipedia.org/wiki/Soda-lime_glass.
- n.d. *UCI Machine Learning Repository*. <http://archive.ics.uci.edu/ml/datasets/Glass+Identification>.
- Woodford, Chris. 2016. September 17. <http://www.explainthatstuff.com/glass.html>.