Pea Quality Analysis: A Principal Component Approach

Olatunji Akinbule

Master of Science, Data Analytics

School of Engineering and Applied Sciences

The George Washington University

Table of Contents

Introduction

The client has presented a dataset consisting of 17 pea attributes of a 60-pea variation.

The objectives are to:

- Analyze the data using Principal Components Analysis (PCA)
    - Decide on the number of Principal components to retain with reasons
    - Interpret the components
- Design a Pea metric
- Fit an appropriate theoretical distribution to score the results
- Select the top 10 percent of Peas that outperform the others

Principal Component Analysis (PCA)

When dealing with dataset with a substantial number of features - highly dimensional data- it can be difficult to comprehend patterns of association. PCA is a method of re-expressing multivariate data. It allows a better understanding by reorienting the data so that the first few dimensions account for as much of the available information as possible.

To interpret the data in a more meaningful form, it is necessary to reduce the number of variables to a few, interpretable linear combinations of the data with each linear combination corresponding to a principal component.

PCA constructs orthogonal – mutually uncorrelated – linear combinations that explains as much common variation as possible in a given dataset. In other words, PCA seeks the direction that maximizes the variance providing independent pieces of the information puzzle represented in the larger dataset.

Preprocessing

The original dataset contains information for 60 pea variations consisting of 17 features with varying measurement metrics in the range of hundreds, tens and others in units.

Standardization of the data is critical to performing Principal Component Analysis (PCA) to avoid the impact of measurement units across the different features in the analysis.

That is the research would like each feature to receive equal weight in the analysis. Standardizing the pea features entails subtracting the mean and dividing by the standard deviation

$$Z_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j}$$

where

- X_{ij} = Data for variable j in sample unit i
- $\bar{x}_j$= Sample mean for variable j
- s_j = Sample standard deviation for variable j

Recaps:

Principal Components Analysis (PCA) is a data analysis tool used to reduce the dimensionality (number of variables) of many interrelated variables, while retaining as much of the information (variation) as possible.

PCA calculates an uncorrelated set of variables called factors or principal components.

Increasingly ordered Principal Components retain most of the variation present in all the original variables.

Table 1 – Original Data

	Pea ID	Tenderometer	Dry matter	Dry matter after freezing	SucrosePercent	TotalGlucose1	TotalGlucose2	Flavour	Sweet	Fruity	Off-flavour	Mealiness	Hardness	Whiteness	Colour1	Colour2	Colour3	Skin
2	**1**	110.00	15.10	19.09	5.40	3.30	3.00	6.48	6.66	4.56	2.20	2.91	3.47	4.72	5.59	5.73	5.99	4.26
3	**2**	120.00	16.80	20.52	5.00	4.00	3.80	5.75	6.09	3.81	2.32	4.03	3.77	4.17	5.73	5.75	5.32	3.82
4	**3**	150.00	20.10	22.77	3.90	4.00	3.70	3.94	4.12	2.44	3.63	5.77	5.39	4.77	6.66	5.11	4.60	3.50
5	**4**	109.00	17.50	20.79	4.90	3.50	3.30	6.60	6.12	4.44	1.93	3.31	4.46	4.86	5.16	5.74	6.57	2.12

Table 2 – Standardized Data

Pea ID	Tenderometer	Dry matter	Dry matter after freezing	Sucrose Percent	TotalGlucose1	TotalGlucose2	Flavour	Sweet	Fruity	Off-flavour	Mealiness	Hardness	Whiteness	Colour1	Colour2	Colour3	Skin
1	-0.80	-1.41	-1.25	1.00	-0.75	-1.00	0.98	1.04	0.97	-0.66	-1.09	-0.89	0.31	0.29	-0.11	0.66	3.15
2	-0.48	-0.86	-0.70	0.64	-0.18	-0.31	0.36	0.56	0.25	-0.55	-0.27	-0.68	-0.66	0.49	-0.09	-0.01	2.40
3	0.47	0.19	0.18	-0.33	-0.18	-0.39	-1.18	-1.09	-1.06	0.71	1.02	0.44	0.39	1.77	-1.29	-0.76	1.86
4	-0.83	-0.64	-0.59	0.56	-0.59	-0.74	1.07	0.59	0.86	-0.92	-0.80	-0.20	0.56	-0.28	-0.10	1.25	-0.50
5	-0.64	-0.83	-0.56	0.20	-0.67	-0.57	0.30	0.47	0.24	-0.74	-0.40	-0.43	0.87	0.36	-1.08	0.15	-0.04

Table 3 – Correlation Matrix

Correlation Matrix of PEA Attributes among 17 descriptive features

Threshold = 0.8

	Tenderomet	Dry matter	Dry matter after freezing	Sucrose Percent	TotalGlucose	TotalGlucose	Flavour	Sweet	Fruity	Off-flavour	Mealiness	Hardness	Whiteness	Colour1	Colour2	Colour3	Skin
Tenderom	1																
Dry matte	0.941367	1															
Dry matter after freezing	0.949758	0.980364	1														
Sucrose Percent	-0.87261	-0.85456	-0.861	1													
TotalGluc	0.862229	0.873711	0.879363	-0.84793	1												
TotalGluc	0.882789	0.881094	0.896609	-0.85542	0.962906	1											
Flavour	-0.89819	-0.88626	-0.88442	0.920271	-0.85296	-0.84202	1										
Sweet	-0.90772	-0.89088	-0.89145	0.959344	-0.8351	-0.83931	0.951248	1									
Fruity	-0.9004	-0.89736	-0.89875	0.927273	-0.85528	-0.84699	0.975098	0.949181	1								
Off-flavou	0.819603	0.824766	0.818103	-0.87725	0.772931	0.750832	-0.95234	-0.90145	-0.90339	1							
Mealiness	0.877943	0.906163	0.898913	-0.8877	0.84951	0.836272	-0.93298	-0.91381	-0.96886	0.836218	1						
Hardness	0.946544	0.928529	0.934129	-0.92269	0.871619	0.87516	-0.9242	-0.94519	-0.94385	0.854454	0.926803	1					
Whitenes	-0.01332	-0.005	-0.01788	-0.21531	0.098643	0.101226	-0.12684	-0.15247	-0.08846	0.107395	0.084279	0.029945	1				
Colour1	0.016509	0.004649	-0.01517	-0.20717	0.060547	0.063248	-0.21852	-0.17179	-0.14936	0.226689	0.113584	0.0182	0.837965	1			
Colour2	-0.13301	-0.12019	-0.10774	0.372277	-0.20641	-0.1969	0.342898	0.3308	0.277767	-0.34777	-0.2338	-0.16935	-0.89299	-0.90784	1		
Colour3	-0.26844	-0.22459	-0.22061	0.418688	-0.24392	-0.23677	0.448025	0.39291	0.398	-0.50663	-0.32695	-0.30279	-0.41408	-0.63462	0.64673	1	
Skin	-0.63919	-0.70437	-0.69571	0.539527	-0.53516	-0.57368	0.478384	0.556216	0.522858	-0.41824	-0.54513	-0.62797	0.109676	0.226299	-0.12406	-0.01645	1

Table 4 –

PEA	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15	Z16	Z17
Eigenvalues	11.42	3.247	0.756	0.509	0.301	0.183	0.149	0.122	0.083	0.055	0.049	0.033	0.028	0.027	0.018	0.013	0.008
Percentage	67.20%	19.10%	4.40%	3.00%	1.80%	1.10%	0.90%	0.70%	0.50%	0.30%	0.30%	0.20%	0.20%	0.20%	0.10%	0.10%	0.00%
Cumulative	67.20%	86.30%	90.70%	93.70%	95.50%	96.60%	97.40%	98.20%	98.60%	99.00%	99.30%	99.40%	99.60%	99.80%	99.90%	100.00%	100.00%

From Table 4

The second column (Percentage) shows the proportion of variation explained by each eigenvalue and the third column (cumulative) percentage adds the successive proportions showing the total variation of the data each principal component.

The first and second eigenvalues explain 67.20% and 19.10% of the variation of Pea data respectively. The first two eigenvalues explain 86.30% of the variation in the Pea dataset.

The results so far support our ideal objective in reducing the dimensionality of our data while minimizing the loss of information in a bid to understand the underlying patterns.

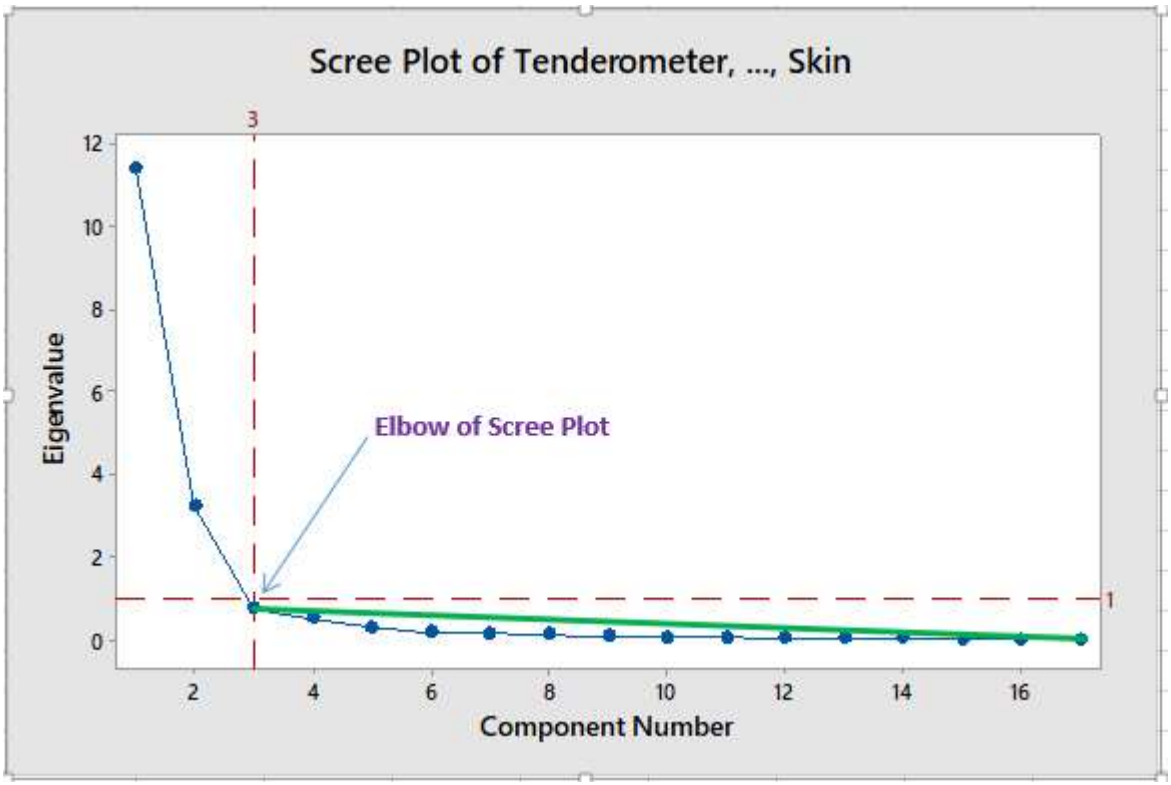


Figure 1: Scree plot

The elbow of the Scree Plot helps in determining the number of Principal Components to retain in our analysis. By inspecting the differences between eigenvalues, we see that the first inflection point occurs between the third and fourth eigenvalues. Furthermore, there is an almost linear behavior of the eigenvalues after the third Principal Component highlighted with the dotted reference lines and shown with the green line in Figure 1.

Jackson reports that the most common stopping rule in PCA depends on the average value of the eigenvalues positing that the sum of the eigenvalues equals the number of variables. The Kaiser-Guttman criterion selects eigenvalues greater than the average eigenvalue (i.e, X > 1) because those axes summarize more information than any single ordinary variable (Jackson, 1993).

In essence, Kaiser's rule stipulates that principal components that extract at least as much as the equivalent of one original variable should be selected. Therefore, this report retains two eigenvalues greater than 1 (PC1 & PC2) with eigenvalues at 11.42 and 3.25 respectively.

Interpretation of Principal Components

The interpretation of the principal components shows variables strongly correlated with each component on their respective axis based on their position in the PC1 – PC2 co-ordinates.

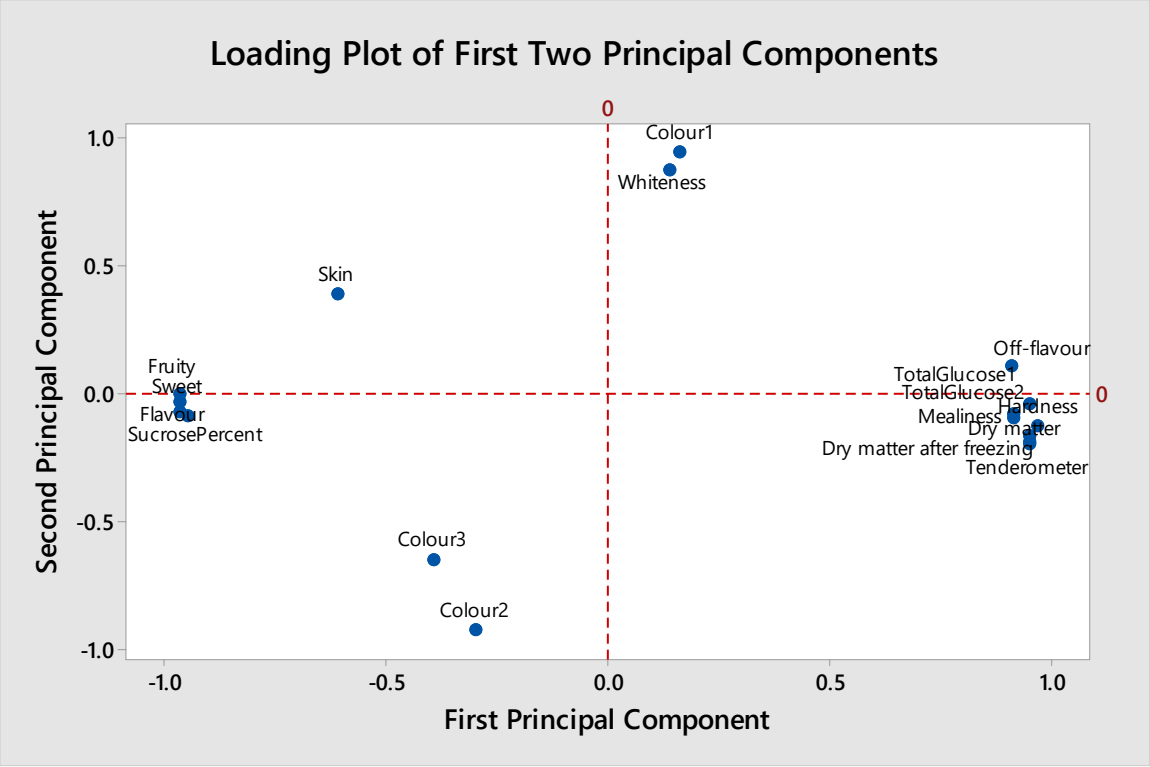


Figure 2: Loading Plot for PC1 and PC2

From the Loading Plot in Figure 2 above the first component is expressing information on the texture, content and taste qualities. Generally, there is a separation between the taste on the extreme negative end of the first component (*Fruity, Sweet, Flavor*) and the texture-content

features on the right positive end (*Tenderometer, Mealiness, Dry matter after Freezing, Dry matter, Sucrose Present, TotalGlucose1, TotalGlucose2*)

Therefore, the First PC, generally, is the index for taste/ texture-content quality while the second principal component appears to be an appearance classification component because it polarizes the features with a positive distinction on Skin, Colour1 and Whiteness and a negative distinction with Colour2 and Colour3

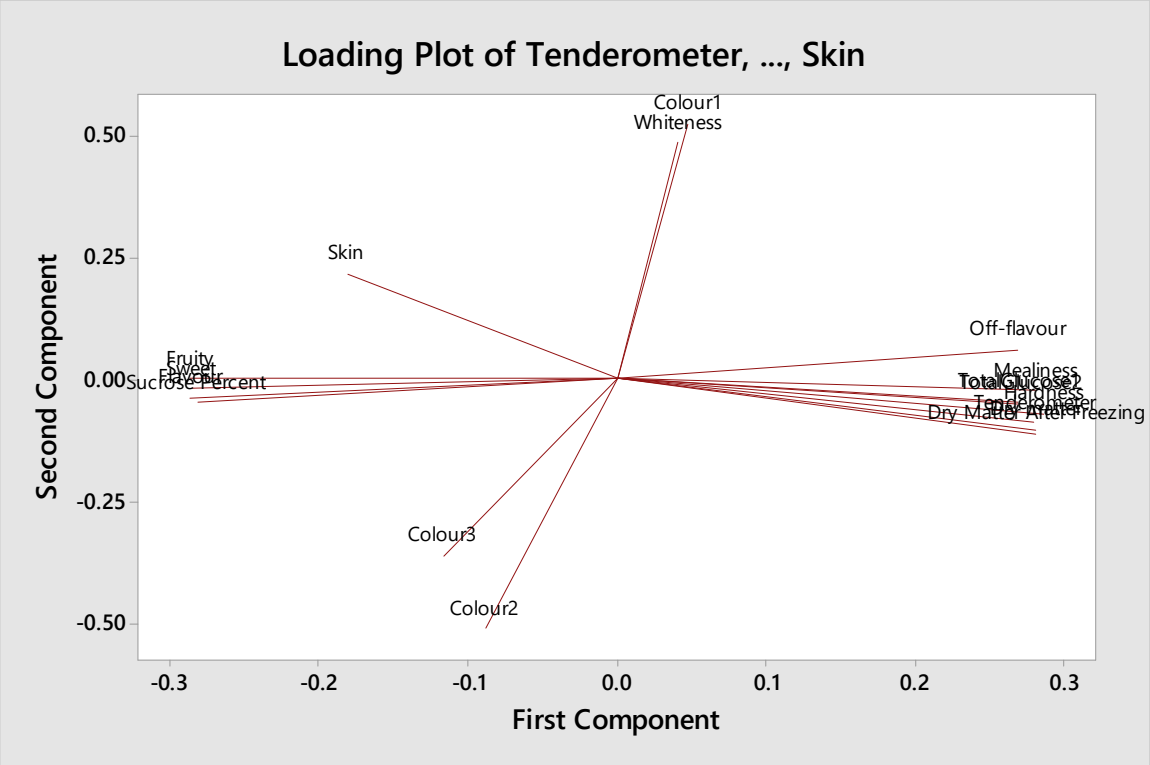


Figure 3: Loading Plot showing Pea features

The Loading plot presents an additional pictorial representation of our feature positioning based on the First and Second principal components.

The analysis of the correlations between the original variables and Principal components reinforce the information from the loading plot.

Correlations between Original Variables Xi and Principal Components Zi					
Loading Analysis on PC1 & PC2					
(Threshold = 0.8)					
Loadings	Z1	Z2	Z3	Z4	Z5
Tenderometer	0.947663	-0.16398	-0.01699	-0.01304	-0.02798
Dry matter	0.949596	-0.18852	-0.0933	-0.04523	-0.00218
Dry matter after freezing	0.950516	-0.20236	-0.08605	-0.02773	-0.02933
SucrosePercent	-0.95325	-0.08856	-0.00563	-0.02374	-0.0683
TotalGlucose1	0.911298	-0.0841	-0.10804	0.212063	-0.2794
TotalGlucose2	0.913489	-0.0947	-0.15213	0.170167	-0.29082
Flavour	-0.96871	-0.06989	-0.15979	-0.05626	-0.08581
Sweet	-0.96762	-0.03701	-0.04607	-0.00738	-0.14327
Fruity	-0.97026	-0.00046	-0.12371	-0.04972	-0.10558
Off-flavour	0.908591	0.106703	0.269132	-0.01194	0.103777
Mealiness	0.948127	-0.04387	0.046469	0.069042	0.111856
Hardness	0.967696	-0.12812	0.01523	0.00067	0.05138
Whiteness	0.135856	0.87093	-0.42533	0.07318	0.04559
Colour1	0.160241	0.943053	-0.07125	-0.01616	0.011822
Colour2	-0.30091	-0.92503	0.128478	-0.02106	-0.04615
Colour3	-0.39553	-0.65598	-0.4372	0.3846	0.249422
Skin	-0.61235	0.387181	0.439701	0.516668	-0.02295

Table 5 – Loading Analysis on Principal Components 1 & 2

First Principal Component Analysis – PCA 1

Based on the selected Threshold of 0.8, the first principal component Z1 is strongly correlated with 12 of the variables highlighted in yellow increasingly with *Tenderometer*, *Dry matter*, *Dry matter after freezing*, *TotalGlucose1*, *TotalGlucose2*, *Off-flavor*, *Mealiness*, *Hardness* – and decreasingly with *SucrosePercent*, *Flavor*, *Sweet*, *Fruity* and *Skin*.

Second Principal Component Analysis – PCA 2

Component Z2 shows a positive correlation with *Whiteness* and *Colour1* and a negative correlation with *Colour2* and *Colour3*.

It is also noteworthy to point out the mutually uncorrelated nature of the linear combinations PC1 and PC2. They represent and explain distinct linear combinations that explain as much common variation as possible in the dataset.

PEA Quality Metrics

The individual signs of each variable determines the corrleation between the variable and its respective Principal Component.

Figure 4: Variation of feature changes with First component

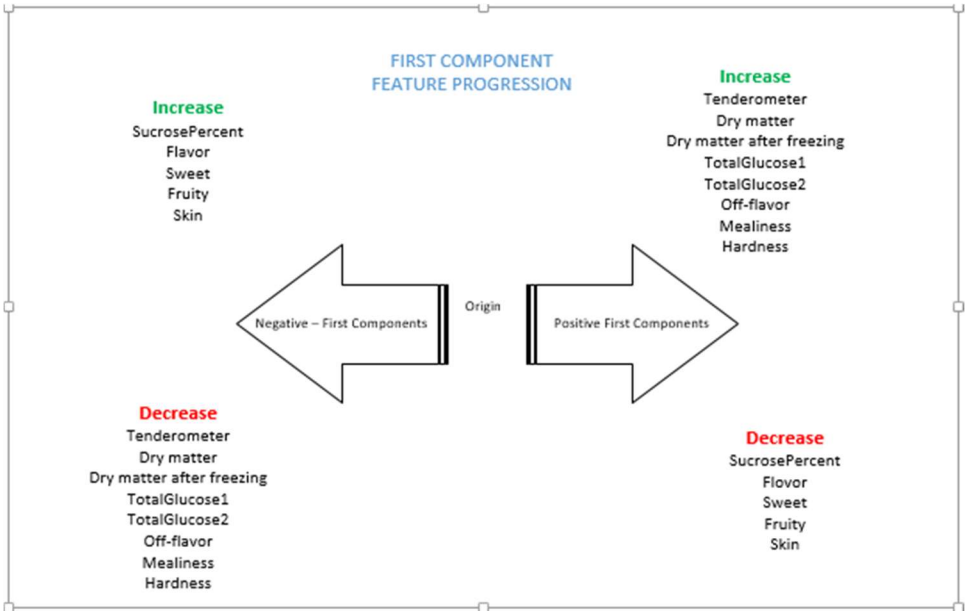
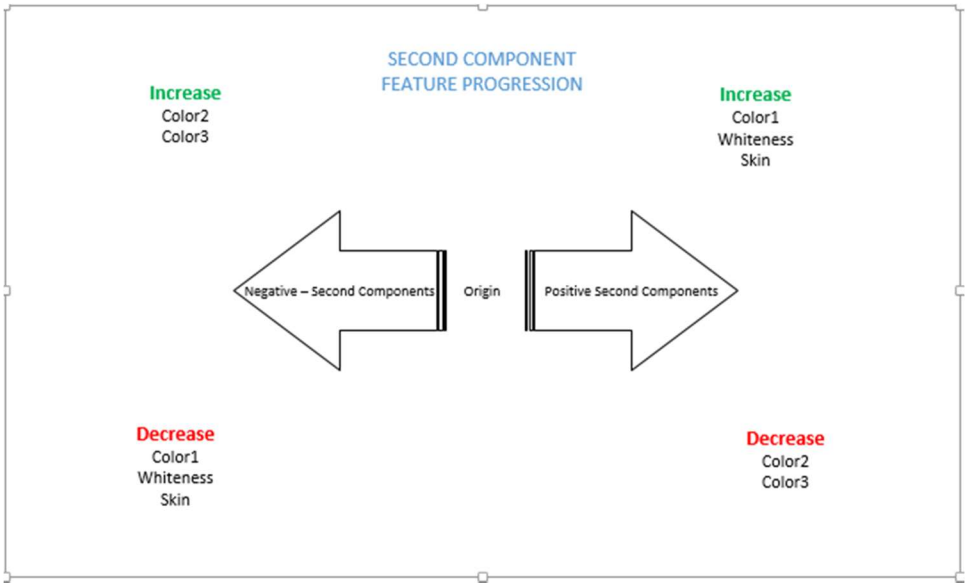


Figure 5: Variation of feature changes with Second component



Metrics Construction

My judgment of the best peas will be ones high values in the following properties: maturity and quality (Tenderometer), dry matter before and after freezing, TotalGlucose1, Total Glucose 2, Off-flavor, Mealiness and Hardness but low on sweetness, flavour, sucrose, fruity taste – PC1 is an apt representation of this expectation.

Additionally, Brianna Elliot posits the best peas are green peas since they possess a fair amount of antioxidants and fiber (not white as PC2 is positively correlated with). A key assumption regarding the color variables is that either Colour2 or Color 3 represents a green color since high values of PC 2 represent white and colour1.

The aforementioned stipulated metrics translate to peas with high PC1 values and low PC2.

Metric Formulation

$$\textbf{Metric} = \begin{bmatrix} x1,1 & \dots & x1,17 \\ . & . & . \\ . & . & . \\ . & . & . \\ x60,1 & \dots & x60,17 \end{bmatrix} \begin{bmatrix} pc1,1 & pc2,1 \\ . & . \\ . & . \\ . & . \\ pc1,17 & pc2,17 \end{bmatrix} = \begin{bmatrix} met1,1 & met1,2 \\ . & . \\ . & . \\ . & . \\ met60,1 & pc60,2 \end{bmatrix}$$

Transforming the entire data set 60 X 17 with the first and second principal components (17 X 2) to a PC weighted metric (60 X 2) with assigned evaluations based on the distinct attributes of both principal components.

The total scoring metric for the peas will be a subtraction of each PC2 score component (**met X, 2**) from its respective PC 1 Score component (**met X, 1**).

$$\textbf{Total PC Score} = \text{met X, 1 - met X, 2}$$

This ensures that values with high met X, 1 and low met X,2 values (better if negative) rank the highest peas.

The results of the overall ranking are presented in Table 6.

Pea ID	PC1Score	PC2Score	TotalScore
1	-13.28344677	2.139587374	-15.42303414
2	-6.99869947	1.275799706	-8.274499176
3	5.717802613	4.453332419	1.264470194
4	-8.056979498	-0.390724654	-7.666254844
5	-5.221990741	2.405812775	-7.627803515
6	5.85930524	1.862055212	3.997250028
7	-12.99064376	2.985673156	-15.97631692
8	-5.842012829	1.836868564	-7.678881393
9	13.87869011	2.893726186	10.98496393
10	-6.381908341	0.806087158	-7.187995499
11	-5.973956504	-1.760345168	-4.213611336
12	21.38105513	-2.519839607	23.90089474
13	0.224047667	1.478213002	-1.254165334
14	0.860982197	0.491851633	0.369130564
15	17.69317528	-0.757387509	18.45056279
16	-13.13401396	-3.954159189	-9.179854772
17	-3.331492054	-1.898296833	-1.43319522
18	12.71900494	-1.016916206	13.73592115
19	-10.53028954	8.277602158	-18.8078917
20	2.9850676	5.611290684	-2.626223084
21	14.27674158	2.402740442	11.87400114
22	-11.52776878	3.539427887	-15.06719666
23	-2.860879547	-6.353162899	3.492283352
24	19.63724689	0.586299644	19.05094725
25	-10.38163027	-2.413998255	-7.967632019
26	-8.095775243	-4.254984499	-3.840790745
27	24.89955266	-1.617396928	26.51694959
28	-1.346145056	1.930031066	-3.276176122
29	14.54641718	0.269462169	14.27695501
30	1.950553286	-2.145141756	4.095695042

Pea ID	PC1Score	PC2Score	TotalScore
31	-12.39133457	-3.654760966	-8.736573606
32	0.311241231	-5.254716761	5.565957992
33	6.429529505	-1.964647496	8.394177002
34	-14.51534727	3.041854239	-17.55720151
35	-4.774211547	-0.194192604	-4.580018943
36	-3.34272542	-4.691317667	1.348592247
37	-11.22266959	2.848890528	-14.07156012
38	-2.519174505	3.153769024	-5.672943529
39	7.413548001	4.351772724	3.061775277
40	-7.60920751	0.838753998	-8.447961508
41	0.274183236	-3.563279564	3.837462799
42	19.0997225	-0.302390918	19.40211341
43	-12.8430967	6.137044016	-18.98014072
44	0.640231914	-4.336736995	4.976968908
45	12.08558175	-1.005240555	13.0908223
46	-3.604339938	-3.071807198	-0.53253274
47	-14.16639503	0.548607903	-14.71500293
48	11.05721478	-0.348961984	11.40617676
49	-8.616572951	3.797134293	-12.41370724
50	1.086629992	-0.729788446	1.816418438
51	30.25324913	6.067914363	24.18533476
52	-16.28820176	-1.26112601	-15.02707575
53	-11.02227399	-2.986454972	-8.035819014
54	4.882160193	-3.278613852	8.160774045
55	-11.15328764	-1.489839205	-9.663448439
56	-2.081556447	-4.538732102	2.457175655
57	13.14381229	-6.299610613	19.4434229
58	-14.27552228	-1.590117644	-12.68540463
59	-2.717545315	1.755744966	-4.473290281
60	15.79434795	1.857341767	13.93700618

Table 6 – Metric Scoring for all Peas

Score Statistics

	PC1Score	PC2Score
Minimum	-16.29	-6.35
Maximum	30.25	8.28

Fitting a Distribution to Total Scores

The distribution of the total scores is a right skewed histogram with a fitted normal distribution having a P-value of 0.104. Therefore, there is not enough evidence to reject the normality of the Pea scores at an alpha level of 0.05.

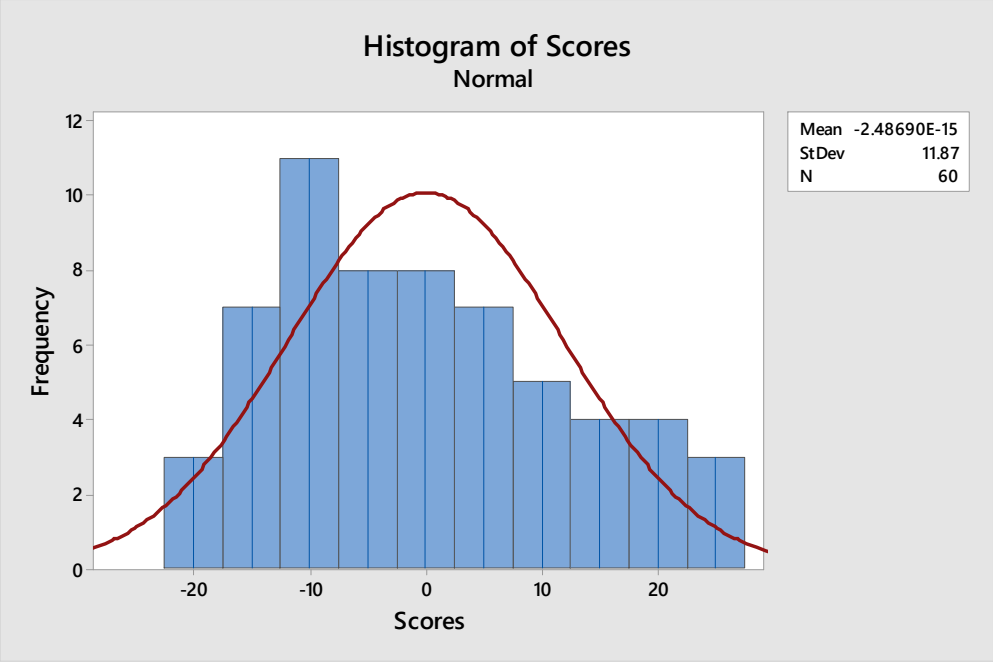


Fig 6 – Histogram of Total Scores

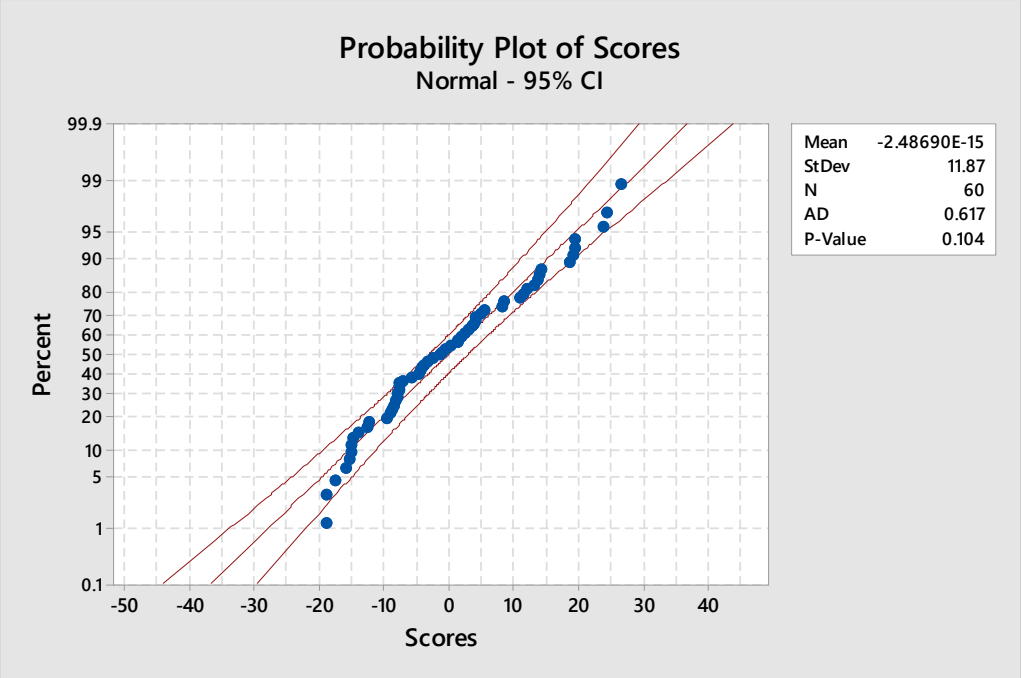


Fig 7 – Probability Plot of Total Scores

Pea Selection

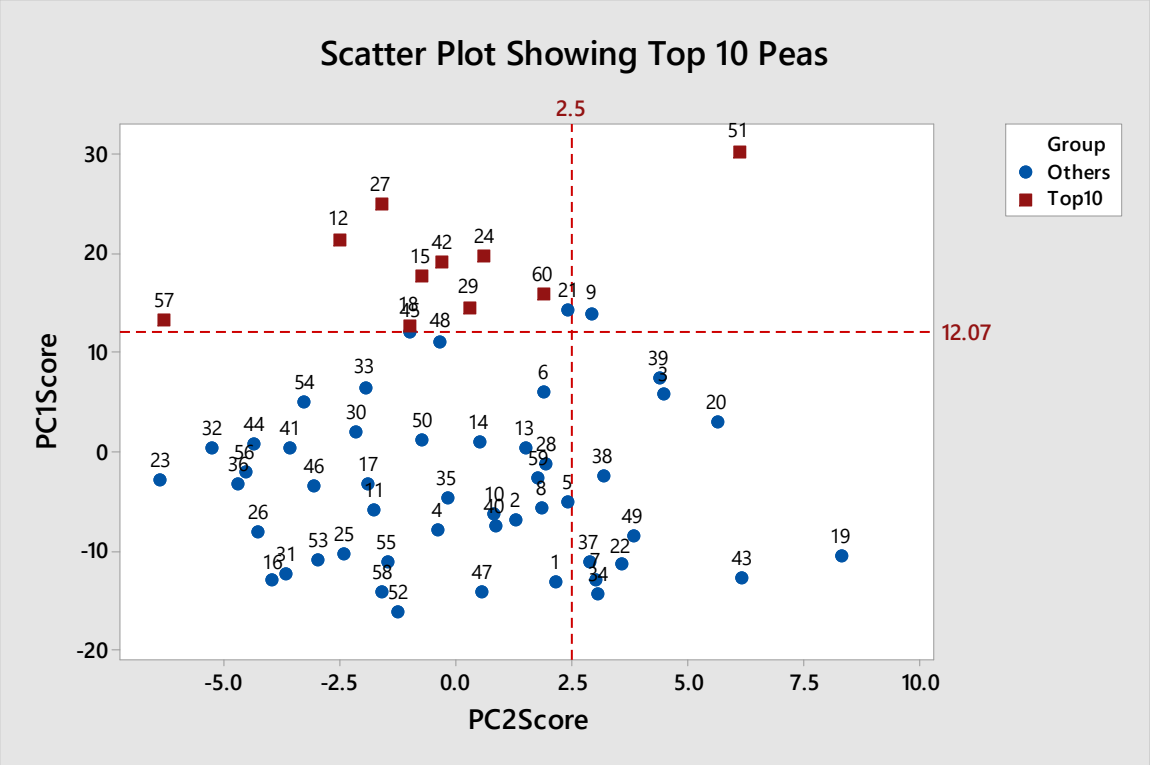
The total scores have a mean of 0 and Standard deviation of 11.87.

Therefore the top 10% of the peas will lie 1.29 Standard Deviations away from the mean which is 0 + 1.29 (11.87) = 15.31 . All Peas with total scores greater than 15.31 are Pea IDs 27, 51, 12, 57, 42, 24 and 15.

Based on my selection criteria of high PC1 and low PC2 values the top 10 peas are expected at the top left corner of the Scatter Plot having high PC1 Score values (Y-axis) and low PC2 (X-axis) magnitudes.

The red labeled peas are the top 10 rated peas.

Selected top 10% of peas based on the metric criteria in decreasing order is: 27, 51, 12, 57, 42 and 24.



Reference lines show the expected quadrant where top rated peas will be concentrated.

Pea 51 the visible outlier making it into the top selection because it possesses the highest PC1 score that provides it a cushion enough to rank high despite its relatively high PC2 score which gives it a lesser overall value. However, Pea 51 resultant score is still higher than that of other peas with less higher PC1 scores and much lower PC2 scores like Peas 48 and 21. Pea 51 is for individuals who care for overall contents and less about the appearance of the Pea.

Also notice that Pea 57 despite having a relatively low PC1 compared to the top 10's scores is ranked in the top 4 peas because it has a very low PC2 score, with much less whiteness and Colour1, which is this report's metric (high on PC 1 and low on PC2).

Pea ID	Tenderome	Dry mat	tter afte	crosePe	talGluc	talGluc	Flavou	Swee	Fruity	Off-flav	Mealin	Hardne	Whiteness	Colour1	Colour2	Colour3	Skin	Total Scores
27	200.00	25.50	27.97	1.80	6.70	5.80	2.66	2.66	1.42	6.10	6.67	7.75	4.27	4.97	5.63	4.53	1.65	26.52
51	200.00	28.10	28.75	1.00	6.50	6.60	2.28	2.23	1.29	6.45	6.70	7.83	5.53	7.30	4.36	3.50	1.63	24.19
12	196.00	25.00	27.86	2.40	6.70	7.00	3.41	3.18	1.82	4.64	6.24	7.43	4.26	4.84	5.95	4.55	1.85	23.90
57	200.00	24.10	26.30	3.20	6.40	6.40	4.14	4.91	2.49	3.50	5.37	6.58	3.89	4.45	6.57	6.96	1.71	19.44
42	193.00	24.30	25.79	2.20	5.50	5.80	3.10	3.43	1.80	4.86	6.22	7.07	4.14	5.28	5.58	3.70	2.05	19.40
24	196.00	24.10	26.98	2.50	5.40	6.10	3.22	3.21	1.95	4.41	6.24	7.27	4.60	6.03	5.60	4.16	1.67	19.05
15	188.00	23.90	25.43	2.80	6.40	5.50	3.39	3.28	1.98	4.50	6.04	7.14	4.35	5.09	5.58	4.40	2.03	18.45
29	146.00	23.50	25.14	2.90	6.60	5.70	3.72	4.35	2.20	4.08	6.50	6.27	4.99	5.53	5.56	5.34	1.82	14.28
60	180.00	23.80	24.75	3.10	6.10	5.60	3.70	3.86	2.33	4.11	6.18	6.83	5.15	5.77	5.29	4.41	1.99	13.94
18	176.00	22.50	26.05	3.00	5.20	4.90	3.96	4.48	2.30	3.94	6.23	6.41	4.47	5.11	5.72	4.97	2.05	13.74
45	168.00	22.20	23.74	3.40	5.60	5.10	3.75	4.30	2.22	4.27	6.10	6.27	4.06	5.14	5.87	4.22	2.23	13.09
21	166.00	22.10	24.47	2.70	5.60	5.00	3.77	3.97	2.17	4.37	6.47	6.55	4.95	6.05	5.31	4.39	2.21	11.87
48	168.00	22.20	23.73	3.30	4.50	4.40	3.87	3.88	2.23	4.06	5.99	6.31	4.45	5.53	5.78	4.98	1.92	11.41
9	158.00	21.70	24.53	2.90	6.10	6.40	3.79	3.88	2.31	3.52	6.24	5.73	5.39	6.30	5.13	5.23	2.01	10.98
33	162.00	20.60	24.11	3.50	4.70	4.40	4.71	4.68	2.67	3.19	5.32	5.91	4.32	4.77	6.22	4.86	2.33	8.39
54	160.00	22.00	23.91	3.90	5.20	5.30	5.24	5.09	3.30	2.80	4.67	5.99	4.34	4.93	6.22	6.60	2.15	8.16
32	140.00	21.40	23.15	4.40	4.20	4.00	5.53	5.41	3.68	2.47	4.72	5.78	3.88	4.34	6.47	6.79	1.80	5.57
44	123.00	19.30	22.66	3.90	4.50	4.70	5.46	5.41	3.27	2.97	5.15	4.98	3.61	4.30	6.60	5.43	2.37	4.98
30	153.00	21.50	24.18	4.70	3.80	4.40	5.43	5.19	3.47	2.40	4.43	5.26	4.46	4.78	5.72	5.90	1.61	4.10

Table 7 – Verifying results of the Pea selections

Table above shows the effect of the scoring metric based on PC1 and PC2. All top Peas values rank high on positively correlated values highlighted in orange for PC1 including Tenderometer, Dry matter before and after freezing, Total Glucose 1 & 2 and relatively lower on Flavor, Sweetness and Fruity qualities.

The Colour2 and Colour3 (Constituents of PC2) do not show a marked reducing pattern in the ranking as do the features for PC1 as there is relatively low variation in the Colour2 and Color 3 data compared to features described by PC1.

Summary

For this research the first 2 Principal components were the basis of the analysis explaining 86.3% of the variation of the pea dataset.

PC1 is the index describing the taste/ texture-content qualities of the peas.

PC2 is the index describing the external features of the pea based on skin, whiteness and Colour1, 2 and 3

The resultant combination of PC1 (positive) and PC2 (negative) applied over the standardized data served as the weighing metric for the top 10% of the peas.

The top 10 % of peas are Peas 27, 51, 12, 57, 42 and 24.

References

Jackson, Donald A. “Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches.” *Ecology*, vol. 74, no. 8, 1993, pp. 2204–2214. *JSTOR*, JSTOR, www.jstor.org/stable/1939574.

Brianna Elliot. “Why Green Peas are Healthy and Nutritious.” https://www.healthline.com/nutrition/green-peas-are-healthy#section1