**Exam 1 –Market Segmentation**

Banana Republic: Expansion to NCAA Audience

**Business Issue**

For the last five years Banana Republic has consistently reported negative sales during the company’s quarterly reports including a negative 1 percent for the third quarter of 2017 compared to a positive 4 percent from Old Navy and positive 1 percent from Gap [1,2]. It is no secret that Banana Republic is staying behind competitors like H&M, Kohl’s, and Macy’s. A study by RBC Capital found that between millennials and non-millennials, the brand of Banana Republic was disliked by at least twice as many people surveyed compared to people that viewed the brand favorably. That same study found that Gap and Old Navy had more positive perceptions [3]. The business approach of Banana Republic requires a drastic change.

One approach to reimage the brand and stop the negative regression of sales is to reach out to a passionate audience and participate in the “Madness” that comes every year in March. The idea is to start a line of university apparel that appeals to college basketball fanatics. A perfect time of the year to launch this new line would be in March when the NCAA has its Division I Basketball Tournament for 3-4 weeks. Ideally, if this is a success then the same idea can be used for other sports and times of the year with the next major events coming in May and June with the NBA playoffs.

**Target Population**

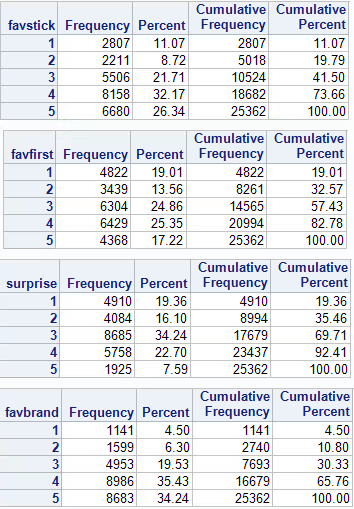
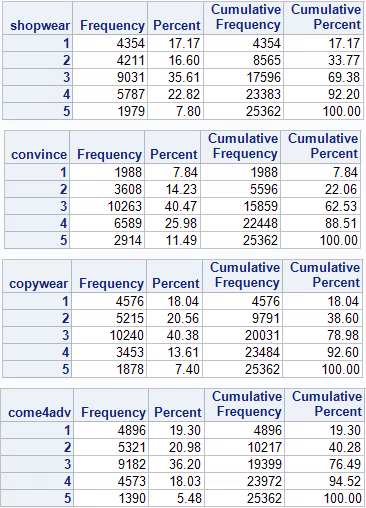
The target population would be everyone over the age of 18 years. NCAA data shows that March Madness Bracket participants and tournament viewers are almost evenly spread throughout ages above 18 years [4]. Because Banana Republic has a negative image with millennials and non-millennials, this age range of 18 and above works best to find out if adding NCAA apparel to the store would help change customers’ opinion of the brand and the increase sales.

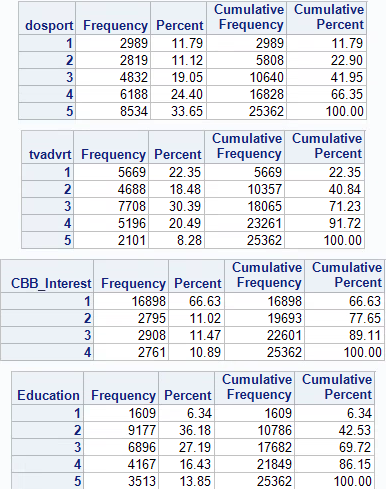
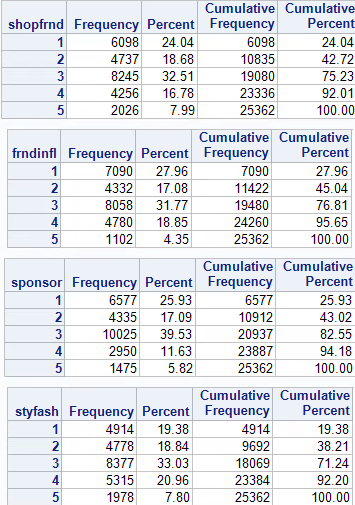
**Variables**

For this segmentation system, I started with reading out 19 variables from the National Consumer Study data set. Due to the fact that three of those variables had more than 75% of missing observations, I ended up with 16 variables. There are two sets of five variables each for factor analysis. The remaining six variables are for cluster analysis.

The data set includes 16 ordinal variables, 15 of them with five levels and one variable with four levels. The number of missing observations ranged from 1406 to 1977 (which is 5.49% to 7.74% of the observations) with the exception of the College Basketball Interest variable with 7431 missing observations. I decided to use the MI procedure in SAS for these ordinal variables to impute them.

There is an arbitrary missing pattern in the data so the available method with the MI procedure is the Fully Conditional Specification (FCS) Logistic Regression method [5]. There were 871 arbitrary missing patterns and 202 cases where 15 variables had no response. The number of observations to work with is 25362 observations.



**Table 1**. A combination of all 16 frequency tables for all the variables used. Education had no missing observations but all other 15 variables did, especially CBB\_Interest. Imputed variables using the FCS Logistic Regression method. The final data set contains 25362 observations.

**Factor Analysis: Loyalty and Influence**

First Set

* I have a few favorite clothing brands that I always stick with
* I always look for favorite brands first
* On several occasions I have surprised myself by buying clothing brands I normally don’t
* I usually only shop at my favorite stores because I know they have brands I like
* I am drawn to stores I normally don’t shop at by sales

I want to find if respondents from the survey are loyal to specific stores and brands.I think this first set of questions will measure that loyalty. Banana Republic wants their current customers to be loyal and it would really help if they can find potential customers that like to be loyal to specific stores and brands.

Second Set

* I am good at convincing others to try new things
* People often copy what I do or wear
* People come to me for advice before buying new things
* I prefer to shop with my friends
* My friends opinion of a store influences whether I shop there

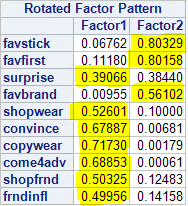
I think this second set of questions will measure whether respondents influence others in shopping habits or are the ones being influenced. It would be great if current Banana Republic customers influence their social network on apparel shopping habits. This ties to the first set of questions. If they are loyal to the brand then they will inadvertently guide new customers to the store they like (hopefully it is Banana Republic).

I chose Principal Component Analysis (PCA) as the extraction technique to use since the goal is to form two uncorrelated linear combinations of the observed variables. The PCA method has the advantage that factor scores produced are correlated with the factors in the analysis so it is easier to interpret the results.

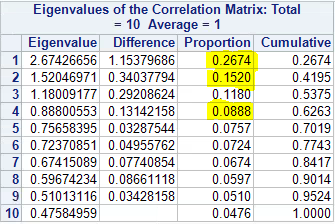
I used the varimax orthogonal rotation because I want the factors produced to be as uncorrelated to each other as possible. This will produce well separated clusters in the cluster analysis.

Two factors were extracted. By looking at the rotated factor matrix, I found the higher absolute number for each variable because this shows there is a stronger association between the factor and the variable. As shown in Table 2, Factor 1 consists of seven variables and Factor 2 consists of three variables. It appears that Factor 1 represents Influence on Others and Factor 2 represents Loyalty to Stores/Brands.

From the Eigenvalues of the Correlation Matrix I calculated the percent of variance that is explained by each factor. Each eigenvalue represents the total variance explained by each factor. By adding the proportions of the three variables that makeup Factor 1 and adding the proportions of the seven variables that makeup Factor 2, the total variance explained by each factor can be calculated. Factor 1 explains 50.82% of the variance and Factor 2 explains 49.17% of the variance.

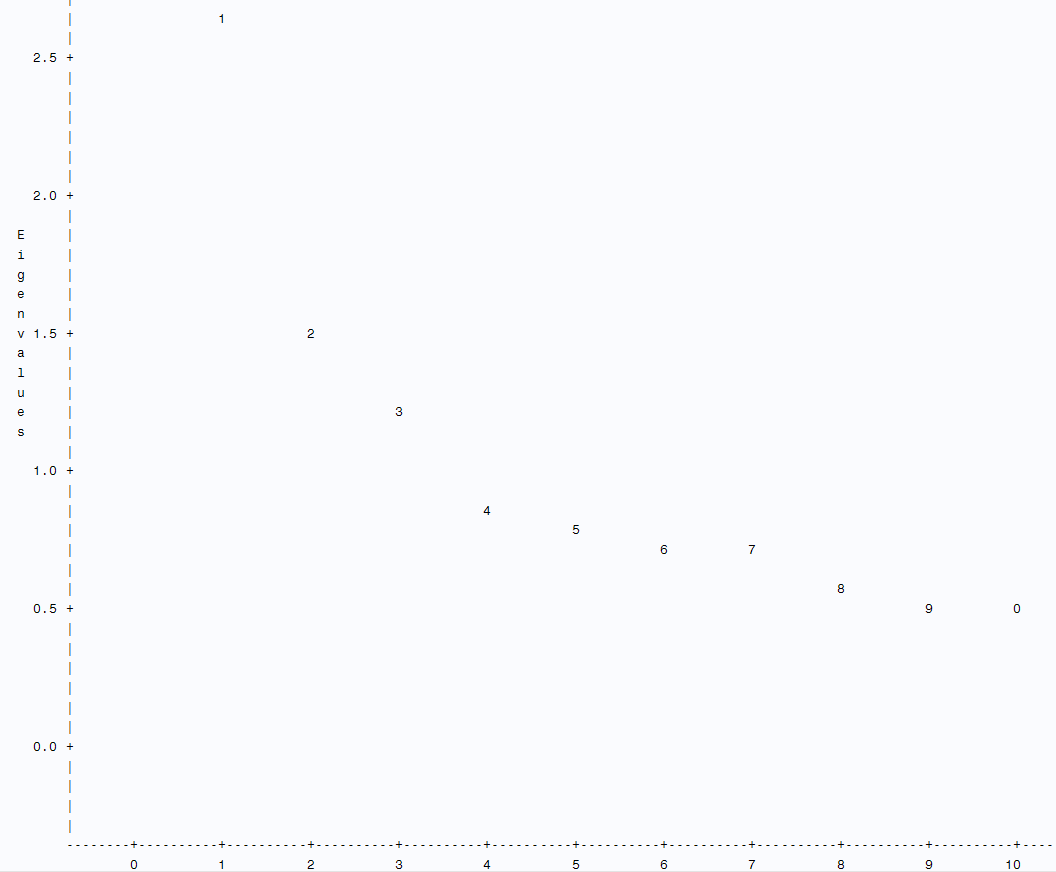


**Table 2**. Factor 1 consists of seven variables of which three are decently good with loadings above 0.65 and one poor loading of 0.39. Factor 2 consists of the remaining three variables of which two have fabulous loadings above 0.80 and the third loading is okay.



**Table 3**. The proportions of Factor 1 are highlighted. By adding those three proportions, the total variance explained by Factor 1 can be calculated. Adding the remaining proportions that make up Factor 2 give the total variance explained by Factor 2.

Scree Plot of Eigenvalues

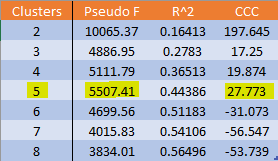


**Figure 1.** Scree plot showing the ideal number of factors on the largest drop and subsequent bend and flat line of the factors.

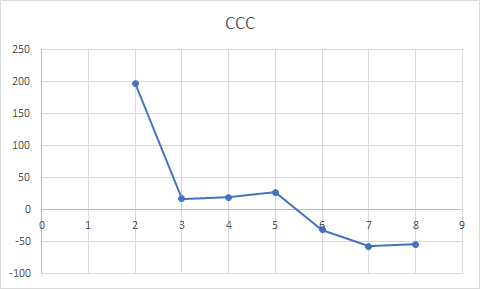
The ideal pattern for a scree plot is to distinctly see a bend that is preceded by a steep slope and followed by an almost flat line. The number of factors are represented by the first point where the flat line trend begins. In this case, it is a little difficult to see because no factor accounts for a drastically larger portion of the variability. However, a bend is distinguishable at factor two which corresponds with the rotated factor matrix. All other factors afterwards are between 0.5 and 1.5.

**Cluster Analysis**

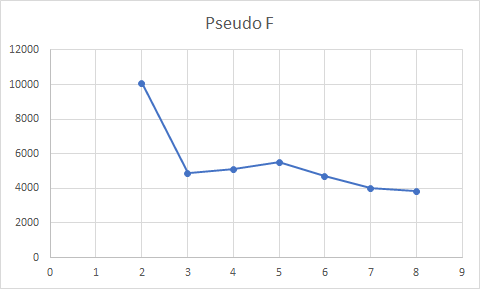
For both the Cubic Clustering Criterion (CCC) and the pseudo F statistic, the first local maximum suggests the number of clusters. In this case, as seen in Figure 2 and Figure 3, both diagnostic statistics show that the first local maximum is with five clusters. So the suggestion from the pseudo F statistic and the CCC is that there are five clusters.



**Table 4**. Diagnostic statistics pseudo F statistic, R^2, and CCC from a k-means cluster procedure.



**Figure 2**. The CCC for two through eight clusters from a k-means cluster procedure. The first local maximum is at five clusters. The large negative numbers suggest there might be outliers.



**Figure 3**. Pseudo F statistic for two through eight clusters. The local maximum is at five clusters, agreeing with the CCC.

After performing the gap analysis, the number of clusters that should be selected for the solution is six clusters. This was obtained by starting with the “gap number” corresponding to the least number of clusters, in this case two clusters. This “gap number” has to be greater than the “gap number” of the next cluster after subtracting the adjusted standard deviation of that next “gap number”. If that rule is not met then we proceed to the next number of clusters until it is met. The cluster number that meets that requirement is the number of clusters that the gap analysis suggests.

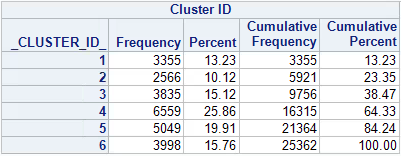
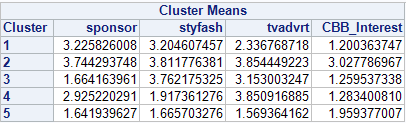
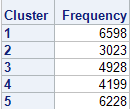


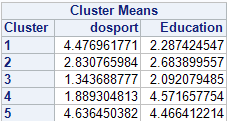
Table 5.

Taking a look at the unstandardized cluster means of the four driver variables used with the five clusters (number of clusters suggested by diagnostic statistics) we can see can see some differences and similarities between the segments. Two clusters pop right out: Cluster 2 and Cluster 5. The former seems to be the cluster that Banana Republic should be looking for since the means are above three across the table. People in this cluster like to keep up with the latest styles and fashions, they pay attention to TV advertising and talk about it, they like sports enough that they look at what companies sponsor teams and events, and they are interested in college basketball specifically. If Banana Republic were to release a new line of sports apparel for this time of year with the NCAA Tournament and March Madness, this is the cluster to focus on. This cluster can be called the Fashionable Talkers. The only issue is that it is the smallest cluster in terms of frequency but none of the clusters have too many observations either. The other cluster that sticks out is Cluster 5. This cluster is on the other side of the spectrum compared to Cluster 2. People in this cluster do not care about the latest styles and fashions, they do not pay attention to TV advertising (so they most definitely do not talk about it), and they do not care about companies sponsoring sport related things. However, this cluster is somewhat interested in college basketball and it is the second largest in terms of frequency. Banana Republic might have some success with this cluster only if the people are enticed by a sport that they are moderately interested in. This cluster can be called the Gloomy Potential because there might be a lot to gain from this cluster.

**Table 6.** Table on the left shows the unstandardized cluster means for driver variables using k-means clustering. Clustering method found five clusters. Cluster 2 and Cluster 5 pop out as being on opposites sides of the spectrum. Table on the right has the frequency for each cluster.

Using two more variables that are relevant to the business problem, the pattern stated above involving Cluster 2 and Cluster 5 is present. For people that do sports every week and with a range of their education, both clusters are on opposite sides again. Cluster 2 is indifferent about exercising every week and their education includes high school and some college. Cluster 5, on the other hand, takes exercises every week seriously and their education includes graduate school.



**Table 7.** Unstandardized cluster means for two descriptor variables using k-means clustering. The clustering method found five clusters.Cluster 2 and Cluster 5 are, again, on opposite sides.

**Summary**

In this segmentation for Banana Republic it has been found that there are two unobserved factors from the variables obtained from the NCS data set. The first factor contains seven variables and is associated with the influence that people have on others with shopping habits. The second factor contains three variables that are associated with being loyal to a certain store or brand for apparel. Even though the first factor has more variables associated with it, the second factor has much better loadings (much stronger association between variables and factor). This is important because if Banana Republic can get to people that are loyal to specific stores and brands then the image of the company can start going in a positive direction in sales and marketing. It was also found through gap analysis that the suggested number of clusters in the data set was six clusters and through diagnostic statistics the suggested number of clusters was five. None of the clusters are in the extreme in terms of having too many or too few observations. By evaluating the means of the driver variables in the cluster analysis, it was found that Cluster 2 is composed of great candidates that can help Banana Republic in two ways: the company would gain customers that are not only loyal to specific stores/brands but also like to share their ideas about apparel and are influential in others’ opinions when it comes to shopping.

**References**

1. Business Wire. Gap Inc. Reports Third Quarter Results. November 16, 2017. <https://www.businesswire.com/news/home/20171116006321/en/Gap-Reports-Quarter-Results>
2. Duprey, Rich. Motley Fool. Should Gap Just Kill the Banana Republic Brand. October 17, 2017. <https://www.fool.com/investing/2017/10/17/should-gap-kill-the-banana-republic-brand.aspx>
3. Trefis Team. Nasdaq. Banana Republic Continues to be a Drag on Gap Inc. October 03, 2016. <https://www.nasdaq.com/article/banana-republic-continues-to-be-a-drag-on-gap-inc-cm688030>
4. The Nielson Company. Method Behind March Madness: Why and How Americans Engage with the NCAA Basketball Tournament. March 16, 2017. <http://www.nielsen.com/us/en/insights/news/2017/method-behind-march-madness-why-and-how-americans-engage-with-ncaa-basketball-tournament.html>
5. SAS. The MI Procedure, Imputation Methods. <https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mi_sect019.htm>