MARKET SEGMENTATION: XBOX NOW17

1. Select a problem or business issue to apply your market segmentation to. This may be a particular brand, a category of products or services, or a particular industry that you create the market segmentation for. Be sure to state your business problem.

I performed an entertainment market segmentation utilizing Simmons' National Consumer Survey. My primary variables are all psychographic questions addressing the population that are early technology adaptors who spend their money and time on the latest video games.

BUSINESS PROBLEM: I was recruited by XBOX to see where they should target an advertising campaign for the forethcoming release of their newest XBOX gaming system.

2. Select your target population that you are going to segment. Many times this just turns out to be the U.S. adult population 18 years of age or older (the entire NCS data set). Other times it might be something different such as adults aged 18-24 years or women 18-49 or people with digital tablets or frequent movie goers or tequila drinkers – the list is pretty endless. If you are feeling adventurous, you might take on a trend analysis in addition to your market segmentation – I am providing you with all four quarters of 2012 NCS data. Be sure to tell me what your target population is and why you picked it.

The target population will be the entire NCS data set (25000) because there are so many different genres and modes of enjoying video games being introduced. Additionally, purchasers of video games and cutting edge technology are often parents/guardians of the end users, so there is no upper bound to the age range.

3. Clean and prepare the variables you are going to use in your segmentation system. HOW ADD THE DRIVER (OTHER 2 VARS TO THE EQUATION) - This includes both driver variables in the statistical procedures as well profile variables. Please produce a frequency listing of all the variables you are using for the analysis so that I can see that the data are clean.

top_graphics_agree_alot	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	25065	98.05	25065	98.05
1	499	1.95	25564	100.00

top_graphics_agree_alittle	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	24940	97.56	24940	97.56
1	624	2.44	25564	100.00

top_graphics_neither	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	23860	93.33	23860	93.33
1	1704	6.67	25564	100.00

top_graphics_disagree_alittle	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	24840	97.17	24840	97.17
1	724	2.83	25564	100.00

top_graphics_disagree_alot	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	22802	89.20	22802	89.20
1	2762	10.80	25564	100.00

topgraphics	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2762	43.75	2762	43.75
2	724	11.47	3486	55.22
3	1704	26.99	5190	82.21
4	624	9.88	5814	92.10
5	499	7.90	6313	100.00
Frequency Missing = 19251				

spendmoney	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3426	54.13	3426	54.13
2	737	11.64	4163	65.78
3	868	13.71	5031	79.49
4	662	10.46	5693	89.95
5	636	10.05	6329	100.00
Frequency Missing = 19235				

spendtime	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3571	56.34	3571	56.34
2	861	13.58	4432	69.93
3	814	12.84	5246	82.77
4	522	8.24	5768	91.01
5	570	8.99	6338	100.00
Frequency Missing = 19226				

mainent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3410	53.43	3410	53.43
2	826	12.94	4236	66.37
3	811	12.71	5047	79.08
4	734	11.50	5781	90.58
5	601	9.42	6382	100.00
Frequency Missing = 19182				

first	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	10617	43.86	10617	43.86
2	4876	20.15	15493	64.01
3	5846	24.15	21339	88.16
4	2015	8.33	23354	96.49
5	850	3.51	24204	100.00
Frequency Missing = 1360				

pay	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	12102	50.17	12102	50.17
2	4323	17.92	16425	68.09
3	4966	20.59	21391	88.68
4	1838	7.62	23229	96.30
5	892	3.70	24121	100.00
Frequency Missing = 1443				

keepup	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2877	11.98	2877	11.98
2	2791	11.62	5668	23.60
3	7669	31.92	13337	55.52
4	7094	29.53	20431	85.05
5	3591	14.95	24022	100.00
Frequency Missing = 1542				

love	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5175	29.31	5175	29.31
2	4450	25.21	9625	54.52
3	8030	45.48	17655	100.00
Frequency Missing = 7909				

buyfriendsoc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3501	32.22	3501	32.22
2	2066	19.01	5567	51.23
3	2798	25.75	8365	76.98
4	1955	17.99	10320	94.97
5	547	5.03	10867	100.00
Frequency Missing = 14697				

maincharuses	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	7142	30.44	7142	30.44
2	4377	18.66	11519	49.10
3	9238	39.38	20757	88.47
4	1938	8.26	22695	96.74
5	766	3.26	23461	100.00
Frequency Missing = 2103				

adskids	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3916	16.84	3916	16.84
2	2719	11.69	6635	28.54
3	10635	45.74	17270	74.28
4	4371	18.80	21641	93.08
5	1610	6.92	23251	100.00
Frequency Missing = 2313				

Males	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	14454	56.54	14454	56.54
1	11110	43.46	25564	100.00

Marital	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	7645	29.91	7645	29.91
1	17919	70.09	25564	100.00

Nintendo	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	21385	83.65	21385	83.65
1	4179	16.35	25564	100.00

Xbox	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	23148	90.55	23148	90.55
1	2416	9.45	25564	100.00

Playstation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	22502	88.02	22502	88.02
1	3062	11.98	25564	100.00

CNN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	24937	97.55	24937	97.55
1	627	2.45	25564	100.00

Netflix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	22408	87.65	22408	87.65
1	3156	12.35	25564	100.00

FACTOR ANALYSIS

4. Be sure in your segmentation system to pick at least two sets of questions that measures two different constructs so you have to apply a factor analysis to the set of questions so that you can demonstrate your knowledge of factor analysis. Once that is done, the next step is to do a factor analysis of the set(s) of questions that you selected that measures a latent, unobserved construct.

    a. List out the questions that you selected to do the factor analysis on:
        (1) I need games on computer with top of the line graphics.
        (2) I spend more money on video games than music or movies.
        (3) I spend more time playing video games than watching television.
        (4) Video games are my main source of entertainment.
        (5) I'm the first g my friends to have the latest in electronic equipment.
        (6) I'll pay about anything for an electronic product that I really want.
        (7) I try to keep up with developments in technology.
        (8) I love to buy new gadgets and appliances.

    b. Tell me what latent unobserved construct(s) you think they measure.

The 1st construct that I believe will be measured is whether or not the person spends their money, time and interest on video games with good graphics.

The 2nd construct will measure whether or not the person is an early adaptor of technology – i.e. buys the latest video game systems, etc.

    c. Decide which extraction technique to use and tell me why

I will use Principal Component Analysis (PCA). PCA reduces data complexity by analyzing and retaining key components. PCA is a variable reduction technique that will convert observations of possibly correlated variables using an orthogonal transformation in to principal components, where the first one or two components explain a large proportion of the data's variance.

    d. Decide which rotation method you are going to use and why

I am using the Varimax Rotation technique in my cluster analysis because the factors are orthogonal and are more likely to yield well divided. The orthogonal rotation positions each variable to become strongly associated with just one factor, and thus reducing the amount of data needed to be manipulated in the cluster analysis.

    e. Run the factor analysis

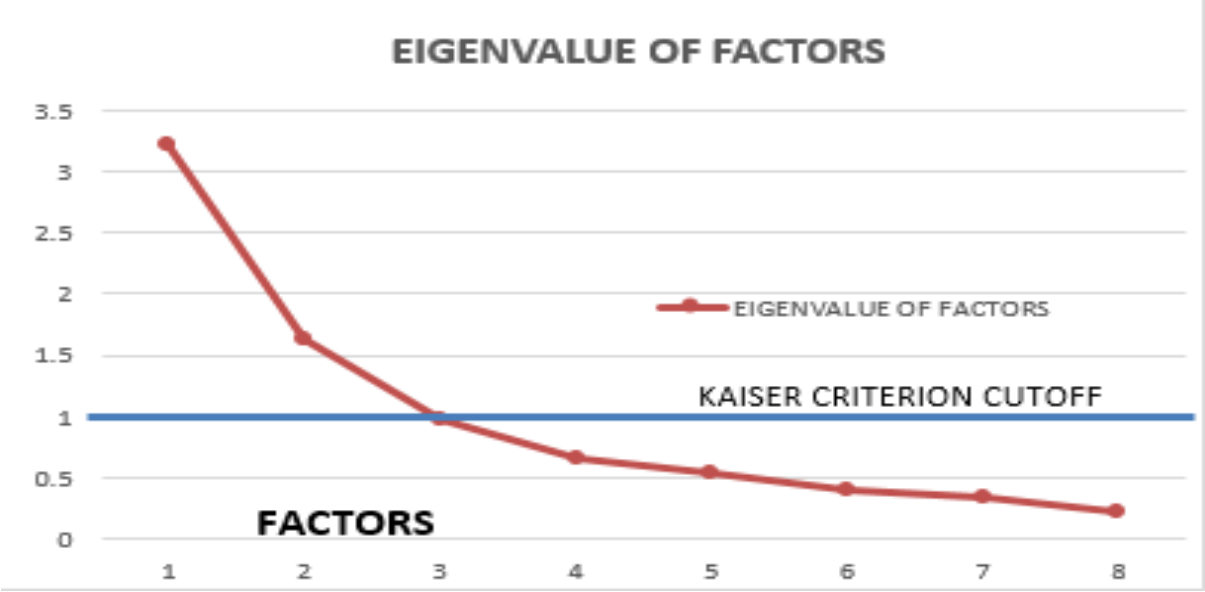
Eigenvalues of the Correlation Matrix: Total = 8 Average = 1	Eigenvalue	Difference	Proportion	Cumulative
1	3.2184859	1.5769707	0.4023	0.4023
2	1.6415152	0.6674768	0.2052	0.6075
3	0.9740384	0.3149412	0.1218	0.7293
4	0.6590972	0.116755	0.0824	0.8116
5	0.5423422	0.1490818	0.0678	0.8794
6	0.3932604	0.045641	0.0492	0.9286
7	0.3476194	0.1239781	0.0435	0.972
8	0.2236413		0.028	1

Final Communality Estimates: Total = 4.860001							
topgraphics	spendmoney	spendtime	mainent	first	pay	keepup	love
0.41723616	0.69146545	0.79519762	0.79974821	0.66686366	0.60644648	0.28804188	0.59500166

f. How many factors were extracted?

Factors with eigenvalues greater than one are extracted, and, according the table in part 4e, only two factors were extracted that met this criterion. These two factors also explain the greatest amount of the variance, approximately 61% (.4023+.2052) .

g. Make a pretty scree plot – how do you interpret it?



I will extract any factor with an eigenvalue greater than 1 (above the Kaiser Criterion cutoff, or KCC). This scree plot show that only 2 factors pass this test and will be extracted. It is a metaphorical mountain, and when the rock of eigenvalues starts rolling downwards, the KCC shows when the variance comes to a rest.

h. What criteria was used to determine number of factors? How does that work?

The KCC, or Kaiser Criterion cutoff (only eigenvalues greater than 1 are extracted) is the criteria used to determine the number of factors. KCC is used because the eigenvalues shown in the scree plot are also the measure of the proportion of variance that each factor explains. Thus, after the first 2 factors, the other factors are not contributing much variance at all, and unimportant.

i. What percent of the variance is explained by the factors?

Eigenvalues of the Correlation Matrix: Total = 8 Average = 1	Eigenvalue	Difference	Proportion Variance	Cumulative
1	3.2184859	1.5769707	0.4023	0.4023
2	1.6415152	0.6674768	0.2052	0.6075

These 2 factors explain approximately 61% (.4023+.2052) of the variance .

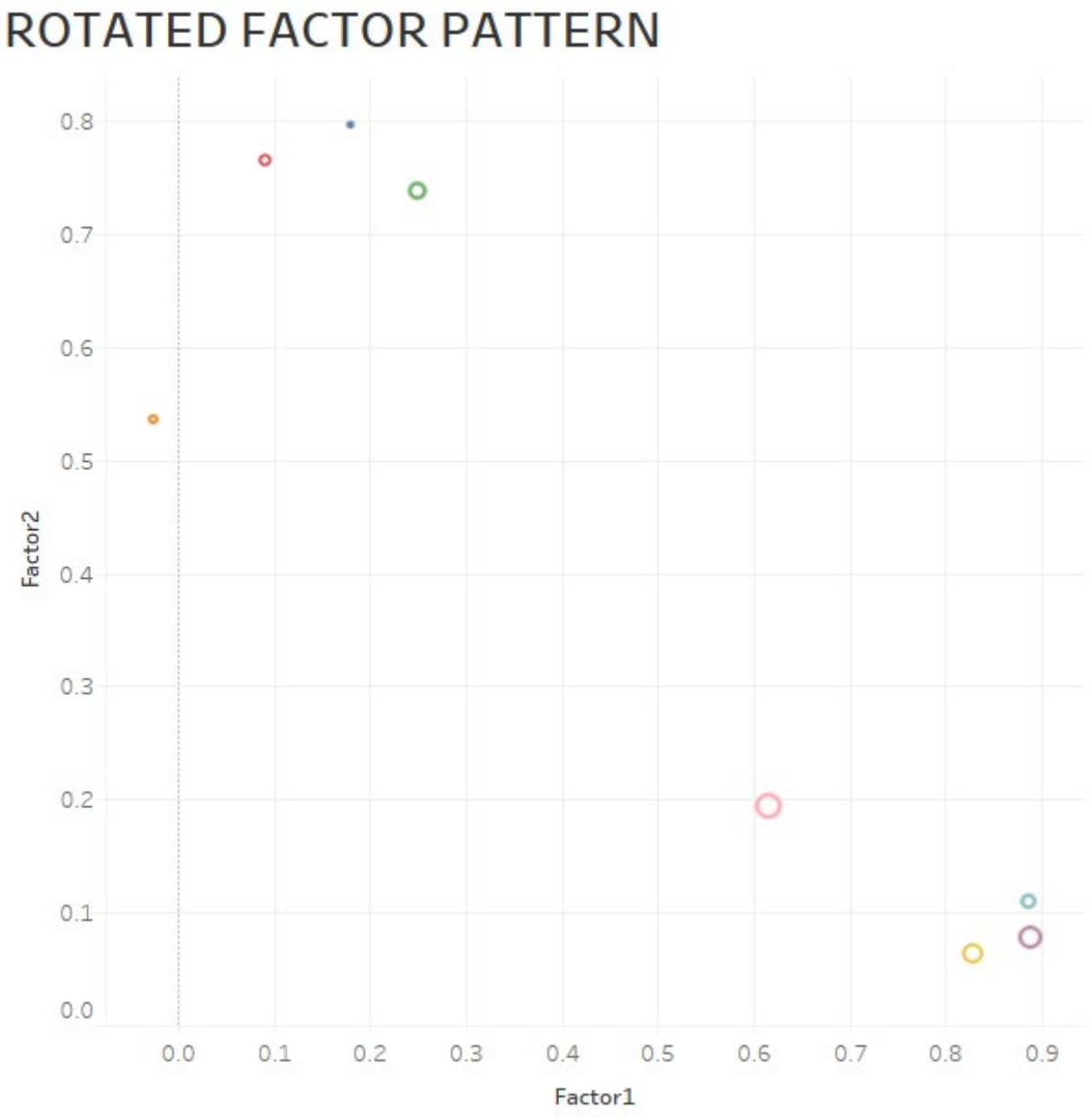
j. Interpret the rotated factor matrix loadings and label the factor(s)

Rotated Factor Pattern		
	Factor1	Factor2
topgraphics	0.61596	0.19449
spendmoney	0.8291	0.0637
spendtime	0.88826	0.07874
mainent	0.88761	0.10907
first	0.17974	0.79659
pay	0.24894	0.73789
keepup	-0.02586	0.53607
love	0.09092	0.76599

The Rotated factor matrix loadings show a distinct separation of variable in to 2 factors. I would interpret the 2 factors to be: (Factor 1) a video gamer who spends all of their money and time on video games with the best graphics – aka vgamejunkie, and (Factor 2) an early adaptor of technology who loves to buy all gadgets the day that they are released no matter what the cost – aka earlyadopt.

The variable scores highlighted in red loaded highest on the latent variable, factor 1. The variable scores highlighted in blue loaded highest on the latent variable, factor 2.

In more detail, for example, looking at the variable *keepup* that scored a -0.026 for Factor 1 could mean that people who are really in to playing video games are really not interested in keeping up with new developments in technology – the negative score represents the negative relationship between the variable and the factor, and vice versa. Factor 2 scored a 0.77 for the variable *love* representing the fact that this factor generally feel good about loving to buy new gadgets and appliances.



k. Output the factor scores for the next step

The following is a sample of the factor score for the first 100 respondents out of the 25000 total:

Obs	my_id	vgamejunkie	earlyadopt
1	1985951	.	.
2	1985952	.	.
3	1985954	.	.
4	1985955	.	.
5	1985960	.	.
6	1985961	.	.
7	1985963	.	.
8	1985964	.	.
9	1985966	.	.
10	1985967	.	.
11	1985969	.	.
12	1985970	.	.
13	1985977	.	.
14	1985979	.	.
15	1985980	.	.
16	1985981	.	.
17	1985986	.	.
18	1985987	.	.
19	1985989	.	.
20	1985990	.	.
21	1985992	.	.
22	1985994	.	.
23	1985995	.	.
24	1985997	.	.
25	1985998	.	.
26	1986022	-0.76221	-1.20773
27	1986023	.	.
28	1986027	.	.
29	1986030	.	.
30	1986031	.	.
31	1986032	0.33687	0.86412
32	1986033	.	.
33	1986034	.	.
34	1986038	.	.
35	1986041	.	.
36	1986043	.	.
37	1986049	.	.
38	1986050	.	.
39	1986052	.	.
40	1986053	.	.
41	1986058	.	.
42	1986059	.	.

Obs	my_id	vgamejunkie	earlyadopt
43	1986289	.	.
44	1986290	.	.
45	1986294	.	.
46	1986315	.	.
47	1986316	-0.46694	0.93733
48	1986319	-0.96503	-0.05927
49	1986320	-0.77970	-0.75234
50	1986321	.	.
51	1986338	.	.
52	1986340	0.05962	-0.33905
53	1986342	.	.
54	1986343	.	.
55	1986345	.	.
56	1986346	.	.
57	1986348	.	.
58	1986349	.	.
59	1986355	.	.
60	1986360	0.86131	-1.23506
61	1986362	.	.
62	1986366	.	.
63	1986367	1.75100	0.04192
64	1986370	.	.
65	1986371	.	.
66	1986372	-0.60522	-1.73775
67	1986373	2.48483	-2.42842
68	1986374	.	.
69	1986375	.	.
70	1986386	.	.
71	1986388	.	.
72	1986393	.	.
73	1986394	.	.
74	1986395	.	.
75	1986397	.	.
76	1986398	2.82897	-0.46730
77	1986402	.	.
78	1986405	.	.
79	1986407	.	.
80	1986413	.	.
81	1986418	.	.
82	1986424	.	.
83	1986425	.	.
84	1986432	.	.
85	1986434	.	.
86	1986446	.	.
87	1986447	.	.
88	1986448	.	.

Obs	my_id	vgamejunkie	earlyadopt
89	1986450	.	.
90	1986467	.	.
91	1986468	.	.
92	1986470	.	.
93	1986473	.	.
94	1986474	.	.
95	1986476	.	.
96	1986491	.	.
97	1986492	.	.
98	1986507	.	.
99	1986509	.	.
100	1986510	-0.80930	-0.24322

CLUSTER ANALYSIS

5. The next step is the cluster analysis. Here I would like you to use at least one factor from the factor analysis in step 4 plus a few more relevant variables (say 3-4) to utilize in your factor analysis. Here is how I would like you to proceed with the cluster analysis.
    a. You may wish to standardize the variables (other than the factor scores) to remove any influence due to scaling.

I did not standardize any of the variables that I chose.

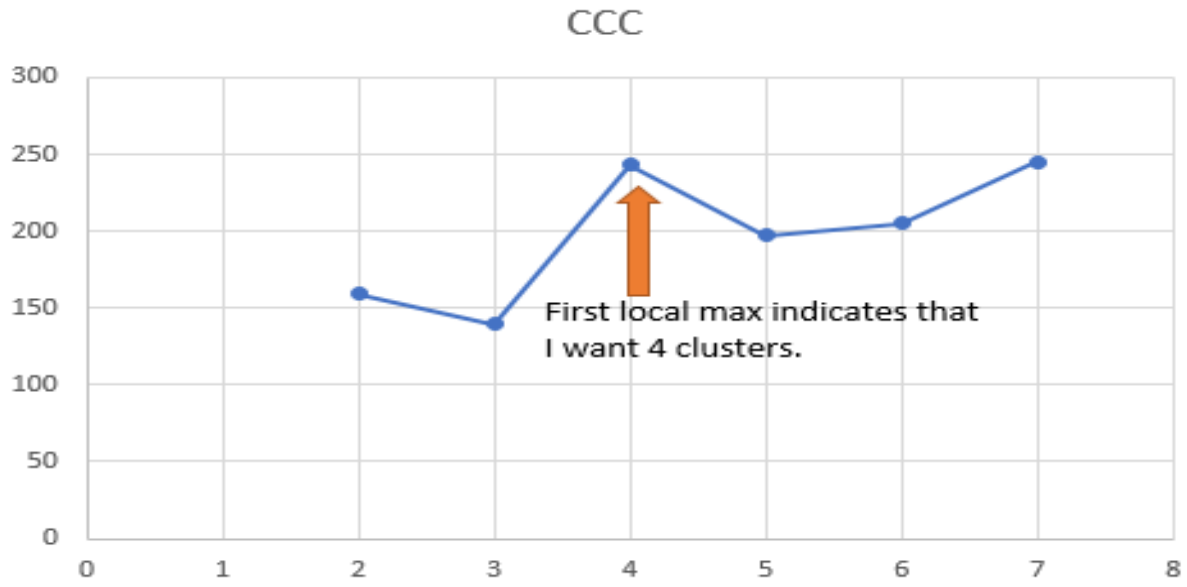
    b. Pick a bracket of number of clusters (say from 3 to 7) and then run a k means cluster procedure to collect the following diagnostic statistics from the runs: R square, cubic clustering criterion (ccc) and pseudo F statistic (the proc FASTCLUS in SAS is the easiest way to obtain these but you can also get at least the ccc from R).

I ran PROC FASTCLUS to get the following results:

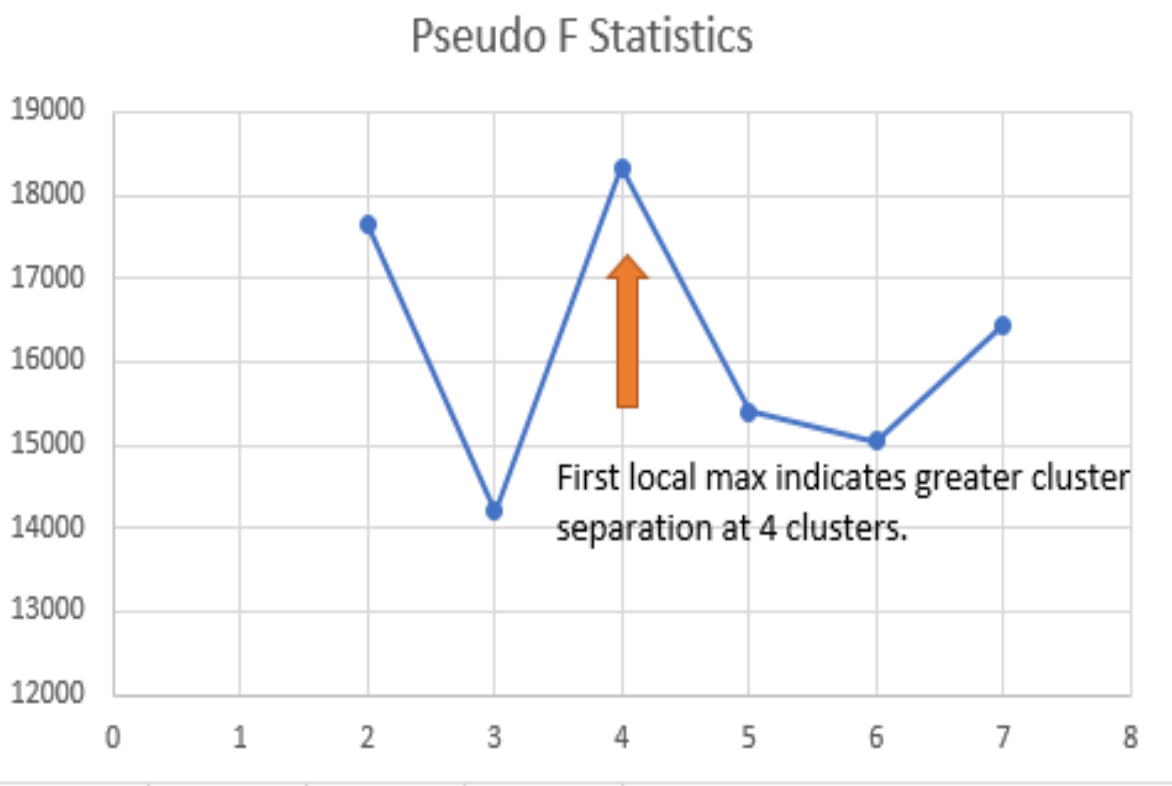
# of CLUSTERS	CCC	Pseudo F Statistics	R squared
2	158.887	17650.64	0.24524
3	139.596	14195.7	0.38777
4	243.255	18332.85	0.48146
5	196.972	15399.09	0.56041
6	204.87	15046.53	0.59874
7	244.892	16450.08	0.62854

c. Plot the diagnostic statistics against the number of clusters and apply the rules we discussed in class about the number of clusters that this graph might suggest to use (note that it may be the case that these diagnostics might not suggest any of the ones you ran. If so, note that).

I began by plotting the Cubic Clustering Criterion (CCC) to find the first local maximum that will indicate an optimal number of clusters to use:



I then plotted the pseudo f scores to find the first local maximum – this also infers the optimal number of clusters is 4:



Both psuedo-f and CCC suggest that I use 4 clusters in my analysis, however, I am going to run a GAP analysis first to see what the more up to date Monte Carlo estimate produces.

d. Now perform a gap analysis on the data. Tell me how you know how many clusters you should select as your solution. Don't point at the SAS output box that says here is the number of clusters – that's too easy. Tell me the logic of how it works and then do it for your data.

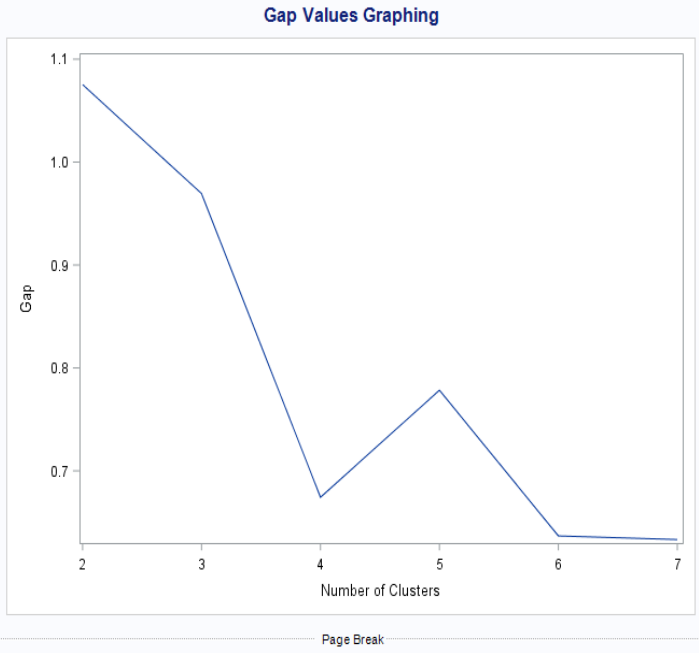
I wanted to see how many clusters that the GAP analysis would output using proc HPCLUS. I got the following results:

ABC Statistics

Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	9.0679	10.1431	1.0752	0.0139	1.0613
3	8.9192	9.8886	0.9694	0.0109	0.9585
4	8.7858	9.4601	0.6743	0.00735	0.6670
5	8.6536	9.4320	0.7784	0.0228	0.7555
6	8.5678	9.2046	0.6368	0.0116	0.6252
7	8.4801	9.1134	0.6333	0.0133	0.6199

Estimated Number of Clusters

Criterion	Number of Clusters
FIRSTPEAK	5



HPCLUS uses the Monte Carlo estimate versus the FASCLUS heuristic and is the preferred method by statisticians, thus I am going to proceed using the HPCLUS to cluster my segments in to the 5 optimal segments.

I then assigned each survey participant to a cluster using HPCLUS, and ran HPCLUS to cluster with 5 clusters I got:

the following results: .

Descriptive Statistics		
Variable	Mean	Standard Deviation
vgamejunkie	-0.012882	0.998365
earlyadopt	0.046367	0.954017
buyfriendsoc	2.408878	1.178786
maincharuses	2.247119	1.000836

Within Cluster Statistics			
Variable	Cluster	Mean	Standard Deviation
vgamejunkie	1	-0.5582	0.5242
	2	1.6673	0.6223
	3	-0.4967	0.5130
	4	-0.5129	0.4548
	5	0.9077	0.8382
earlyadopt	1	0.5867	0.8474
	2	-0.4620	0.8976
	3	-0.1575	0.8135
	4	-0.5539	0.7586
	5	0.7228	0.6952
buyfriendsoc	1	1.8198	0.7081
	2	1.7103	0.8226
	3	3.5890	0.6320
	4	1.2814	0.4501
	5	3.4219	0.7099
maincharuses	1	2.9109	0.6938
	2	2.0794	0.9199
	3	1.9359	0.8356
	4	1.3221	0.5415
	5	3.1339	0.6755

When I began to compare my HPCLUS results with my cluster numbers from PROC FASTCLUS, I realized that I had used a colon in my HPCLUS statements. ERROR WILL ROBINSON! I removed the colons in the HPCLUS, reran it and got the following results:

Estimated Number of Clusters	
Criterion	Number of Clusters
FIRSTPEAK	2

Cluster Summary								
		Distance from Cluster Centroid to Observation						
Cluster	Frequency	Maximum	Minimum	Average	SSE	Standard Deviation	Nearest Cluster	Distance to Nearest Cluster Centroid
1	1181	4.1157	0.5531	1.5913	3390.3	1.6943	2	2.2128
2	1162	4.3406	0.4645	1.6903	3808.0	1.8103	1	2.2128

It returned only 2 optimal clusters using HPCLUS

ABC Statistics					
Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	8.8816	10.0296	1.1480	0.00589	1.1421
3	8.7102	9.7576	1.0474	0.00948	1.0379
4	8.5413	9.4593	0.9180	0.0116	0.9064
5	8.4229	9.2057	0.7828	0.00791	0.7749
6	8.3686	9.0996	0.7310	0.00846	0.7226
7	8.2507	8.9581	0.7073	0.00652	0.7008

The local max adjusted gap (Gap – simulation standard deviation) is at 2 clusters, and none of the adjusted gaps are further away than 1 standard deviation from the previous gap number.

I realize now that I can not continue with only 2 clusters. My client, XBOX, would not be happy with me knowing that they spent all of their money and I have lumped an entire population in to 2 major groups that could almost have been guessed correctly with logical conclusions minus the data. Two clusters does not tell you the client enough information about their population.

Since I will not use 2 clusters, I will run the cluster analysis with k means PROC FASTCLUS with 4 clusters.

e. Run the cluster analysis with your choice of procedures.

I used PROC FASTCLUS to assign each person to a cluster based on their survey answers.

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	3430	vgamejunkie	541	-0.4451032	0.6337588	-1.2470494	2.6567832
		earlyadopt	541	-0.2491501	0.8997069	-2.1049792	3.2549463
		buyfriendsoc	1753	4.0228180	0.6271480	2.0000000	5.0000000
		maincharuses	3317	3.8794091	0.7383522	2.0000000	5.0000000
2	7998	vgamejunkie	1049	-0.2352975	0.7934205	-1.2470494	2.7893678
		earlyadopt	1049	0.7948204	0.7239764	-2.0556091	3.2549463
		buyfriendsoc	2692	1.6530461	0.5897505	1.0000000	3.0000000
		maincharuses	7899	2.9725282	0.4187669	1.0000000	5.0000000
3	5438	vgamejunkie	934	1.0528576	0.9951294	-1.0936786	3.1975241
		earlyadopt	934	0.0796485	0.9119627	-2.4284163	2.8742418
		buyfriendsoc	3385	3.3376662	0.6406833	1.0000000	5.0000000
		maincharuses	5286	2.0902384	0.6755573	1.0000000	5.0000000
4	7141	vgamejunkie	1206	-0.4110622	0.6594700	-1.1678458	3.1975241
		earlyadopt	1206	-0.6412671	0.7985943	-2.3734722	1.5526678
		buyfriendsoc	3037	1.2453079	0.4951163	1.0000000	3.0000000
		maincharuses	6959	1.1201322	0.3398361	1.0000000	3.0000000

Pseudo F Statistic =	19933.57
Approximate Expected Over-All R-Squared =	0.48146
Cubic Clustering Criterion =	270.107

WARNING: The two values above are invalid for correlated variables.

Cluster Means				
Cluster	vgamejunkie	earlyadopt	buyfriendsoc	maincharuses
1	-0.445103197	-0.249150065	4.022818026	3.879409105
2	-0.235297549	0.794820411	1.653046062	2.972528168
3	1.052857628	0.079648481	3.337666174	2.090238365
4	-0.411062244	-0.641267087	1.245307870	1.120132203

Cluster Standard Deviations				
Cluster	vgamejunkie	earlyadopt	buyfriendsoc	maincharuses
1	0.6337588037	0.8997069489	0.6271480343	0.7383521823
2	0.7934204849	0.7239763794	0.5897505038	0.4187669277
3	0.9951294387	0.9119626768	0.6406832666	0.6755573472
4	0.6594700037	0.7985942815	0.4951163466	0.3398360946

Distance Between Cluster Centroids				
Nearest Cluster	1	2	3	4
1	.	2.751753070	2.454090486	3.934855301
2	2.751753070	.	2.405658926	2.385552719
3	2.454090486	2.405658926	.	2.825216786
4	3.934855301	2.385552719	2.825216786	.

f. Produce a table of means for the driver variables used in the cluster analysis. You should use the unstandardized variables for means wherever possible. Discuss how well the cluster analysis worked from your evaluation of this table.

Cluster Means				
Cluster	vgamejunkie	earlyadopt	buyfriendsoc	maincharuses
1	-0.445103197	-0.249150065	4.022818026	3.879409105
2	-0.235297549	0.794820411	1.653046062	2.972528168
3	1.052857628	0.079648481	3.337666174	2.090238365
4	-0.411062244	-0.641267087	1.245307870	1.120132203

The unstandardized table of means worked well for the cluster analysis. I get in to more detain in the following table:

Cluster Means								
Cluster	vgamejunkie		earlyadopt		buyfriendsoc		maincharuses	
1	-0.4451032	LOW	-0.249150065	MEDIUM	4.022818	HIGH	3.8794091	HIGH
2	-0.2352975	MEDIUM	0.794820411	HIGH	1.653046	LOW	2.9725282	MEDIUM
3	1.05285763	HIGH	0.079648481	MEDIUM	3.337666	HIGH	2.0902384	MEDIUM
4	-0.4110622	MEDIUM	-0.641267087	LOW	1.245308	LOW	1.1201322	LOW

g. Select at least 3-4 more variables that are relevant to your business problem that are not drivers in the solution (be sure to include at least one demographic and one media variable) and construct a means table for each of these variables by cluster number. Note if you take the mean of a (0,1) variable it provides you with the proportion of individuals whose answered with the response coded 1.

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	3430	Males	3430	0.394461	0.488806	0	1
		adskids	3252	3.534748	1.101308	1	5
		Marital	3430	0.687755	0.463477	0	1
		Nintendo	3430	0.217201	0.412401	0	1
		Playstation	3430	0.160933	0.367523	0	1
		Xbox	3430	0.123324	0.328856	0	1
		CNN	3430	0.033236	0.179279	0	1
		Netflix	3430	0.146356	0.353514	0	1
		ComedyCentral	3430	0.087755	0.28298	0	1

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
2	7998	Males	7998	0.447862	0.497305	0	1
		adskids	7747	3.020266	0.939275	1	5
		Marital	7998	0.698675	0.458862	0	1
		Nintendo	7998	0.134409	0.341112	0	1
		Playstation	7998	0.098525	0.298041	0	1
		Xbox	7998	0.07827	0.268612	0	1
		CNN	7998	0.022131	0.147117	0	1
		Netflix	7998	0.09815	0.297535	0	1
		ComedyCentral	7998	0.057764	0.233312	0	1

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
3	5438	Males	5438	0.445936	0.497114	0	1
		adskids	5194	2.875241	0.973474	1	5
		Marital	5438	0.709636	0.453972	0	1
		Nintendo	5438	0.203935	0.402959	0	1
		Playstation	5438	0.178926	0.383326	0	1
		Xbox	5438	0.143435	0.350548	0	1
		CNN	5438	0.02078	0.142659	0	1
		Netflix	5438	0.17249	0.37784	0	1
		ComedyCentral	5438	0.092313	0.289494	0	1

Cluster	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
4	7141	Males	7141	0.429772	0.495078	0	1
		adskids	6813	2.383091	1.173957	1	5
		Marital	7141	0.717267	0.45036	0	1
		Nintendo	7141	0.159361	0.366038	0	1
		Playstation	7141	0.092144	0.289249	0	1
		Xbox	7141	0.073239	0.260547	0	1
		CNN	7141	0.026747	0.161354	0	1
		Netflix	7141	0.119591	0.324506	0	1
		ComedyCentral	7141	0.069038	0.253536	0	1

h. Name the different clusters in your solution and provide a half-page summary of the differences and similarities between the segments.

Cluster 1 is very strong in stating that ads help them choose what product to buy for their kids.

Analysis Variable : adskids						
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	3430	3252	3.5347478	1.1013084	1.0000000	5.0000000
2	7998	7747	3.0202659	0.9392754	1.0000000	5.0000000
3	5438	5194	2.8752407	0.9734741	1.0000000	5.0000000
4	7141	6813	2.3830911	1.1739572	1.0000000	5.0000000

Analysis Variable : Marital						
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	3430	3430	0.6877551	0.4634767	0	1.0000000
2	7998	7998	0.6986747	0.4588624	0	1.0000000
3	5438	5438	0.7096359	0.4539721	0	1.0000000
4	7141	7141	0.7172665	0.4503595	0	1.0000000

Cluster 4 has a higher percentage of married participants, however this is not a good discriminating variable because the means are too similar.

Analysis Variable : Males						
Cluster	**N Obs**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
1	3430	3430	0.394461	0.488806	0	1
2	7998	7998	0.447862	0.497305	0	1
3	5438	5438	0.445936	0.497114	0	1
4	7141	7141	0.429772	0.495078	0	1

Cluster 2 has the highest concentration of male participants, followed close behind by Cluster 3. Cluster 1 has the highest number of female participants.

Cluster	N Obs	XBOX Mean	Playstation Mean	Nintendo Mean
1	3430	0.123324	0.1609329	0.217201
2	7998	0.07827	0.0985246	0.134409
3	5438	0.143435	0.1789261	0.203935
4	7141	0.073239	0.092144	0.159361

Holding true to the statistics, Cluster 3 own more XBOX and Playstation systems, but Cluster 1 surprisingly owns the most Nintendo and is close to owning the most Playstation and XBOX systems.

Analysis Variable : CNN						
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	3430	3430	0.0332362	0.1792788	0	1.0000000
2	7998	7998	0.0221305	0.1471172	0	1.0000000
3	5438	5438	0.0207797	0.1426592	0	1.0000000
4	7141	7141	0.0267470	0.1613543	0	1.0000000

Cluster 1, more women, watches a LOT more CNN than the others, and luster 3 is last – gamers don't watch as much tv.

Analysis Variable : Netflix						
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	3430	3430	0.1463557	0.3535140	0	1.0000000
2	7998	7998	0.0981495	0.2975353	0	1.0000000
3	5438	5438	0.1724899	0.3778404	0	1.0000000
4	7141	7141	0.1195911	0.3245055	0	1.0000000

Cluster 3 has the largest market share of NETFLIX viewers, and Cluster 1 is in 2nd.

Analysis Variable : ComedyCentral						
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	3430	3430	0.0877551	0.2829797	0	1.0000000
2	7998	7998	0.0577644	0.2333121	0	1.0000000
3	5438	5438	0.0923134	0.2894944	0	1.0000000
4	7141	7141	0.0690379	0.2535364	0	1.0000000

Comedy Central viewing holds the same pattern as NETFLIX, only there is not as much of a separation. The difference would be that this is a cable program versus an internet based service where in NETFLIX commercials and advertising opportunities are more difficult to manipulate towards the viewers to ensure that they see your ad. This is important, because now I can tell Comedy Central that I can find them almost twice as many viewers in Cluster 3 than Cluster 2.

	Cool Parents	Average Person	Gamers	Married No Kids
Top Rated Segments	AdsKids		XBOX	Married
	Nintendo		Playstation	
	CNN		NETFLIX	
		Comedy Central		
	Xbox		Nintendo	CNN
	Playstation	AdsKids	AdsKids	
	NETFLIX	Main Character Uses	SOCIAL MEDIA	NETFLIX
Middle Rated Segments	Comedy Central			

Cluster Means								
Cluster	**vgamejunkie**		**earlyadopt**		**buyfriendsoc**		**maincharuses**	
1	-0.4451032	**LOW**	-0.249150065	**MEDIUM**	4.022818	**HIGH**	3.8794091	**HIGH**
2	-0.2352975	**MEDIUM**	0.794820411	**HIGH**	1.653046	**LOW**	2.9725282	**MEDIUM**
3	1.05285763	**HIGH**	0.079648481	**MEDIUM**	3.337666	**HIGH**	2.0902384	**MEDIUM**
4	-0.4110622	**MEDIUM**	-0.641267087	**LOW**	1.245308	**LOW**	1.1201322	**LOW**

CLUSTER NAMES:

Cluster 1: Cool Parent - Hip, Up to date, watching movies for ideas of what new products to use, using social media to stay current, watching Comedy Central and NETFLIX, buying video game systems, keep current on world affairs with CNN and watch ads for their kids is a good source of information. Highest concentration of female survey participants. Cluster 1 doesn't seem interested in spending money on the latest video games unless it is marketed in a way that they won't feel as if it is a new technology purchase, but more as a "my friend says its great". They are likely to use a product if they see a star in a movie use that product.

Cluster 2: Techies – Want the latest tech products when they are released and will pay anything almost anything to get it. Watch ads for their kids sometimes, but don't use social media for purchase ideas, don't spend a lot of money on video game systems, however they are slightly impressionable when a main character uses a product - they want to use that product.

Cluster 3: Gamers – They like to spend their time playing video games on the computer with great graphics or on their choice of one of the gaming systems they own. They relax by watching NETFLIX and Comedy Central and get ideas of purchases from their friends on social media. Majority are women.

Cluster 4: Childless – Lowest scoring for having kids by quite a lot. They watch CNN and NETFLIX more than some of the other clusters. Interesting that Cluster 4, with the lowest mean for watching ads for

their children, did not have more video game systems since they would have extra capital income, which leads me to believe that this is the business professional, career-oriented cluster, frugal, or just do not care about video game systems.

i. Write a one paragraph executive summary of the segmentation system.

EXECUTIVE SUMMARY:

I was recruited by XBOX to see where they should target an advertising campaign for the forthcoming release of their newest XBOX gaming system. I recommend that XBOX focus their advertisements on Cluster 1, Cool Parent. Cool Parents are influenced by advertisements for their children, what their friends think on social media and what the main character uses on a movie. XBOX should produce a cross-marketing advertising plan of action for their new game system – have an actor using the system in an upcoming movie, and having that upcoming movie as one of the new games that will be released soon after the movie release. I recommend that XBOX' social media team start up an advertisement of the new system that will entice the public to share the ad across platforms – increasing the opportunity that it reaches the eyes and ears of the highly suggestable Clusters 1 and 3 (Gamers).

XBOX can reach Cool Parents and Gamers audiences alike through a targeted advertising campaign aimed at Comedy Central audiences which are almost twice as large as they are with the Techies (Cluster 2).

FUTIRE ANALYSIS: This study puts to rest the myth that all gamers are male, as a majority are women. I would recommend that XBOX focus advertising on the female clientele, however that will be a topic for future analysis, as the question still begs as to whether or not the women merely the primary purchasers or the actual end-users of the gaming systems.