**Assignment 3, MGMT643, Spring 2016**

*R.L. Andrews*

The assignment will involve group work and each group will submit an Excel file electronically via e-mail to randrews@vcu.edu with the work for the entire group putting Assignment 3 in the subject line. You are to use the data from Assignment 1 found on my homepage http://www.people.vcu.edu/~randrews/ under the title **U.S. Census Bureau Data (Excel 2007)**. Use the tab in this data set file with 82 variables using the numbers 1 through 82 for the columns with numerical data. You are to consider using 40 numerical variables in the columns numbered: 2-4, 10-12, 29, 30, 33, 35, 36, 38-49, 51-57, 60-63, 75, 76, 78, 80-82. (Note that you may create appropriate new variables that use one or more of these variables.) Also use column **71, Total women-owned firms of 100 or more, percent 2002**, and convert it into a new categorical variable, **71X**, with four categories: **1.** **None** for all localities with a numerical value of 0; **2.** **Low** for values 0 < V71 < 23; **3.** **Medium** for values 23 ≤ V71 < 30; and **4.** **High** for values 30 ≤ V71. Hence numeric data for columns numbered 1, 5-9, 13-28, 31, 32, 34, 37, 50, 58, 59, 64-70, 72-74, 77 & 79 will not be analyzed. You are to perform the analyses below using the rows and numerical variables you decide are appropriate to include. **Deadline for submission is 9 AM Thursday, April 28.**

Each team will:

1. Consider their experience for the work done on Assignment 1 with these data and report the appropriate cases and variables that will be used for the parts of this assignment. **Exclude 11000&1, District of Columbia!** Clearly define new variables. For excluded variables and cases, give a reason for their exclusion.
2. Use the four categories for 71X and perform a multivariate analysis of variance to test for a significant difference between the four centroids of all the variables in your final data. Report your p-value & conclusion. Next determine which variable of all the variables in your final data set has the most significant difference between the means of the four categories and report the ANOVA p-value plus identify which of the four categories are significantly different from each other using α = .10.
3. Perform a principle component analysis on these data and calculate the scores for the first four principal components labeling them PRIN1 to PRIN4. Use the four categories for 71X and perform a multivariate analysis of variance to test for a significant difference between the four centroids of the four principal components. Report your results for this MANOVA test, including the MANOVA p-value and your conclusion about the existence of any differences. Comment on the similarity or dissimilarity between the p-values of the two preceding MANOVAs.
4. Build what you believe to be the best model using discriminant analysis to classify localities into one of the four categories for 71X using your original variable values (do not use the principal component scores). Describe how you arrived at your best model and report its classification rate. Use the data for the District of Columbia from the original data set and use your model to predict its classification into one of the four categories. Use the model to estimate the probability of it being in each of the four categories of None, Low, Medium and High given its data values and tell how you obtained these estimates.
5. Using your original variable values, build what you believe to be the best logistic regression model to predict if the value of 71X for a locality is None. Describe how you arrived at this model and report its classification rate. Use the data for the District of Columbia from the original data set and use your model to predict its classification as being None or not based on these variable values. Use the model to estimate the probability of it being None given its data values and tell how you obtained this estimate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group A | Group B | Group C | Group D | Group E | Group F |
| Liu | Bunnell | Masters | Smith | Olson | Bristow |
| Sleeth | Abed | Gupta | Kotak | Wood | Sink |
| Abraham | Mudigonda | Pahari | Russell | Banjo |   |