DAPT 622 Assignment 2

Daniel Erb

February 1, 2019

```r
library(data.table)

# file path for candy data
candy.file <- "G:/My Drive/vcu/Spring 2019/Statistics/Assignments/Assignment 2/DAPT_622_Assignment_2/dat

# 75 observations, not 69 like the read.table showed...
cd <- read.csv(candy.file)

# file path for demographics data
demo.file <- "G:/My Drive/vcu/Spring 2019/Statistics/Assignments/Assignment 2/DAPT_622_Assignment_2/data

# 440 observations for this data set
dd <- read.csv(demo.file)

rm(candy.file)
rm(demo.file)
```

Part I. Consider the candy bar data again. A study was conducted to investigate the nutrition content of a variety of candy brands and products. The data set includes the following variables (most of which should be self-explanatory):

Perform a Factor Analysis on this data by carrying out the following steps:

1) How many factors should be retained (based on eigenvalues and a Scree plot)? You may use any "Factoring method" or "Rotation method" you wish. The defaults, of course, are fine.

2) Provide the rotated factor loadings matrix. You should notice that (as compared with Principal Component Analysis) fewer variables load onto each factor. Interpret (as best as possible) your factors. Can you give an overall name to each factor?

3) Plot factor scores for your first two factors. Color the points by Brand. Are there any natural groupings of observations?

Part II. Consider the demographics data (Demographics.jmp or Demographics.txt). Each of the 440 rows contains demographic information on particular localities in the United States. The following is a brief description of the variables.

1) Perform a linear discriminant analysis using the highlighted variables for discriminating the variable Pop_Size_Group.

Pct_Age18_to_34 % = (# aged 18-34) / (population) Pct_65_or_over % = (# aged 65 & up) / (population) Num_physicians Active Nonfederal Physicians Num_hospital_beds Hospital Beds, Cribs, Bassinets Num_serious_crimes Serious Crimes Pct_High_Sch_grads % = (# 12 yrs school) / (# 25 & over) Pct_Bachelors % = (# bachelor's) / (# 25 & over) Pct_below_poverty % = (# < poverty) / (population) Pct_unemployed % = (# unemployed) / (population) Per_cap_1990income Per Capita Income Total_personal_income Total Personal Income

2) Show a plot of the two discriminant functions with points colored by Pop_Size_Group. Do both discriminant functions appear necessary to describe group separation?

3) Give the scoring coefficients for the two discriminant functions. Which variables appear most important for describing group separation?

4) Provide a classification matrix and the percent of observations misclassified.