BusMgmt 744, Project 2

Kelly Markor

03/16/2020

install.packages(“tidyverse”) library (tidyverse) install.packages(“ggplot2”) library(ggplot2) install.packages(“readxl”) library(readxl) install.packages(“dplyr”) library(dplyr)

### 1. Data Preparation (20pts)

### Randomly select a subset of 2000 observations from the data file using the R dplyr “${\tt sample\\_n}$” command. sample\_n(tbl, size, replace = FALSE, weight = NULL, .env = NULL, …)

Profit\_by\_Product <- read\_excel(“Profit by Product.xlsx”) View(Profit\_by\_Product)  
Profit\_by\_Product <- Product %>% sample\_n(Profit\_by\_Product, 2000, replace = FALSE, weight = NULL, .env = NULL, …)

### Create R-friendly variable names using the “${\tt clean\\_names}$” function found in the “${\tt janitor}$” package.

install.packages(“janitor”) Profit\_by\_Product <- as.data.frame(matrix(ncol = 6)) Profit\_by\_Product %>% clean\_names(dat, case = c(snake)

### Create a new variable “${\tt log\\_rev}$” by taking the log of “${\tt revenue}$”.

${\tt log\\_rev}$ <- log(${\tt revenue}$)

### 2. Numeric Description (30pts)

##For variables ${\tt revenue}$, “${\tt profit}$” and ``${\tt product\\_cost}$’’ desribe each variable by finding: Profit\_by\_Product <- read\_excel(“Profit by Product.xlsx”) ##Mean avg\_revenue <- mean(“Revenue”) avg\_profit <- mean(“Profit”) avg\_product <- mean(“product\_cost”) ##Mode(s) mode\_revenue <- mode(“Revenue”) mode\_profit <- mode(“Profit”) mode\_productCost <- mode(“product\_cost”)

##Variance varRevenue = Profit\_by\_ProductProfit  
var(varProfit)  
varProduct\_cost = Profit\_by\_Product$Product\_cost  
var(varProduct\_cost)

##Sample Standard Deviation varRevenue = Profit\_by\_ProductProfit  
sd(varProfit)  
varProduct\_cost = Profit\_by\_Product$Product\_cost  
sd(varProduct\_cost)

##5 Number Summary varRevenue = Profit\_by\_ProductProfit  
fivenum(varProfit)  
varProduct\_cost = Profit\_by\_Product$Product\_cost  
fivenum(varProduct\_cost)

### 3. Visualizing the data (50pts)

### Draw a histrogram for “${\tt customer\\_service\\_cost}$” and describe the plot.

ggplot(data=Profit\_by\_Product, aes(${\tt customer\\_service\\_cost}$)) + geom\_histogram()

### The graph is pretty standard and shows that the customer service cost is within a range of values.

###Create a scatter plot of “${\tt revenue}$” vs. “${\tt profit}$”. What relationship, if any, do you see? ggplot(data=Profit\_by\_Product, aes(${\tt profit}$)) + geom\_point()

### There is a direct correlation between revenue and profit. As one increases, the other increases also.

###Create box-whisker plots of “${\tt log\\_rev}$” by “${\tt quarter}$”. Describe the realtionships of ${\tt log\\_rev}$ within and between quarters. ggplot(data=Profit\_by\_Product, mapping=aes(x=${\tt log\\_rev}$, y=${\tt quarter}$)) + geom\_boxplot()

### Ther relationship of the revenue within and between quarters remains fairly consistent.