Noah Tesfaye (6279877)

ASSIGNMENT 3 PART II

Note: Your submission will consist of two steps. First, use the answer-area to provide a statement to the TA alerting them that a PDF document containing your answers to this assignment's Part II questions has been submitted (via document sharing discussed next). Second, use the shared-documents tool within MyStatLab to upload your PDF document. Do not forget to include a statement of academic integrity within the statement that you provide in the answer area to part II. Finally, note that you are required to show your work for full credit --- correct numeric answers may earn you little credit unless you show your work.

QUESTION PREFACE[25 marks]: You will be examining the data-set labelled ***ch05_Car_discounts*** **(see doc-depot)**, which has tracked automobile purchase activity, keeping track of some information about the purchaser (Sex, Income, and Age) as well as the *Discount* (advertised price less the final sale price) that the purchaser received upon their purchase. Please answer the following questions, by taking advantage (in part) of a statistics package such as Minitab or R or etc (see page 72, 117-119 and 183-185 for some brief guidance for producing graphical displays). Bonus 10 marks, for following the Plan/Do/Report style when producing your solution --- note that the Guided examples which appear in every chapter provide illustrations of how to apply to Plan/Do/Report style.

1. [4 marks] Using an appropriate display examine the distribution of Sex (male=0's and female =1s) among …

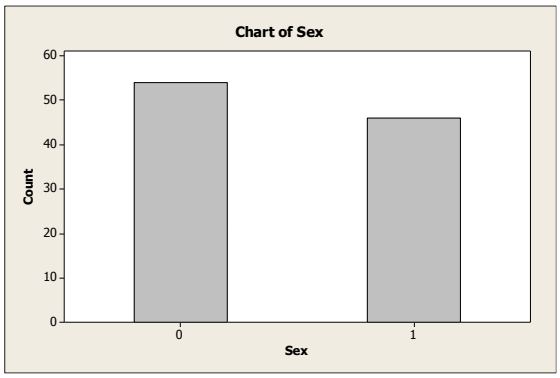
    a) … all purchasers

    Plan

    - We want to find the number of all female and male purchasers. Since we have a raw data we would want use a graphical display to show our work.
    - The data are categorical variables and we are interested how they are distributed, so will make a bar chart.

    Do

    - We can make a bar chart using Minitab or excel to illustrate the number of male and female purchers.



Report

- Our report displays the distribution of sex (Female=1 and male=0) among all purchases. To be specific we can use the count if function in excel to check our answer, which consists of 46 female purchasers and 54 male purchasers as displayed in our barchart.

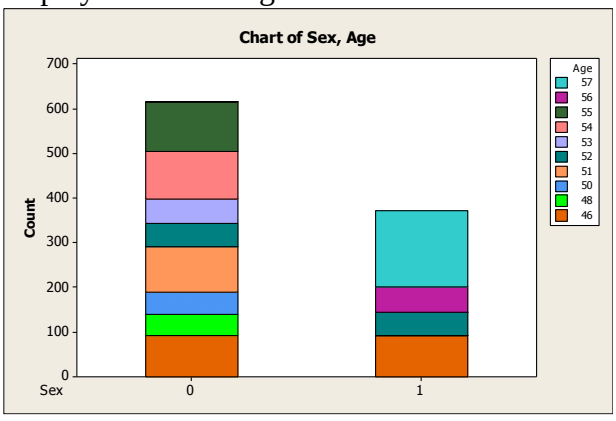
b)… purchasers 45 years of age or older

Plan

- We want to find the number of female and male purchasers 45 years of age or older. Since we have a raw data we would want use a graphical display to show our work.

Do

- We can make a stacked bar chart using Minitab to show the distribution of male and female purchaser age of 45 and above.
- We set the categorical variables to sex and age, then we set the data option to exclude value(age) below 45 for our bar chart to display the following:



Report

- Our bar chart to show the distribution of sex (male=0's and female=1's) that purchasers of 45 years and above years old. Our legend shows number (count) of each age number of purchasers. We

can also see that the distribution of the required age among male and female purchasers. Ages 46 and 55 have increased distribution among male purchasers. Ages 57 and 46 have increased distribution among female purchasers.

2. [6 marks] Examine the distribution of the variable *Discount.*

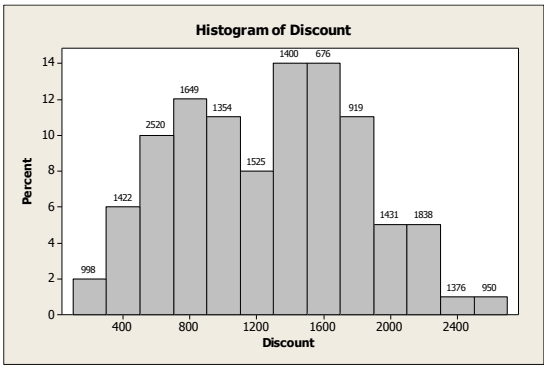
a) Using an appropriate graphical display, being sure to label carefully, and being sure to describe its shape.

Plan

- We want to find the distribution of the variable *Discount*. Since discounts are in any currency, we can conclude that it's a quantitative variable. So we'll use a histogram and describe its shape.

Do

- We will make histogram using minitab by specifying our graph variable to discount.



Report

- The histogram shows us the distribution of the variable discount over percent (y-axis). The vertical axis scale isn't clear in this case but has no effect on the shape of the histogram. The currency of the discount is not specific so the graph does not contain any. The shape our histogram is unimodal and skewed to the right.

b) Summarize the distribution quantitatively by providing relevant descriptive statistics.

Mean 1266.8
Standard Deviation 537.4
Min 131.0
Q1 817.8
Median 1327.5
Q3 1673.5
Maximum 2520.0
Upper fence 2957.05
Lower fence -467.75
IQR 855.5

3. [6 marks] Re-examine question (2) by comparing the distribution of *Discount* secured by females *versus* that secured by males. Use a graphical display most appropriate for comparison.

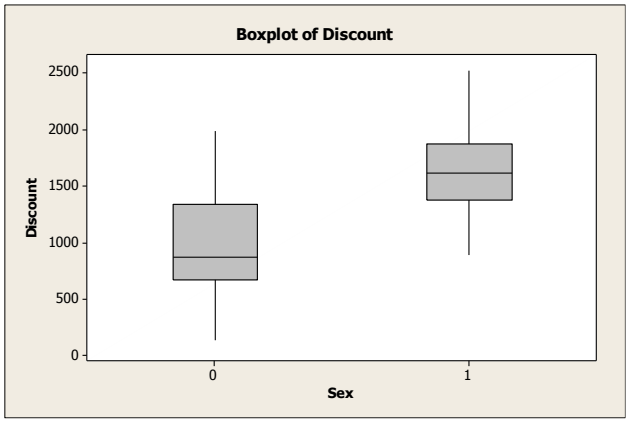
a) comment on the result
b) Provide descriptive statistics to support your comments

Plan

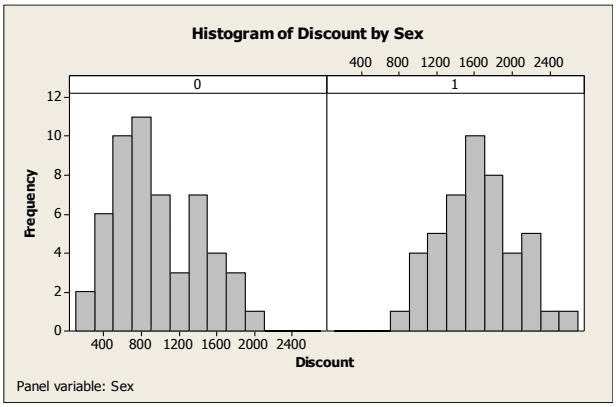
- We are asked to graphical show the comparison between male and female distribution of discount. So, we should have two histograms to show the distribution and comment on the results.

Do

- We will use Minitab to display the comparison of discount among male and female purchasers using histogram or boxplot in which case sex is the variable.



-



Variable	Sex	N	Percent	Mean	SE Mean	StDev	Variance	CoefVar
Discount	0	54	54	962.1	62.3	457.7	209527.3	47.58
	1	46	46	1624.6	56.4	382.4	146197.7	23.54

Variable	Sex	Minimum	Q1	Median	Q3	Maximum	Range	IQR
Discount	0	131.0	666.8	870.5	1336.0	1990.0	1859.0	669.3
	1	892.0	1370.5	1614.5	1868.5	2520.0	1628.0	498.0

Male (Boxplot chart)

Upper fence= Q3+1.5IQR=1336.0+1.5(669.3) =2339.95

Lower fence=Q1-1.5IQR=666.8-1.5(669.3) =-337.15

Female (Boxplot chart)

Upper fence= Q3+1.5IQR=1868.5+1.5(498.0) =2615.5

Lower fence= Q1-1.5IQR=1370.5-1.5(498.0) =623

By calculating the upper fence and lower fence clearly see that there are no outlier (value that fall above or below the fences)

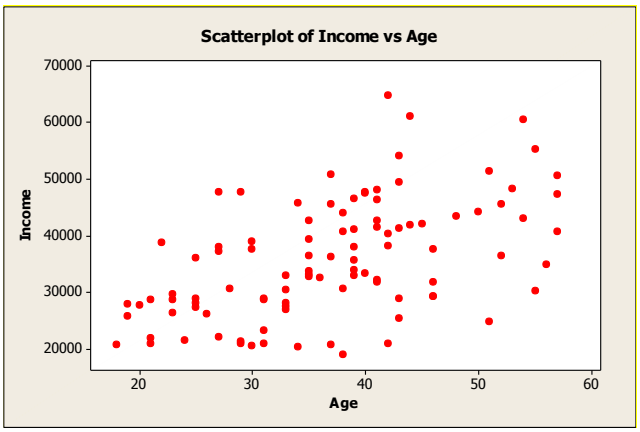
Report

- Both the boxplot and histogram show the same results. Females (1's) tend to get more discount than mean as they mean is higher than that of male (0's).
- Female (0's) histogram is unimodal and symmetric.
- Male (0's) histogram is unimodal and skewed to right.

4. [5 marks] Examine the relationship between *Age* and *Income* using an appropriate graphical display. Provide appropriate quantitative summaries. Discuss and interpret the result (does it make sense). Recommend how one might examine whether the relationship differs between males and females.

Plan

- Check the conditions.
  - ✓ Quantitative variables: Income is measured in currency and age is measured in years.
  - ✓ Linearity (straight enough) condition: The scatterplot is moderately straight (not fully).
  - ✓ Outlier condition: There are no obvious outliers.



Do

- The conditions for the scatterplot have been satisfied. So can calculate the correlation from the graph
- The correlation is 0.525.
- The positive correlation coefficient confirms the direction of our scatterplot.

Report

- It possible that age and income have moderate correlation as when customers (people) grow older they tend to earn more money might be due to seniority or work experience.

5. [4 marks] Produce a correlation matrix, including all variables except *Sex*. Interpret the correlation matrix. What would happen if you included *Sex* as one of the variables in your correlation matrix computation?

Plan

- Correlation measures the strength of the linear association between the two variables.

We asked to produce the correlation matrix, excluding sex. Analyze the correlation matrix and discuss the impact of sex inclusion in the correlation matrix.

Do

- We will use Minitab and go under stats then select correlation under basic variable to find the correlation matrix.

Correlations: Discount, Age, Income

	Discount	Age
Age	-0.265	
Income	-0.727	0.525

Cell Contents: Pearson correlation

- The correlation coefficient gives the direction of association. In this correlation matrix, the correlation between age and discount has is –0.265, which is a moderately strong and indicates substantial variation.
- The correlation between income and discount has is –0.727, which is a very strong variation among the two.
- The correlation between income and age has is 0.525, which is a moderate correlation among the two.
- When we include sex as one of the variables in our correlation matrix, we understand that sex and discount have a positive stronger correlation (0.618) than in comparison to discount and income or age. Sex and age have a better correlation than sex and income.

Report

- We have seen the various correlation coefficients of our data. We can speculate what might be the cause of correlation or variation among the variables. It possible that age and income have moderate correlation as when customers (people) grow older they tend to earn more money might be due to seniority or work experience. Income and discount have a strong variation might be mainly because marketing campaigns in the organization know their customer earnings and spending patterns in order to offer them a discount. Sex and discount have moderately strong relation might be because the items that is being discount is highly dependent upon the purchaser's sex.