Helena Devins
6300497

Assignment 2 Part II

Note: Your submission will consist of two steps. First, use the answer-box 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 (document sharing). 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. [DUE DATE March 15th]

PLAN: Throughout this assignment I plan to analyze the data of a car sales person looking into the statistics around sex, age, income, and discounts. By manipulating this information in different charts, graphs and tables there should be evidence of different segments having an advantage in the purchasing of a car. It should be predicted that one sex over the other, customers of different ages, or customers with different incomes will be able to command different pricing. Through the creation of bar graphs, histograms, box-plots, scatter plots, and a matrix correlation there will be visual evidence of negative and positive associations as well as the frequency of customers and current discounts.

DO:

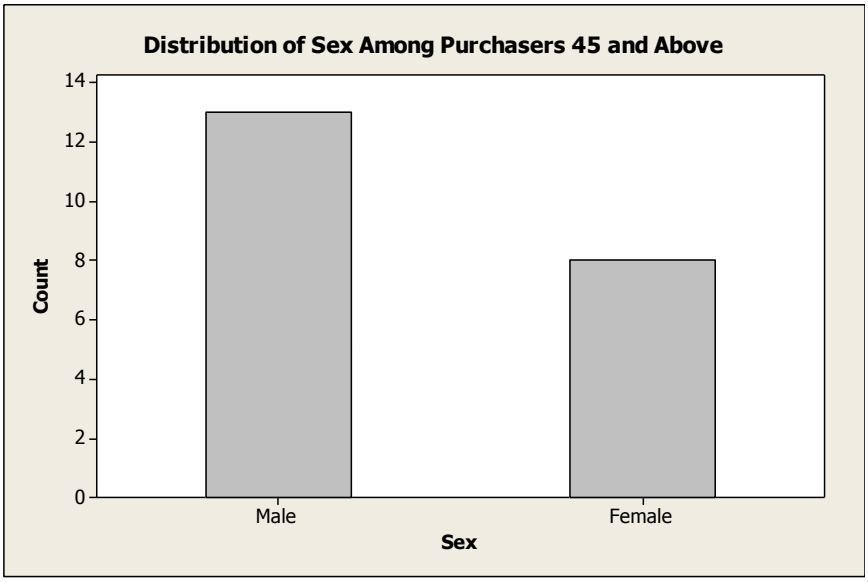
QUESTION PREFACE[25 marks]: You will be examining the data-set labelled ***ch05_Car_discounts*** **(see doc-depot)**, which has tracked automobile purchase activity, keeping track of some information about the purchaser (Sex, Income, and Age) as well as the *Discount* (advertised price less the final 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

Examining the distribution of purchasers of cars it is evident that there are more males than females purchasing cars at this time.

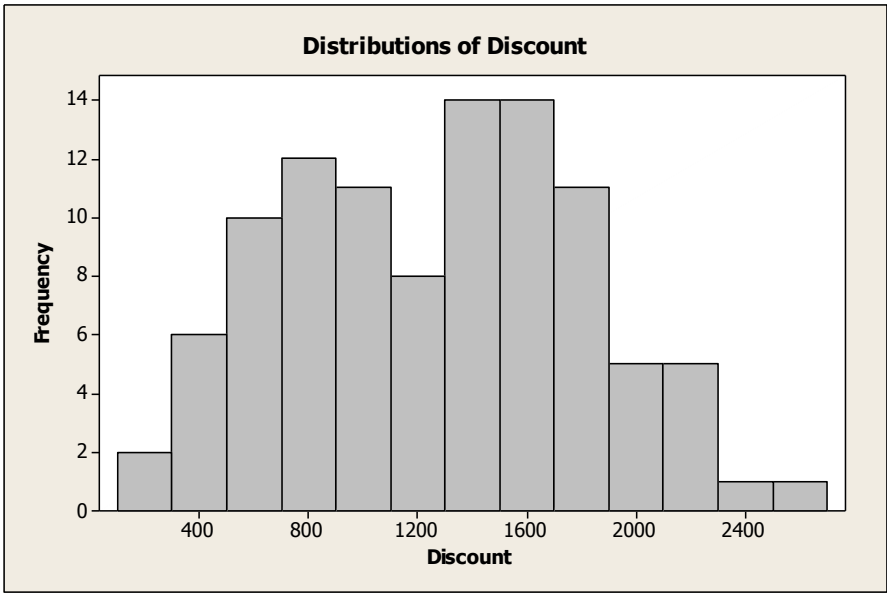
b) … purchasers 45 years of age or older



When examining the distribution or purchasers 45 and over there are about two-thirds more male purchasers in this age bracket.

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.

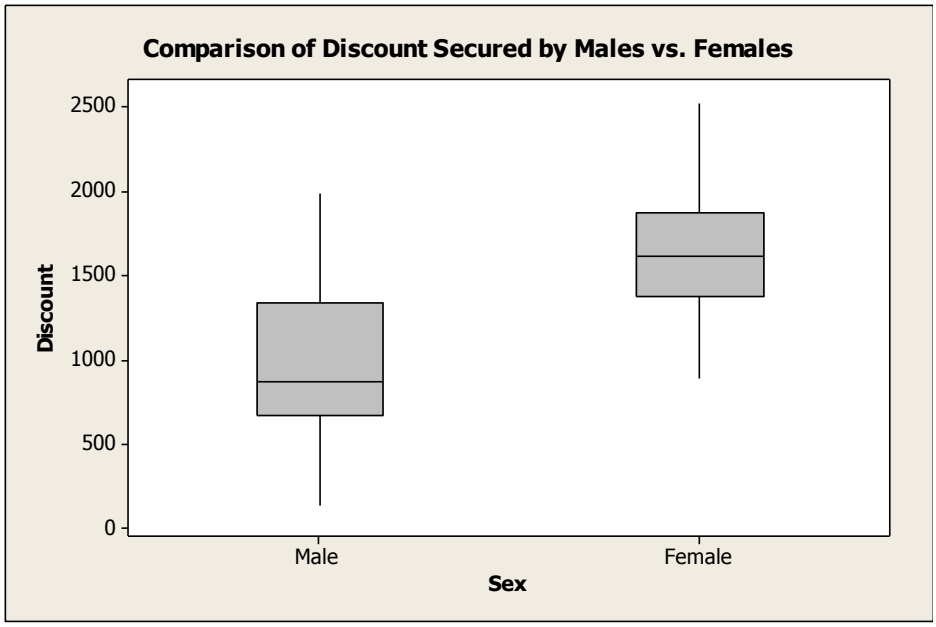


This graph describes discounts among all purchasers; looking at the amount of discount and the frequency of each. This graph is relatively symmetrical with a bimodal pattern; this can be concluded by seeing that there are two prominent humps (also known as modes) in the chart and if you were to fold the graph in half it is relatively the same pattern. The spread of this graph is 2,389.

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

Variable	Mean	St.Dev.	Min.	Q1	Median	Q3	Max	IQR
Discount	1266.8	537.4	131.0	817.8	1327.5	1673.5	2520.0	855.7

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 results

The median for females is a higher value in this plot. In comparison, the IQR for males is larger distribution however the female value is higher.

b) Provide descriptive statistics to support your comments

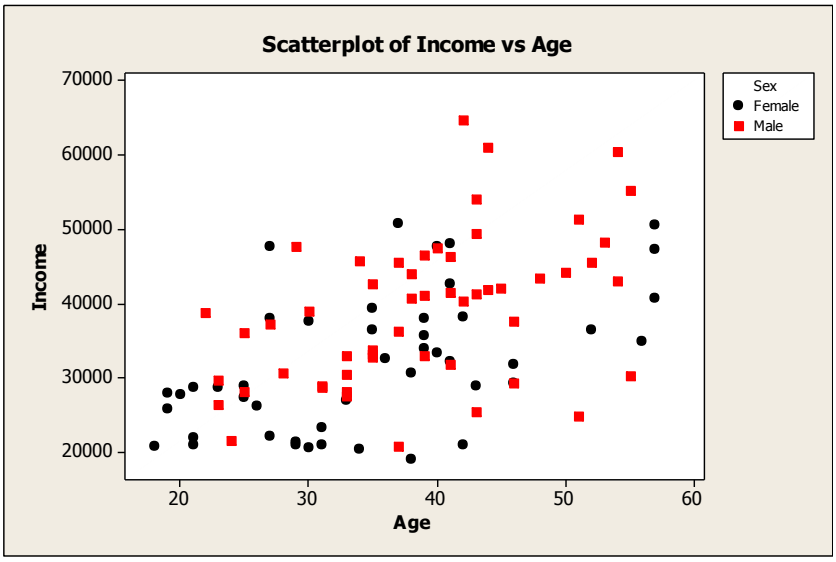
Variables	Males - Discount	Females - Discount
Mean	962.1	1624.6
St.Dev.	457.7	382.4
Min	131.0	892.4
Q1	666.8	1370.5
Median	870.5	1614.5
Q3	1336.0	1868.5
Max	1990.0	2520.0
IQR	669.2	498

Helena Devins
6300497

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.

This graph makes sense as it is a weak positive relationship between increasing income as the purchasers get older. It can be concluded that it is positive as it runs from bottom right to top left in a vaguely linear pattern. It can be assumed that as a person gets older they have a better paying job. The correlation between age and income is 0.525.

It would be recommended that to examine the relationship between age and income, differing between males and females, you can create a scatterplot grouping males versus female. The graph displayed below is an example of the graph suggested. It would be specified that if you are going to find the correlation of this graph you must look at the correlation still between income and age in terms of male and in terms of female. It is not proper to do correlation of sex and income or sex and age as it is trying to do correlation with a quantitative value and categorical value.



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?

Matrix Correlation			
	Age	Discount	Income
Age	1.00000	-0.26531	0.52521
Discount	-0.26531	1.00000	-0.72749
Income	0.52521	-0.72749	1.00000

Interpreting the table above you can conclude that there is a weak negative linear association between discount and age. On the other hand there is a strong negative linear association between income and discount. However, with age and income there is a moderate positive linear association. When data is correlating with itself it produces perfect correlation of one.

If you were to include sex as one of the variables in your correlation matrix this would result in incorrect results as it is not proper to do correlation between quantitative data and categorical data. This correlation is found as male and female were given numerical values making them seem like quantitative data which is not the true case.

Matrix Correlation				
	Age	Sex	Discount	Income
Age	1.000			
Sex	-1.198	1.000		
Discount	-0.265	0.618	1.000	
Income	0.525	-0.351	-0.72749	1.000

REPORT: By reflecting on the data, charts, tables, and graphs processed above you can conclude that there are more men purchasing cars within all purchasers and those customers age 45 and above. A car sales person distributes the most discounts at a mid-range around $1300 to $1700. With the least distribution around $2400 and only slightly more distributed at around $200. It is evident that women get higher discounts however males have a larger IQR. Very vaguely it is evident that there is a positive relationship between the age of a purchaser and how much income they earn; those old have a higher income. It can be assumed that an older purchaser is more distinguished in their career and therefore has a larger income. By creating a scatter-plot that presents males and females with different symbols it is also evident that male purchasers earn more income then females in the same age bracket. By producing a correlation matrix it can be concluded that there is a negative association between age and discount; as well the relationship between discount and income. However there is a positive association between age and income, as was suggested in the scatterplot. By discovering these two negative associations it can be concluded that as a purchaser gets older car sales people are less likely to give them a discount. Though this is a weak association it is in a car sales person's best interest to sell to an older purchaser at a price closer to the full value, as older purchasers should have more experience. To that end, those purchasers with a high income also have less chance at getting a discount as a car sales person can take advantage of these purchasers that have a larger amount of disposable income. The association between income and discount is a stronger negative association as car sales people would want to sell more expensive cars to these purchasers with higher incomes, as well as sell cars at a price closer to the full value. To conclude, more purchasers are males they have a higher income and get less of a discount. On the other hand, fewer purchasers are female and they end

up getting a larger discounts, or more discounts overall. You however cannot make a correlation matrix around sex associated to income, discount, and age because sex is categorical and the other three are quantitative. Categorical data should not yield a proper correlation as it is generally not numerical. The only reason you could attempting a creating a correlation with sex in this data set is because male with given a numerical ‘value’ of zero, and female was given a ‘value’ of one; however this was for creating counts of data not to be literally taken as values.