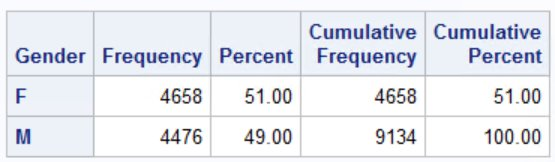
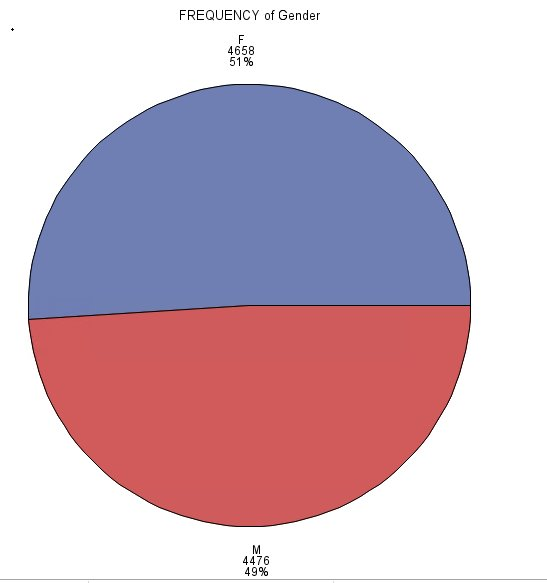
Homework 1

Read the insurance claims dataset.

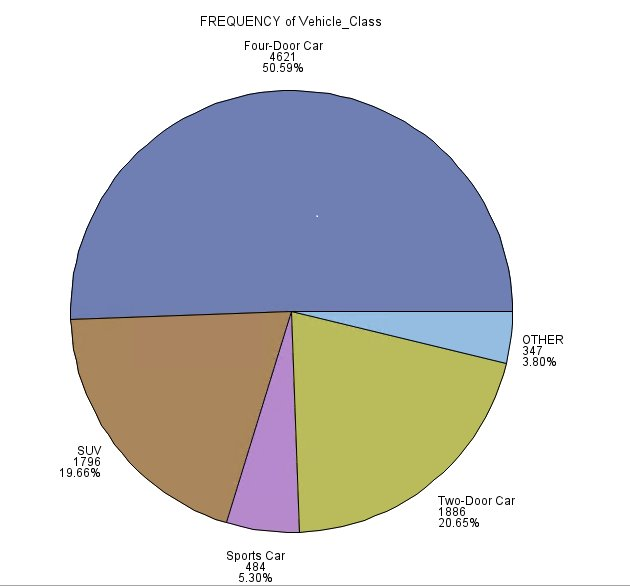
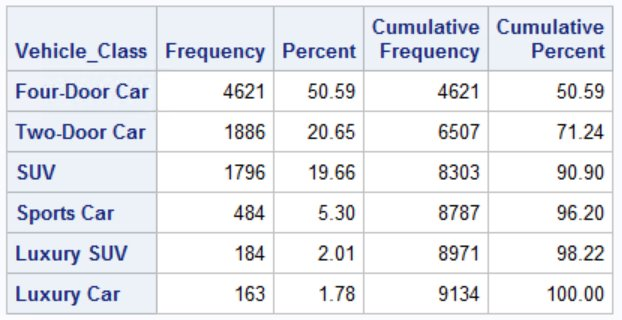
Using SAS answer the following questions and interpret the output in each question. In Other words, what do you learn from the output of each question.

1. What is the distribution of gender, vehicle size, and vehicle class?

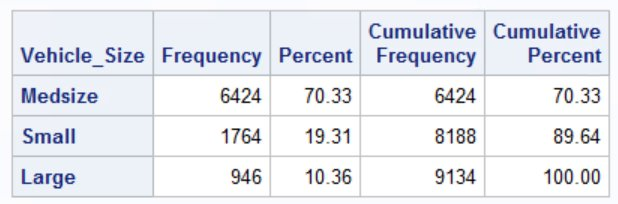
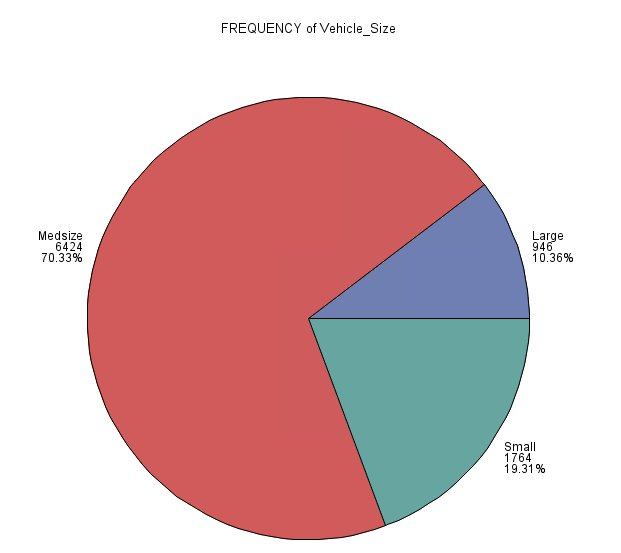
 

**Gender:** As per the frequency distribution, the insurance data is quite evenly distributed among males and females. Female’s share is 51% of the data and males are 49%. This shows female claims car insurance a little more than males.

**Vehicle Class:** As per the given car insurance data, the Four Door car has captured the half of thecustomer share.

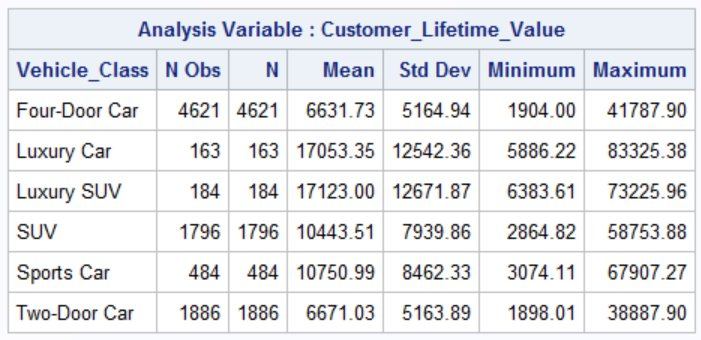


**Vehicle size:**

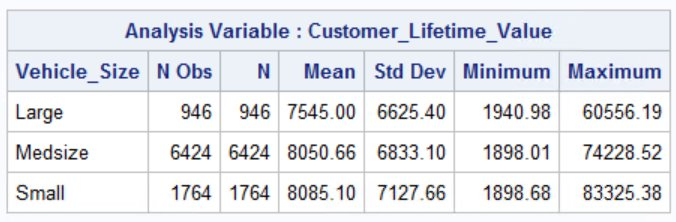
 

As per the given data, the major portion of market is occupied by mid-size car. The market share of mid-size, small and large size are 70.3%, 19.3% and 10.4% respectively.

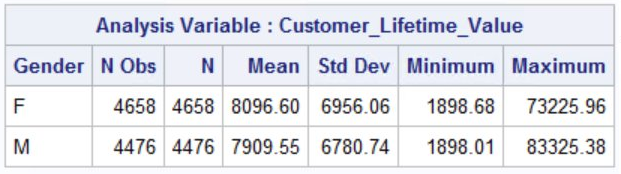
1. What is the average customer life time value of each level of gender, vehicle size, and vehicle class?



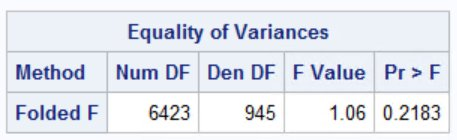
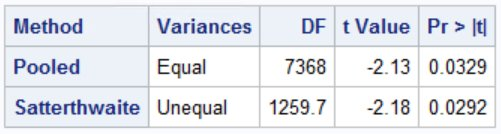
**Vehicle size-** the average customer life time value is higher for Luxury SUV and luxury car segment, which means that the customer purchasing insurance for these segments seems to be more profitable for the insurance company. But another factor to be considered here is standard deviation, this luxury car owner has highest variance as compared to other segments, which means that these customers are not quite risk prone as well. Whereas, the maximum car insurance customer are from four door car segments, which has lowest standard deviation or variability in their insurance values. Hence these customers are least risk prone. The same is evident about the two door car segment that they are also less risk prone.



Vehicle Size: From the above result it is evident that although there is quite deviation from average in lifetime value in all three vehicle size but medium and small size vehicle has almost same average customer life time value.



1. Do Large cars have a higher lifetime value than medsize cars. Do a ttest and report on your findings.

From the mean distribution table, we can see that the average customer lifetime value of large cars is not greater than midsize cars. To statistically test this, we will perform the ttest.

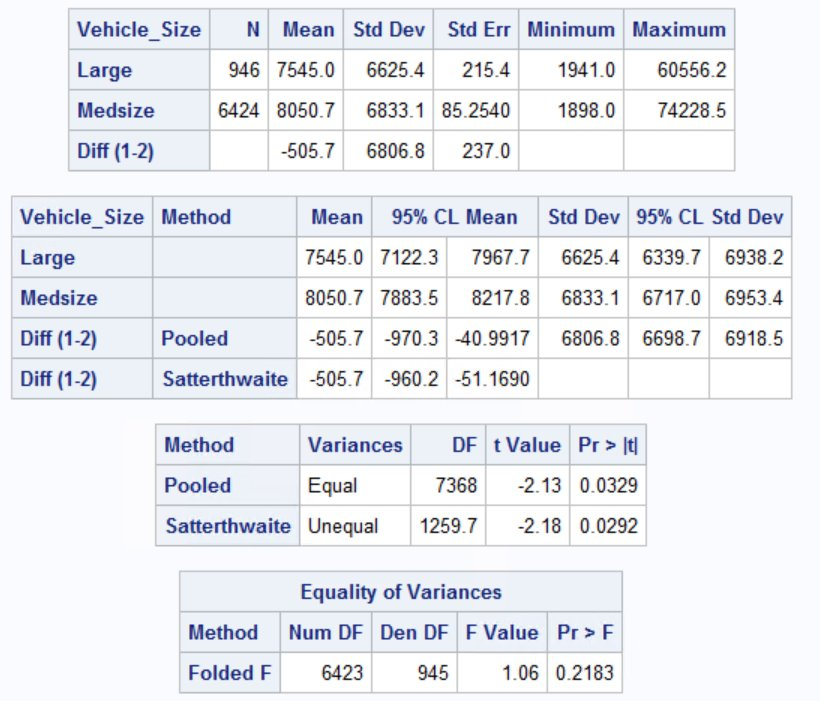
Test of Variance: In this test

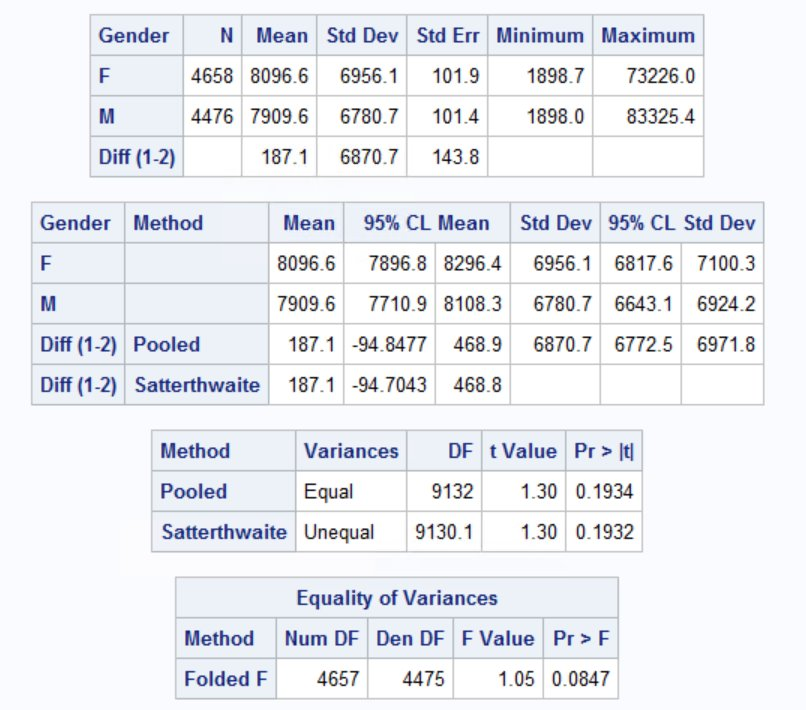
Null hypothesis (H0): variance of both the populations is same, that is the variance of lifetime value for large and medium size cars is equal. This test is important in the sense that it tells whether the lifetime value average difference is not due to noise in the data.

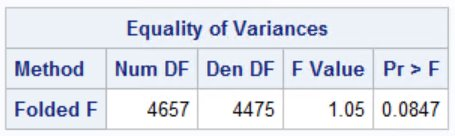
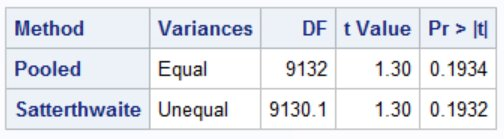
From the below result we can see that the p-value > 0.05 (alpha), hence we cannot reject the null. This means that the variance is same for lifetime value of large and medsize vehicles.

Test of Average: In this test, we assume (Null hypothesis) that the average value of large vehicles is less than or equal to medsize vehicles and alternate hypothesis is that large cars have higher lifetime value than medsize cars.

From the results we can observe that the p-value for equal variance is 0.03 which is less than the threshold value of 0.05, hence we reject the null hypothesis. Therefore, we can conclude that large cars have higher lifetime value than medsize cars.



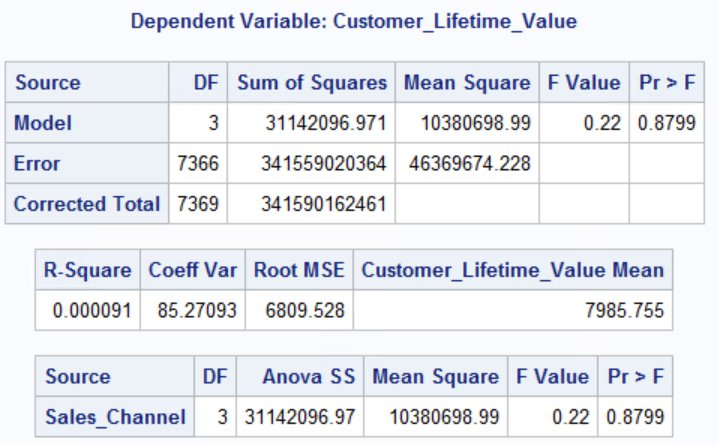
1. Is there a significant difference between men and women in customer life time value?
2. 

To test whether there is significance difference between lifetime value for male and female, we perform a ttest. From the variance test we can see that as the p-value > 0.05, we cannot reject the null hypothesis. Hence with 95% confidence we can say that male and female have equal variance.

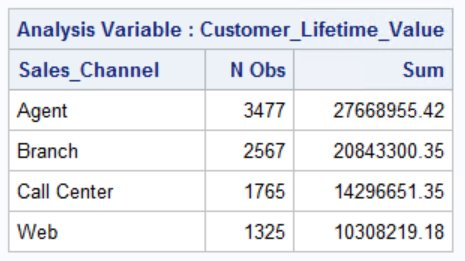
From the ttest with equal variance the p-value (0.1934) > 0.05, hence we cannot reject the null hypothesis. Therefore, the male and female customer average lifetime value difference is significant.

1. Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generates the highest lifetime value?

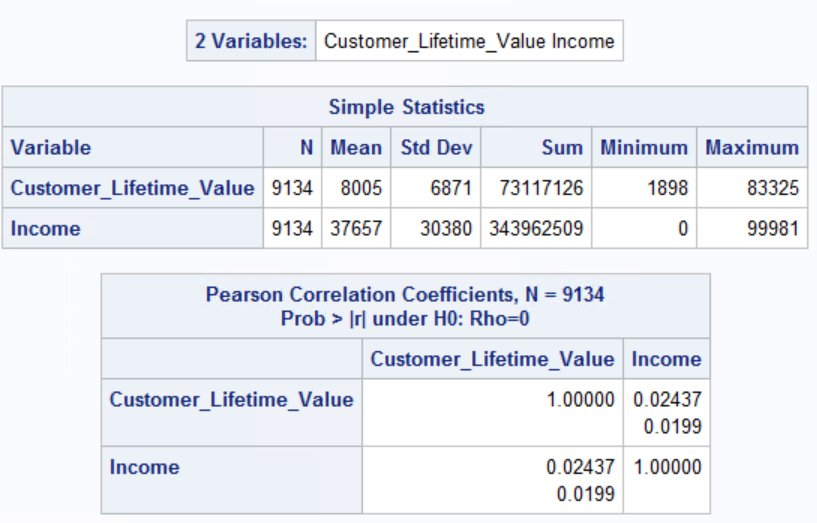


Since the p value >0.05, hence we reject the null hypothesis (i.e.H0- all means are same). Therefore, there is difference in customer life time value across different sales channel.

Among all sales channel, agent generates highest customer life time value.



1. What demographic factors (education, income, marital\_status) affect customer lifetime value?



Since the correlation between income and customer life time value is .02 i.e. (weak correlation) but p value < 0.05 that shows that 0.02 correlation factor is significant.

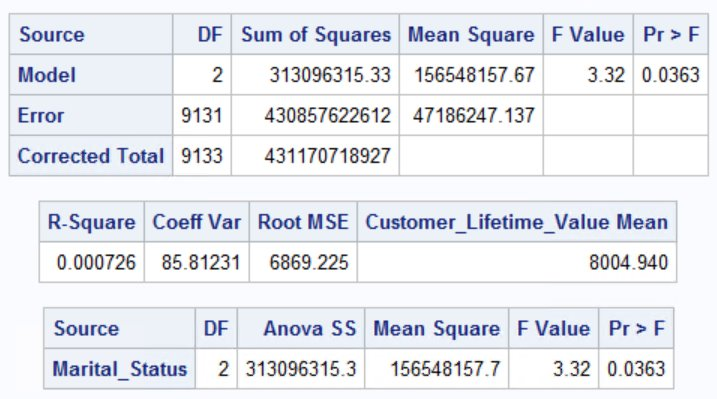


Figure marital

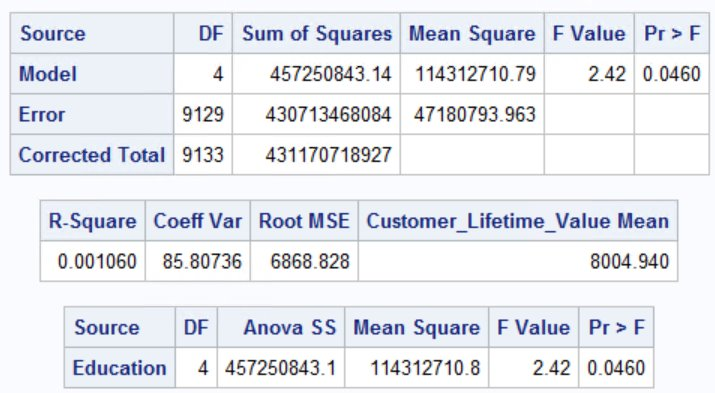
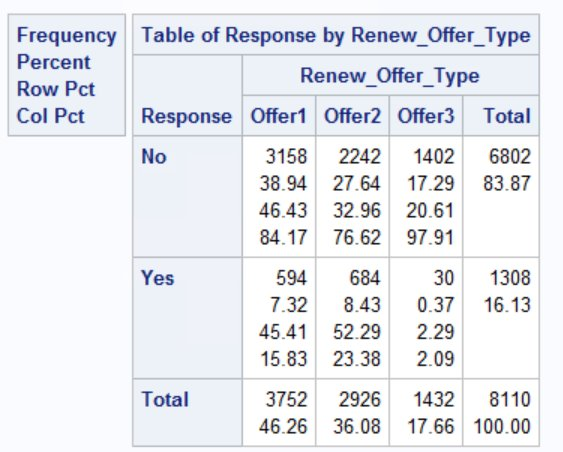
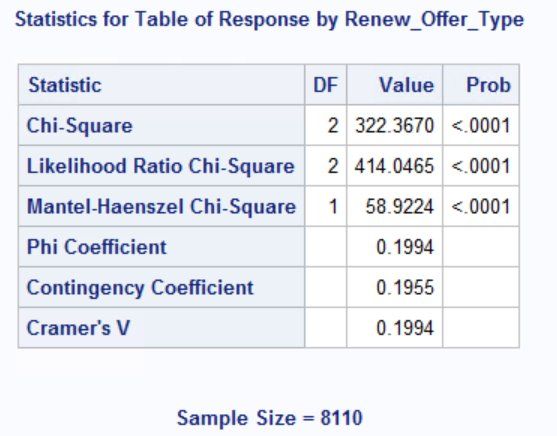


Figure education

1. Is there a relationship between renew\_offer\_type and response (use Chi-sq test)? **Which offer type generates the highest response rate**?

Since the p value< 0.05, we reject the null hypothesis (i.e H0:- the two group are independent) that conclude with 95% confidence that there is a relationship between renew offer type and response.

1. Do different renew\_offer\_types have different lifetime values? **Which offer type is the best**?

annova

1. Is the effectiveness of renew\_offer\_type different across different states with respect to lifetime value?
2. What other interesting insights that are useful to the company in terms of action can be obtained from the data? Write any 3 and indicate which type of analysis is appropriate.