BIA652D Multivariate Data Analysis

Examination of data (Technical Support and Complaint Resolution variables) using Univariate and GLM procedures in SAS

GROUP4 Homework 3

2016

This is a summary of the examination of Technical Support (X8) and Complaint Resolution (X9) variables. The focus is to see if these variables satisfy the assumptions of normality and homoscedasticity. Initially, each variable is examined individually through empirical measures and graphical representations to understand the shape of the distribution and to determine normality. Then the same analyses are done against the Customer Type (X1) variable. Customer Type has three values indicating the length of time a customer has been purchasing from HBAT - less than 1 year, between 1 and 5 years and longer than 5 years. Both the variables are examined for each of these groups to determine whether their variance across all three groups is homogeneous (i.e. they are homoscedastic). The sample size analyzed is 100 for each of the two variables.

The empirical measures of kurtosis and skewness for Technical Support (X8) are -0.55 and -0.20 respectively. Both these values are close to 0 suggesting X8 has normal distribution with no excess kurtosis or skewness. The normality test of Shapiro-Wilk gives p-value of 0.39. At 0.05 level of significance, the null hypothesis of normal distribution of X8 cannot be rejected. The normal probability plot, histogram and boxplot on page 6 show the shape of the distribution for X8. The data points coincide with the normality line in the normal probability plot and the mean and the median in the boxplot also coincide showing the shape of the distribution is fairly normal. The empirical measures of kurtosis and skewness for Complaint Resolution(X9) are -0.59 and -0.14 respectively suggesting X9 is also normally distributed. The p-value from the Shapiro-Wilk normality test is 0.40. Same conclusion as for X8 can be drawn here too at 0.05 level of significance. The normal probability plot, histogram and boxplot for X9 on page 9 show the shape of its distribution is quite normal.

The variance of Technical Support (X8) for each of the three groups of customer type can be seen in the visual representation on page 19 with a boxplot of X8 values for each customer type group. The height of the box for second and third group of customers is almost equal, the height of the box for first group (less than 1 year) is slightly bigger indicating the variance of X8 in first group is slightly more compared to other two groups. Overall though the variance of X8 across all three groups is homogenous. The empirical tests for homogeneity of variance Levene, Brown and Forsythe and Bartlett all have p-values greater than 0.05 (0.48, 0.72 and 0.62 respectively) indicating the null hypothesis of homoscedasticity cannot be rejected at 0.05 level of significance. The coefficient of variation measures for the three customer type groups are close, 32.9, 27.92 and 25.16, again suggesting similar variance across all three groups. The value of 32.9 for first group of customers gives the same information as the boxplot visual that there is slightly more variance across first group as compared to second and third groups.

The boxplot visual representation for Complaint Resolution (X9) across the three customer type groups on page 29 shows homogeneity of variance of X9 across the three groups. Similar observation as for X8 can be made here that the variance for first group of customers appears to be slightly more than that for second and third groups. The empirical tests for homogeneity of variance Levene, Brown and Forsythe and Bartlett all have p-values greater than 0.05 (0.6, 0.66 and 0.65 respectively) meaning the null hypothesis of homoscedasticity stands at 0.05 level of significance. Looking at the coefficient of variation measures for the three customer type groups we can draw the same conclusion - 21.45, 14.94 and 17.43. Although X9 for first group of customers has slightly more variance as shown by the boxplots representation too, overall the measures are quite close indicating homogeneity of variance of X9 across the three groups.