**Slides [11 -12]**

These are the variable which are in this data set - tenure months, monthly charges, avg monthly long distance , total extra , avg monthly download .

Here we have created two table for churn and non-churn containing min q1 median man q3 and max.

One important point to be noted from these tables is for the monthly charges, avg monthly long distance charges avg monthly gb downloaded the mean and median are almost equal or close in both churn and non-churn group.

From this we can infer that the graphs for these variables are symmetric.

Also, from these tables, we can observe that the mean and median for tenure and average monthly charges are different for churn and non-churn group. From this we can infer that the factors tenure and average monthly charges has a weightage over the decision of customers in deciding whether to churn or not.

The point here to be noted here that, these are some few bits of information including mean, median, max, min, and we cannot solely infer conclusion on the basis on this few information alone, but as an initial step we can take these inferences and the subsequent statistical analysis and test are explained in the upcoming slides.

**Dependants:**

From the graph it is evident that customers who has dependants has a low churn rate that those with a dependant, and this is also confirmed by Pearson’s chi square test, giving a low p value, rejecting the null hypothesis, and accepting the alternative hypothesis that there is a relationship between variable dependants and churn label.

**Phone service:**  It is evident from the graph that, equal number of customers chose to churn and not churn, giving no weightage to phone service label in churning or not churning, this is also confirmed by Pearson’s chi square test, giving a high p value greater that 0.05 , and we accepts the null hypothesis that there is no relation between the factor phone service and churn label

**Unlimited data :**

Evident from the graph that customers who an unlimited data has a high churn rate that those who has not.

Pearson’s chi square test on gives a low p value, rejecting the null hypothesis, and taking the alternative hypothesis that there is a relation between phone service and churn label.

**Avg Monthly Gb downloaded:**

From the histogram, for the churn and non-churn group , the distribution looks different, and the means are different, confirming it’s weightage on whether to churn or not. From the qq plot, we test the normality test, we can see that the distribution is not normal for both churn and non-churn group .

We ran a two sample t test on avg monthly download variable, returning a very low p value, rejecting the null hypothesis and accepting the alternative hypothesis that avg monthly gb downloaded effected the customers to churn or not.

**Total extra data charges:**

The histogram is skewed and means are similar. From the test of normality, from the qq plot, we can see that the distribution is not normal.

We ran a two sample t test on these variable and returning a high p value , and accepting the null hypothesis. Total extra data charges ahs no effects on whether customer to churn or not.

**Internet Service , Test of independence :**

Internet services has 3 levels – DSL,Fiber optic and no internet service.

Customers who has fiber optic churned than the other two, evident form the graph.

Since it has 3 levels, we ran a test of independence, created a contingency table of Churn against to internet service and to test our null hypothesis.

The test returned a low p value, rejecting the null hypothesis , Internet service effected whether a customer to churn or not.

**Conclusion :**

I pick dependant variable as a example in building the model

The customers who have a dependency has a very low churn rate than customers who have no dependency.

The low p value from the chi sq test also confirm variable dependants has an weightage on deciding people to stay or leave

Apart from the statistical analysis, this can be explained as :

1. If a customer who has dependency need to leave , everyone in the dependency should have to leave all together ,which is more complicated than customer who has no dependants , and who can easily leave the service.

2. Also the service provider may be providing a better plan for customer who has dependants, than their other service competitors.