In this section we will take a look at how we do a one-way ANOVA tests and two-way ANOVA tests in side R. We will also take a look at Chi square test of goodness of fit and as well as Chi square test of factor associations.

Let's first of all discuss how we will do one-way analysis of variance or one-way ANOVA. Now to do one-way ANOVA let's first of all create a dummy data set. This is the dummy data set that we created to do one-way ANOVA. The data set is about the miles per gallon of different brands of cars. What we want to do now is we want to figure out if there is any difference between average miles per gallon across brands.

To do one-way ANOVA, we use aov command. Now the miles per gallon variables are dependent variable. So the first parameter would be the formula and the first part of the formula would be miles per gallon which will be a dependent variable. Then we put the symbol ~ and write our independent variable which in this case is the brand variable. We will also specify that both these columns are coming from the data called mileage.

Let's execute this command and store the results of analysis of variance procedure in an object called ANOVA. Let's not take a look at the summary of this particular ANOVA object. As you can see the p value is very, very low. So we can say that our ANOVA results are signifying implying that mean mile per gallon differs across all these brands.

Let's now take a look at how we do a two-way ANOVA in side R. For this we will read a text file that I have saved on my system. Let's read this text file and stored it in an object called TW. Let's first of all take a look at the contents of this file and try to understand the data that we have.

Now this data is about the satisfaction level of people depending upon what format of education they are follow. Do they follow an online format of education? Do they follow a hybrid format of education? Or do they follow some other format? Also we have information about what subjects a particular person is enrolled in. Is that person enrolled in a subject like statistics, English, and science etc.

In order to do a two-way ANOVA we will again use the aov function inside R since satisfaction is a dependent variable. So it will form the first part of the formula. So we will right
(Satisfaction~Format+Subject+Format:
Subject,data=tw)

When we do a two-way ANOVA we also talk about the interaction between the two independent variables. So this term here Format: Subject, specifies the interaction term. Then we tell R that the data is coming from an object tw.

Let's execute this command. Now let's take a look at the p values by doing a summary on this model object. As you can see the format factor is turning out to be significant, the subject factor is turning

out to be significant but the interaction term not turning out to be significant.

Let's now discuss the Chi square test of goodness of fit. Now in order to illustrate Chi square test of goodness of fit lets create a dummy data. Let's take a look at the data that we have created in the object dat. This object dat contains data bout the blood type and the expected population prevalence of blood type, the observed number of people with certain blood type as well as the expected number of people with specific blood types based on the population prevalence of that blood type.

In order to do a Chi square test of goodness of fit in R, we need to specify the observed values as well as the expected probabilities are. In this case this particular column pct which specifies the population prevalence rate. This is little different from how we implement Chi square in tools like excel. Because in tools like excel, we supply an observed and expected values but in R what we are doing is we are supplying a vector of observations and another vector of probabilities.

Let's execute this command. As you can see the p value is not less than 5%. In fact it is a very high number, which is 76%. So we cannot reject our null hypothesis here.

Let's take a look at how we do Chi square test of factor independence. First of all let's create a data set. This is the data set we will use to do a Chi square test of factor independence. This data

set essentially talks about the distribution of people based on age and how frequently they consume alcohol.

For example, this cell quantity of 90 represent in our sample, we have observed 90 people were under 45 and who occasionally drink. The question we want an answer to here is the age and the frequency at with someone drinks, are these factors related? So we will do a Chi square test.

The syntax to do a Chi square test is to use this function chisq.test and supply the table to this particular function. Once we execute this command we see that our p value is very, very low implying that there is a dependence between this factor age and the frequency at which someone drink.

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

In this section we have discussed how we do a one-way ANOVA and a two-way ANOVA analysis in site R. We have also discussed to non-parametric tests in the form of Chi square test of goodness of fit and Chi square test of factor independence.

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>