Categorical variables

Suppose we have the file *diamond.xlsx* containing 10 variables. The categorical variables are: *Cut, Color, Clarity*.
Consider the variable *Cut* with categories are in column B in the cells B2 to B341.

To display the absolute frequencies of the categorical variable *Cut* and identify the mode, select the data, then
INSERT ➔ RECOMMENDED PIVOT TABLES ➔ COUNT
INSERISCI ➔ TABELLE PIVOT CONSIGLIATE ➔ CONTEGGIO

To draw a bar plot of the categorical variable *Cut*, select the table, then
INSERT ➔ GRAPHS ➔ ALL GRAPHS ➔ CLUSTERED COLUMN
INSERISCI ➔ GRAFICI ➔ TUTTI I GRAFICI ➔ COLONNE

To draw the pie chart of the variable *Cut*
INSERT ➔ GRAPHS ➔ ALL GRAPHS ➔ PIE
INSERISCI ➔ GRAFICI ➔ TUTTI I GRAFICI ➔ TORTA

Import the dataset in the file *diamond.xls* and define the variables *cut* and *color*.

```
cut=diamond$Cut
color=diamond$Color
```

To display the categories of the categorical variable *cut*

```
cut=as.factor(cut)
levels(cut)
```

To display the absolute frequencies of the categorical variable *cut* and identify the mode

```
table(cut)
```

To draw the bar plot of the categorical variable *cut*

```
plot(cut)
```

To display the percentages of the categorical variable *cut*

```
100*table(cut)/length(cut)
```

or

```
100*prop.table(table(cut))
```

To draw the pie chart of the variable *cut*

```
pie(table(cut))
```

To display the categories of the categorical variable *cut*

```
color=as.factor(color)
levels(color)
```

To display the joint frequency table of the categorical variables *cut* and *color*

```
table(cut,color)
```

To draw the bar plot of the categorical variables *cut* and *color*

```
plot(cut,color)
```

To run the chi-square test to study the association between two categorical variables, *cut* and *color*

Null hypothesis H_0 : Independence (no association)

Alternative hypothesis H_1 : Dependence (association)

The decision of the test depends on the *p*-value.

If *p*-value < 0.05, there is an evidence against H_0 (we reject H_0).

If *p*-value $\geq$ 0.05, there is an evidence in favor of H_0 (we do not reject H_0)

```
chisq.test(cut,color)
```