Numerical variables

Suppose we have the file *diamond.xlsx* containing 10 variables. The numerical variables are: *Carat*, *Depth*, *Table*, *Price*, *x*, *y*, *z*.
Consider the variable *Carat* with values column A in the cells A2 to A341.

To compute the mean of the variable *Carat*
```
=AVERAGE(A2:A341)
=MEDIA(A2:A341)
```

To compute the median of the variable *Carat*
```
=MEDIAN(A2:A341)
=MEDIANA(A2:A341)
```

To compute the minimum value of the variable *Carat*
```
=MIN(A2:A341)
```

To compute the maximum value of the variable *Carat*
```
=MAX(A2:A341)
```

To compute the first quartile of the variable *Carat*
```
=PERCENTILE(A2:A341;0.25)
```

To compute the third quartile of the variable *Carat*
```
=PERCENTILE(A2:A341;0.75)
```

To compute the 90th-percentile of the variable *Carat*
```
=PERCENTILE(A2:A341;0.90)
```

To compute the frequency of the value 0.21 for the variable *Carat*
```
=COUNTIF(A2:A341;"=0.21")
=CONTA.SE(A2:A341;"=0.21")
```

To compute the frequency of the values larger than 0.21 for the variable *Carat*
```
=COUNTIF(A2:A341;"=0.21")
=CONTA.SE(A2:A341;">0.21")
```

To summarize the main characteristics of the variable *Carat*, use the path
```
DATA ➔ DATA ANALYSIS ➔ DESCRIPTIVE STATISTICS
DATI ➔ ANALISI DATI ➔ STATISTICA DESCRITTIVA
```

	To draw the histogram of the variable *Carat,* select the data, then INSERT ➔ GRAPHS ➔ ALL GRAPHS ➔ HISTOGRAM INSERISCI ➔ GRAFICI ➔ TUTTI I GRAFICI ➔ ISTOGRAMMA Consider the variables *Carat* with values column A in the cells A2 to A341 and the variable *Price* with values in column G in the cells G2 to G341. To draw the scatterplot between variables *Carat* and *Price,* select the data, then INSERT ➔ GRAPHS ➔ ALL GRAPHS ➔ SCATTER PLOT INSERISCI ➔ GRAFICI ➔ TUTTI I GRAFICI ➔ DISPERSIONE (XY) To compute the correlation between variables *Carat* and *Price* =CORREL(A2:A341;G2:G341) =CORRELAZIONE(A2:A341;G2:G341)

Import the dataset in the file *diamond.xls.*

To define the variable *carat* from the dataset *diamond*

```
carat=diamond$Carat
```

To compute the mean of the variable *carat*

```
m=mean(carat)
```

To compute the median of the variable *carat*

```
Me=median(carat)
```

To compute the minimum value of the variable *carat*

```
min(carat)
```

To compute the maximum value of the variable *carat*

```
max(carat)
```

To compute the first quartile of the variable *carat*

```
Q1=quantile(carat,0.25)
```

To compute the third quartile of the variable *carat*

```
Q3=quantile(carat,0.75)
```

To compute the p-quantile of the variable *carat*

```
Qp=quantile(carat,p)
```

To compute the frequency of the value 0.21 for the variable *carat*

```
length(carat[carat==0.21])
```

To compute the frequency of the values larger than 0.21 for the variable *carat*

```
length(carat[carat>0.21])
```

To compute minimum value, maximum value, mean, median, first and third quartile of the variable *carat*

```
summary(carat)
```

To draw the histogram of the variable *carat*

```
hist(carat)
```

To draw the histogram of the variable *carat* with 50 bins

```
hist(carat,50)
```

To draw the histogram of the variable *carat* with densities

```
hist(carat,freq=F)
```

To draw the histogram (with densities and 50 bins) and the density plot (in blue) of the variable *carat*

```
hist(carat,10,freq=F)
lines(density(carat,bw=0.1),col="blue")
```

To draw the simplified version of the box plot of the variable *carat*

```
boxplot(carat,range=0)
```

Define the variable *price*.
To draw the scatterplot between variables *carat* and *price*

```
plot(carat,price)
```

To draw the smoothed scatterplot between variables *carat* and *price*

```
smoothScatter(carat,price)
```

To compute the correlation between variables *carat* and *price*

```
cor(carat,price)
```

To run a hypothesis testing on correlation with null hypothesis $H_0 : r = 0$ and alternative hypothesis $H_1 : r \neq 0$. 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)

```
cor.test(carat,price)
```

To draw the coplot (2° version) between the variables *carat* and *price* conditionally on the variable *cut* of the dataset *diamond*

```
cut=diamond$Cut
cut=as.factor(cut)
plot(carat,price,col=cut)
legend("bottomright",legend=levels(cut),pch=1,col=1:5)
```

To draw the scatterplot matrix of the variables *carat*, *price,* and *x* of the dataset *diamond*

```
plot(diamond[,c("carat","price","x")])
```