ANALYSING STUDENT SURVEY DATA: MICROSOFT EXCEL® GUIDE

Dr. Syam Kumar
Dr. Akinlolu Akande
Dr. David Obada

Ollscoil Teicneolaíochta an Atlantaigh
Atlantic Technological University

NTU TORR
Transforming Learning

This document was created to help undergraduate students in understanding data and statistics concepts. Additionally, postgraduate students who are embarking on their research journey and require a quick review of fundamental concepts in data and statistics will also find value in this resource.

Furthermore, we would like to acknowledge funding and support received from the National Technological University TransfOrmation for Resilience & Recovery (NTUTORR). The title of the project funded is "*Implementing and evaluating project-based learning (PBL) in undergraduate introductory statistics modules at ATU Sligo – A Pilot Study*".

akinlolu.akande@atu.ie

syam.kumar@atu.ie

david.obada@atu.ie

VERSION: 01

APRIL 2024

SCHOOL OF SCIENCE, ATLANTIC TECHNOLOGICAL UNIVERSITY SLIGO

AVAILABLE RESOURCES FOR STUDENTS AND INSTRUCTORS

This material is part of our collection of resources that have been developed from over a decade of teaching introductory statistics and data analysis to undergraduate students at ATU Sligo and various Irish tertiary institutions, as well as providing Microsoft Excel classes and trainings as part of our Information Technology Modules. It assumes the use of computer software, particularly recommending MS Excel for beginners before progressing to more advanced statistical tools like SPSS or R. As such, we have used Excel to implement the various analysis and statistical tests.

Primarily aimed at undergraduate modules in science, health, business, and related disciplines that leverage MS Excel for data analysis, our materials are intended to be comprehensive resources. Accompanying these resources is a dedicated repository at https://thewee.link/NTUTORR-PBL-Resources, offering a lot of useful resources for both students and instructors.

- Access to pertinent datasets required for completing exercises within this resource.
- Detailed instructions on conducting data and statistical analysis using MS Excel.
- Provision of MS Excel outputs to aid students in verifying their work during exercises.
- Sample answers to questions featured within this resource.
- A compilation of notes covering introductory statistics and data analysis.
- Additional datasets suitable for project-based learning (PBL) endeavours.

NOTE

This material serves as a comprehensive Microsoft Excel Guide, complementing the exercises featured in the Student Survey dataset available on the afore mentioned repository. It offers a detailed, step-by-step walkthrough tailored specifically for MS Excel users for data/statistical analysis. Each featured exercise within the dataset is systematically demonstrated, ensuring learners have access to clear and concise instructions to effectively carry out the data analysis tasks.

TABLE OF CONTENTS

DESCRIPTIVE STATISTICS

UNIVARIATE ANALYSIS OF CATEGORICAL DATA

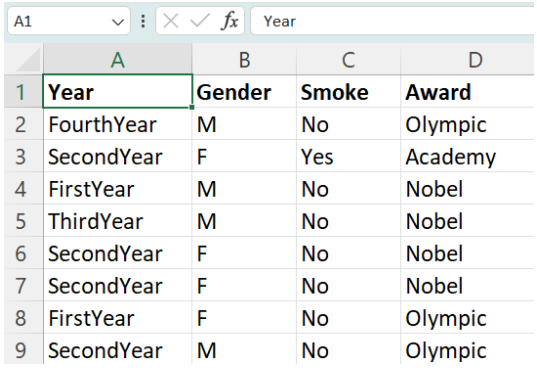
Question: Using the CleanedStudentSurvey Excel file, create a frequency table showing the frequency, relative frequency, and percentage of preferred award *(Award)* among the students surveyed. Also, produce a bar chart and a pie chart to display the preferred award *(Award)* among the students surveyed.

Tabular Summaries: Frequency Distribution

The frequency table and corresponding charts can be created in Excel by using the PivotTable feature. To begin the analysis of the data, one needs to insert the PivotTable for the data under study. For the PivotTable feature to work well, organise the data in a tabular format with column headings.

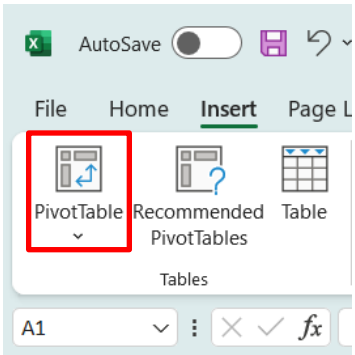
Following are the steps to insert the PivotTable in Excel.

1. Click anywhere within the data range or select the whole data in the worksheet.



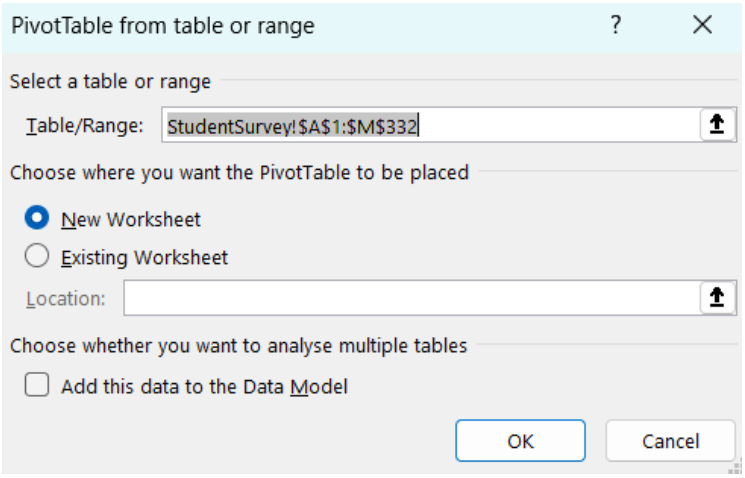
	A	B	C	D
1	**Year**	**Gender**	**Smoke**	**Award**
2	FourthYear	M	No	Olympic
3	SecondYear	F	Yes	Academy
4	FirstYear	M	No	Nobel
5	ThirdYear	M	No	Nobel
6	SecondYear	F	No	Nobel
7	SecondYear	F	No	Nobel
8	FirstYear	F	No	Olympic
9	SecondYear	M	No	Olympic

2. Navigate to the Insert menu on the Excel ribbon and click on the **PivotTable** option located in the "**Tables**" group. This will open the "**Create PivotTable**" dialog box.

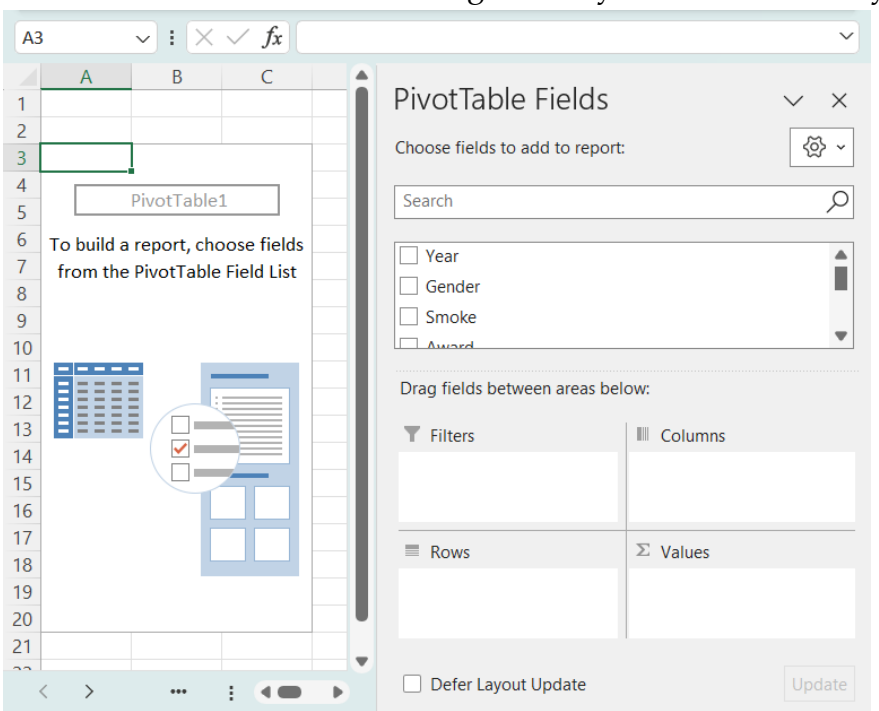


3. In the window that appears, check that the "**Table/Range**" field is displaying the range of data that you are going to analyse. If not, one can manually select the range

at this point. The user then has the option to choose the place to insert the PivotTable. By default, Excel will suggest keeping the PivotTable on a new worksheet, but the user can insert the PivotTable on the same worksheet. For the demonstration, the PivotTable will be placed on a new worksheet. Click the "**OK**" button.

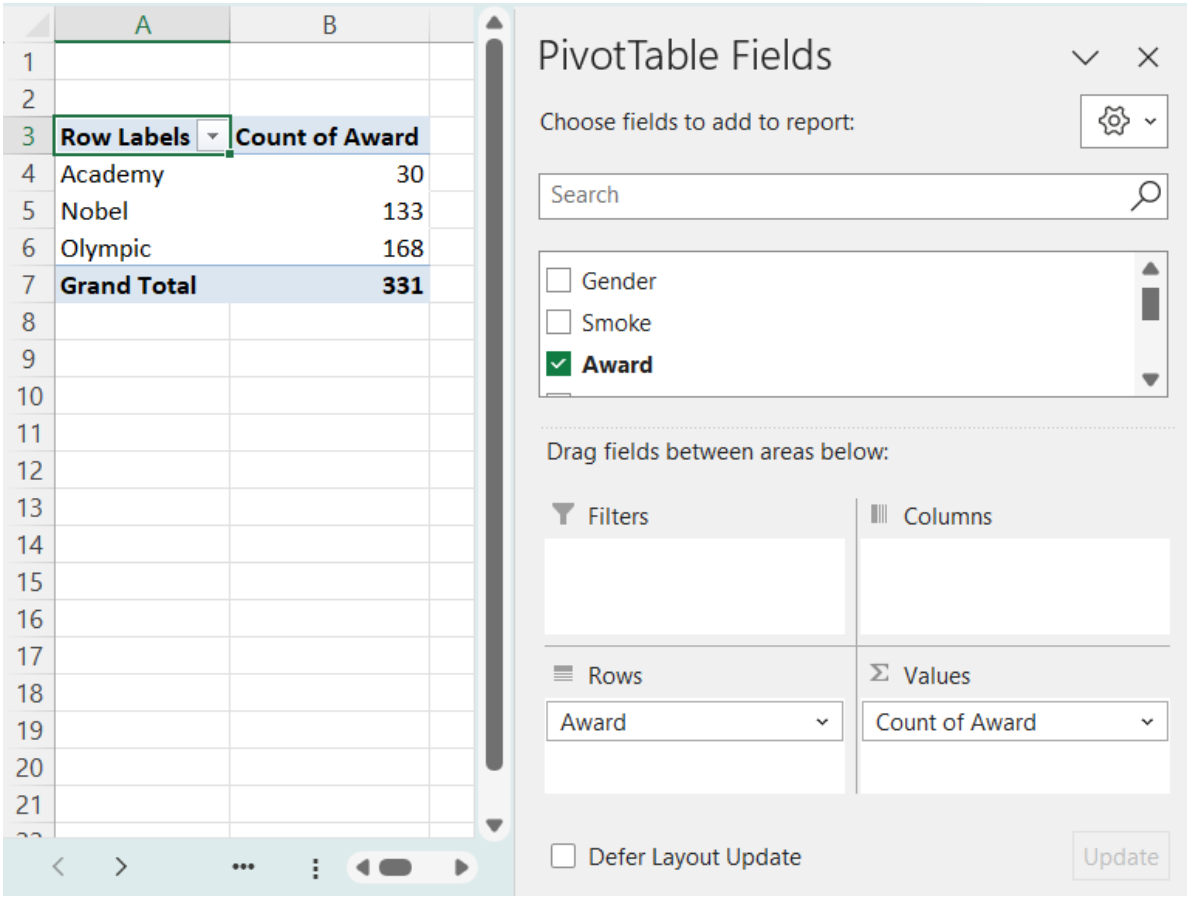


4. A PivotTable will be inserted on a new worksheet named as 'Sheet1' to the left of the current worksheet. There are mainly two sections in this worksheet. On the left, within the range **A3:C20**, a blank PivotTable will be placed and, on the right, just outside of the worksheet, the "**PivotTable Field List**" window can be seen. Note that the name of the worksheet can be changed at any time as deemed by the user.

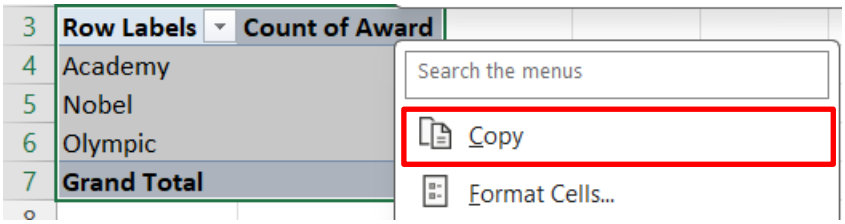


5. In the PivotTable Fields window, all the column headings can be seen as a list. Below that, there are four areas, namely, "**Filters**", "**Columns**", "**Rows**" and "**Values**". By dragging and dropping the column headings to the right areas, one can build a meaningful PivotTable.

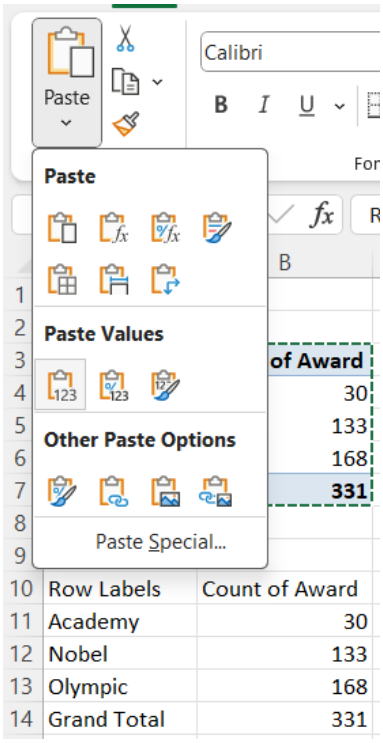
6. To create a frequency table of the preferred award among the students surveyed, one may drag the *Award* field first to the "**Rows**" and then again to "**Values**" areas on the "**PivotTable Field List**" window. Upon this, on the worksheet, user will obtain a table with different *Awards* in the first column and the count of each *Award* in the second column respectively. This is nothing but the Frequency Table.



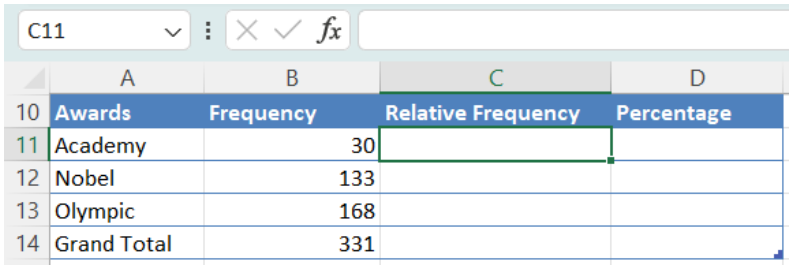
7. One can copy the table values alone to a different location and calculate the relative frequency distribution manually. To do this, select the values. Right click on the selection and click on the "**Copy**" option.



8. Navigate to a location where the data needs to be placed. Click on the cell and then on the "**Paste Values**" option in the Excel ribbon under the "**Home**" tab.



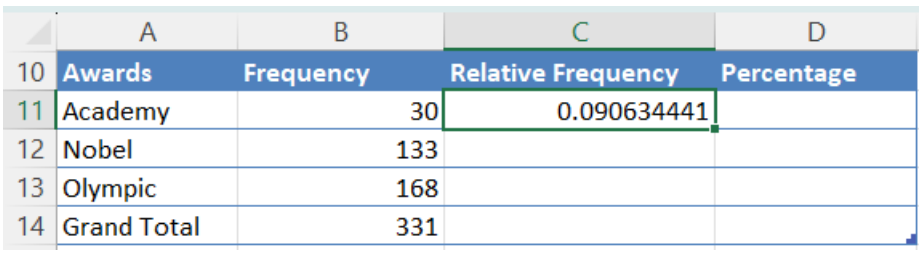
9. Now, one can customise the copied data to make the frequency distribution table as below.



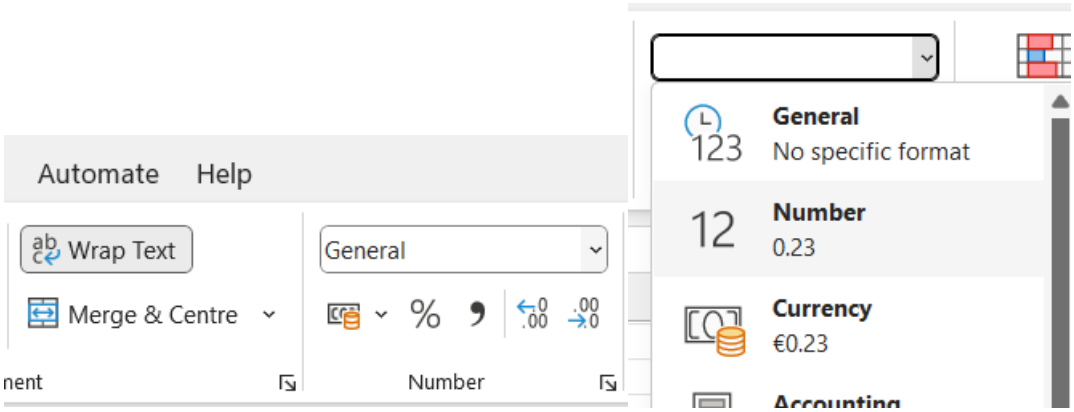
10. To calculate the relative frequency of the first *Award* type (*Academy*) in cell C11, the following formula can be used.

=B11/$B$14

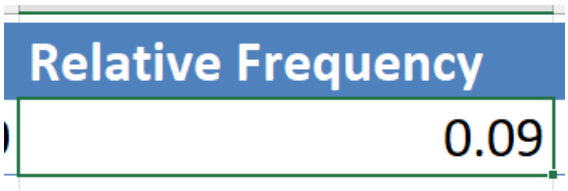
11. The formula will return the relative frequency of the *Academy Award* as shown below.



12. The data type can be converted to *Number* by using the "**Number**" option under the "**Number**" formatting group in the "**Home**" tab. User can keep the desired number of decimal places.



13. Next, the equation can be extended to all the rows by using the "**AutoFill**" option. There are many ways to do this. The easiest way to do this is to double click on the small square that appears in the cell selection.



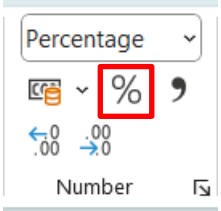
14. Now, the frequency table showing the frequency and the relative frequency of the preferred awards among the students can be obtained as,

	A	B	C	D
10	Awards	Frequency	Relative Frequency	Percentage
11	Academy	30	0.09	
12	Nobel	133	0.40	
13	Olympic	168	0.51	
14	Grand Total	331	1.00	

15. To calculate the percentage, copy the relative frequencies and paste it under the percentage column.

10	Awards	Frequency	Relative Frequency	Percentage
11	Academy	30	0.09	0.09
12	Nobel	133	0.40	0.40
13	Olympic	168	0.51	0.51
14	Grand Total	331	1.00	1.00

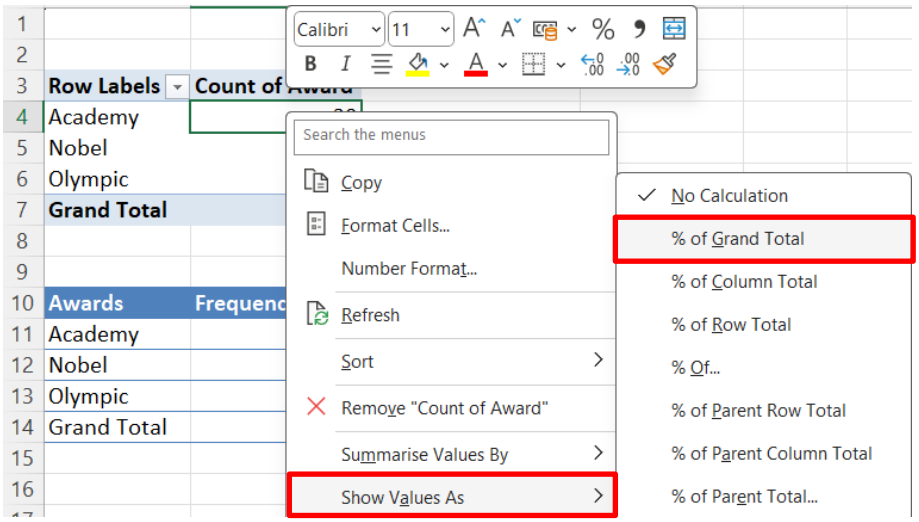
16. Once that is copied to the percentage column, navigate to the "**Number**" group in the "**Home**" tab of Excel. Then, one can simply click on the percentage symbol (%) or manually select the percentage category from the drop-down list.



17. Finally, the frequency table showing the frequency the relative frequency and the percentage of preferred awards among the students surveyed can be obtained as shown below.

Awards	Frequency	Relative Frequency	Percentage
Academy	30	0.09	9%
Nobel	133	0.40	40%
Olympic	168	0.51	51%
Grand Total	331	1.00	100%

18. One can also use the PivotTable feature to calculate the percentage instead of following steps 15 and 16. To do this, right click on one of the cells in the 'Count of Award' column. Click on "**Show Value as**" and then "**% of Grand Total**".



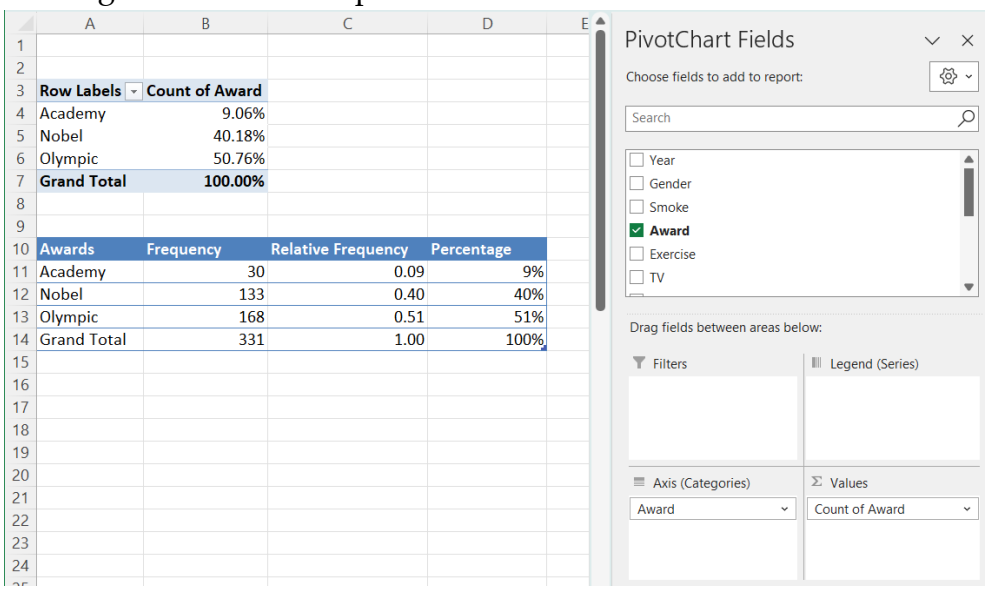
19. The above step will return the percentage value as shown below. These values can be added to the frequency table and the decimal places can be adjusted.

Row Labels	Count of Award
Academy	9.06%
Nobel	40.18%
Olympic	50.76%
Grand Total	**100.00%**

Graphical Summaries: Bar Chart

There are many ways to create a Bar Chart in Excel for a given data. However, if the analysis of the data is carried out using the PivotTable feature of Excel, then a Bar Chart can be created easily. To create a bar chart from a PivotTable in Excel, one can follow these steps.

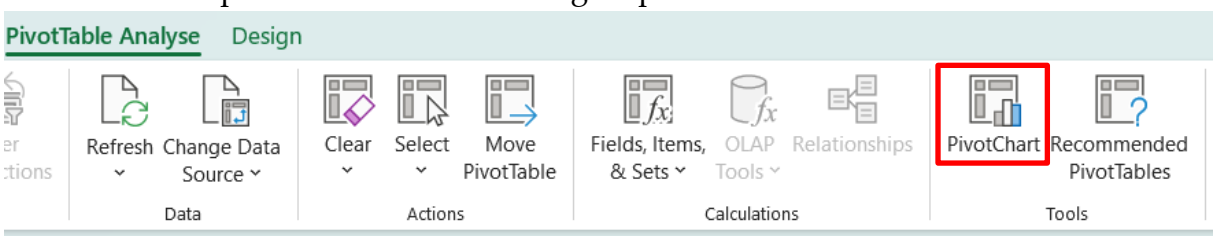
1. Ensure that a PivotTable is already created with the data that needs to be visualised. The data and PivotTable used in the previous section will be considered for demonstration. If the PivotTable is not created before, one may follow the steps 1-7 of the previous section to obtain the following PivotTable and convert the values into percentage as shown in steps 18 and 19.



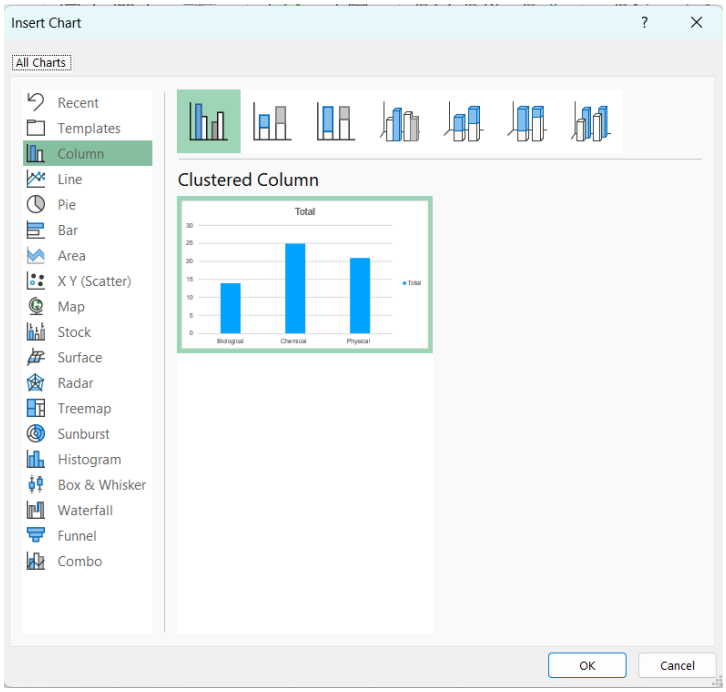
Row Labels	Count of Award
Academy	9.06%
Nobel	40.18%
Olympic	50.76%
Grand Total	100.00%

Awards	Frequency	Relative Frequency	Percentage
Academy	30	0.09	9%
Nobel	133	0.40	40%
Olympic	168	0.51	51%
Grand Total	331	1.00	100%

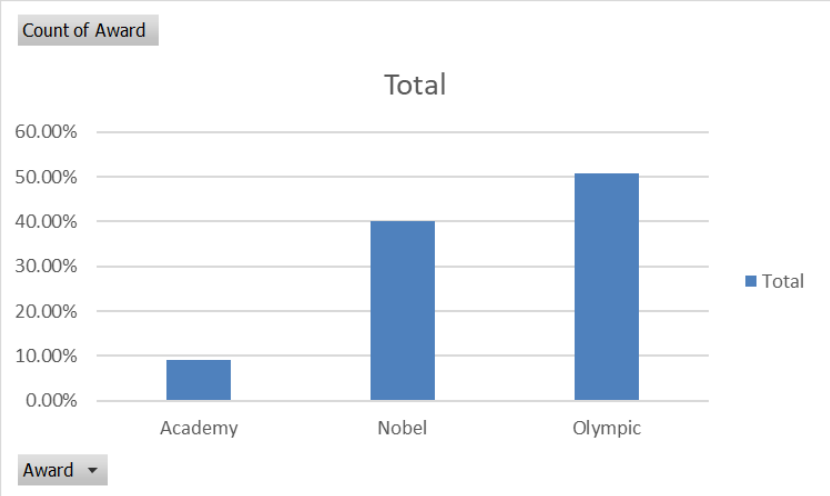
2. Click anywhere inside the PivotTable.
3. Navigate to the "**PivotTable Analyse**" tab in the Excel ribbon and click on the "**PivotChart**" option under the "**Tools**" group.



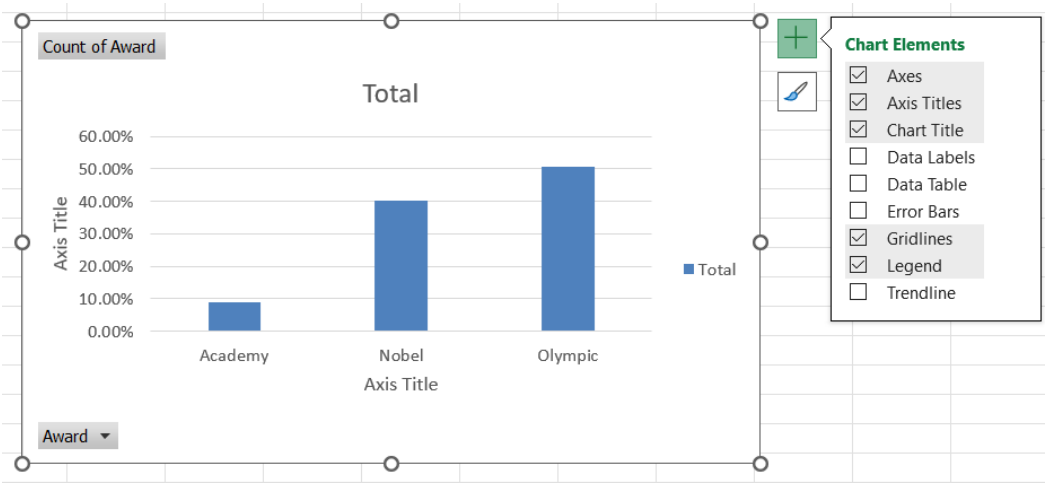
4. "**Insert Chart**" dialog box will be opened. Click on the "**Column**" dropdown button under the "**All Charts**" group, and then click on "**Clustered Column**". Then, click on the "**OK**" button.



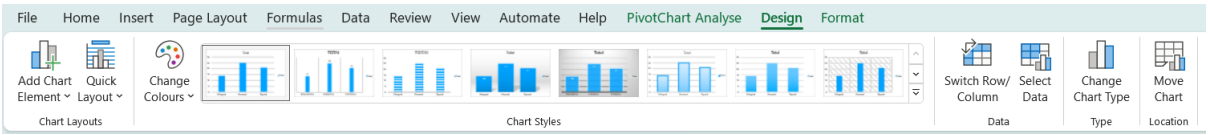
5. A bar chart based on the data currently displayed in the PivotTable will be created in the current worksheet.



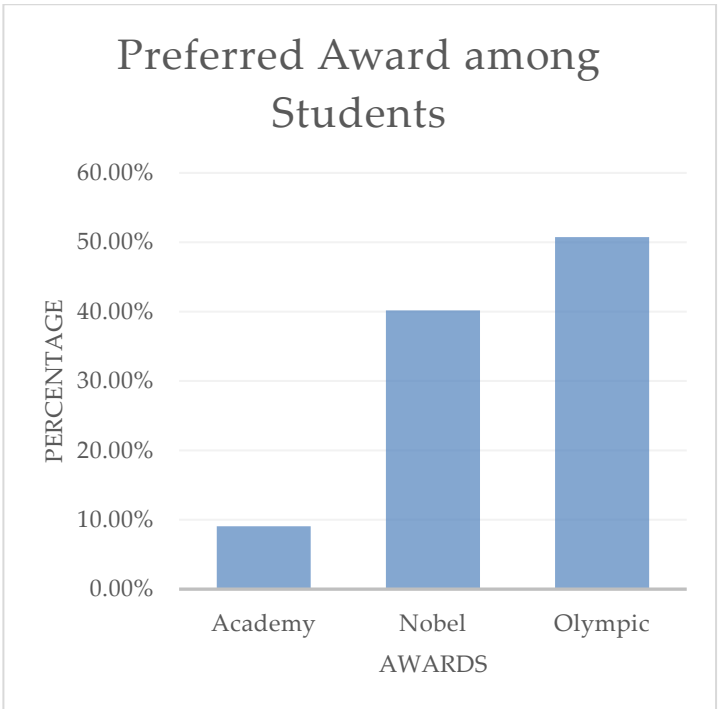
6. The user can now customise or format the bar chart using the "**Chart Elements**" button on the top right of the chart to,
    a. Add data labels,
    b. Show or hide chart title,
    c. Insert Axis labels, or
    d. Add or remove legend.



7. Furthermore, the user can make changes to the design using the tools and templates provided under the "Chart Design" tab in the Excel ribbon.



8. After some formatting, one can obtain the Bar Chart that shows the percentage of *Preferred Awards* among the students surveyed as,

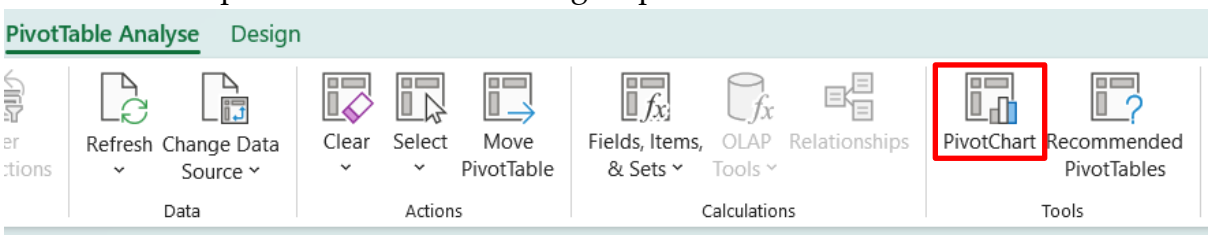


9. If one wants to use the chart in another document, then right-click on the chart and select "**Copy**" or "**Save as Picture**". The chart can then be pasted on to any other document or saved as a picture.

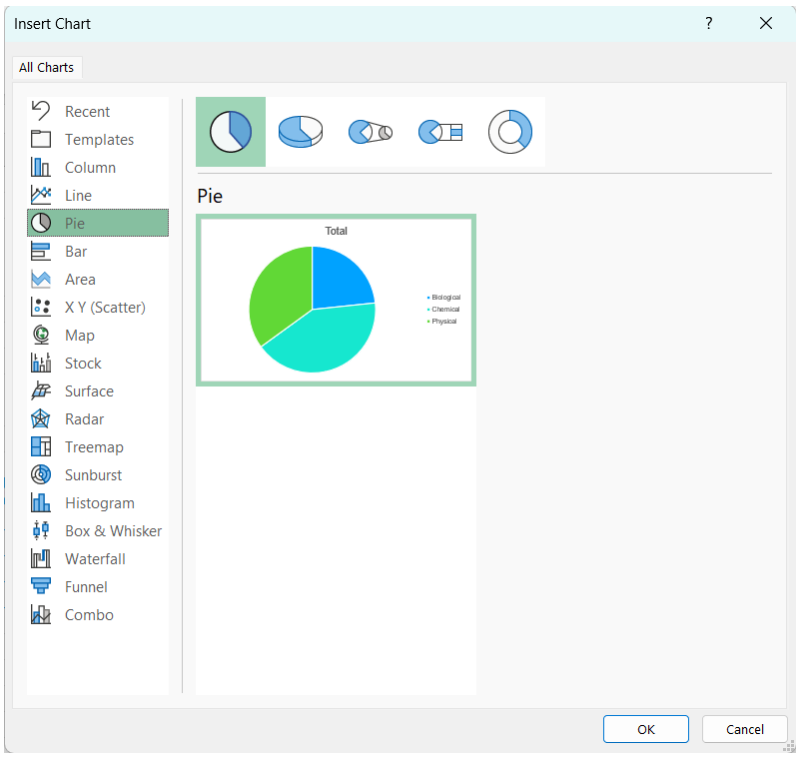
Graphical Summaries: Pie Chart

Using the same PivotTable and data, one can create a Pie Chart as well. To start off, ensure that the PivotTable is containing the data that you want to visualise. If the PivotTable is not created before, one may follow the steps 1-7 of Section 1 to obtain the following PivotTable.

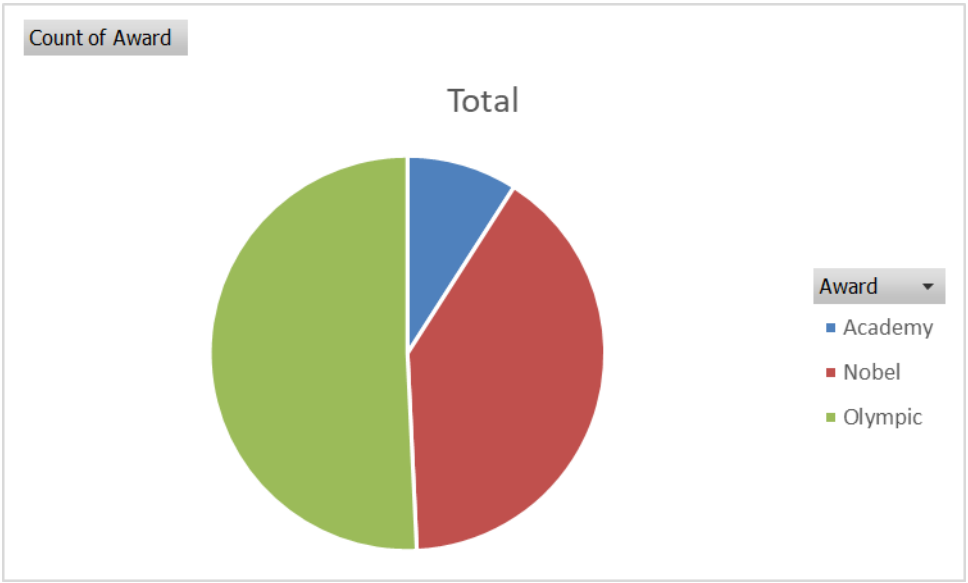
1. Click anywhere inside the PivotTable.
2. Navigate to the "**PivotTable Analyse**" tab in the Excel ribbon and click on the "**PivotChart**" option under the "**Tools**" group.



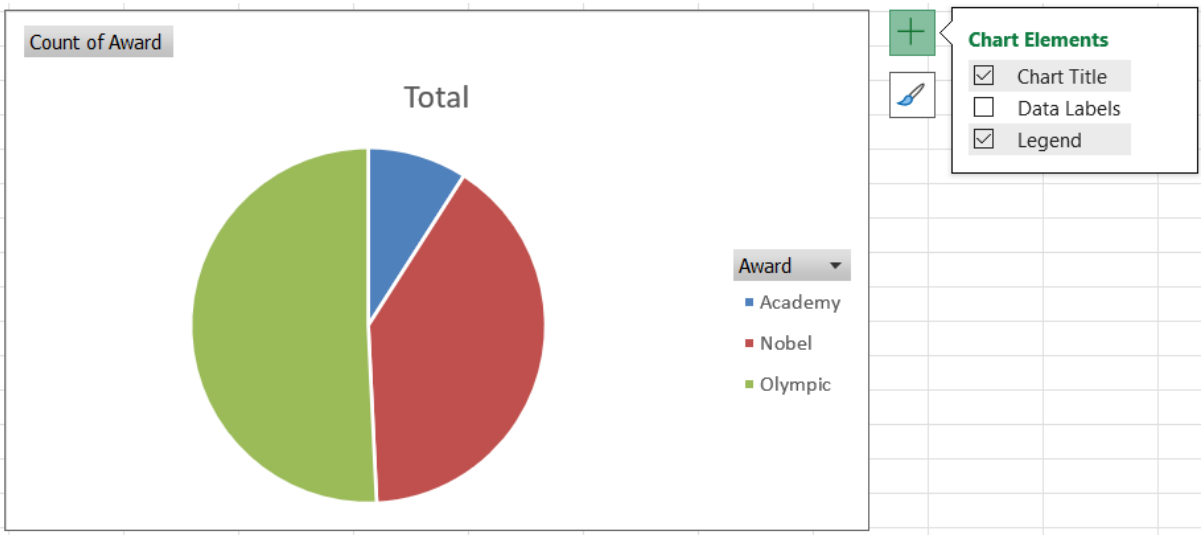
3. In the "**Insert Chart**" dialog box that appears, click on the "**Pie**" dropdown button under the "**All Charts**" group and then click on "**Pie**" chart that is displayed on the right side. Click on the "**OK**" button. Note that there are various pie chart options such as 2-D pie, 3-D pie etc. available to choose from this window. User can select the type that suits the data and preference. For this demonstration, a simple 2-D pie chart will be created.



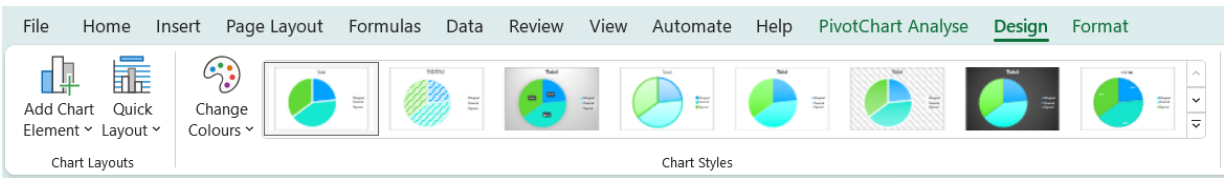
4. In the current worksheet, a pie chart representing the data in the PivotTable will be created.



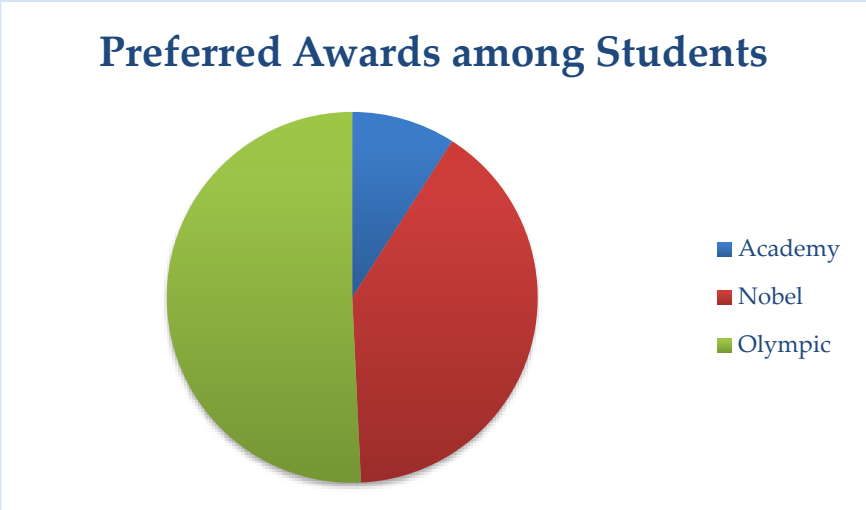
5. The user can customise or format the pie chart using the "**Chart Elements**" button that is provided on the top right of the Pie chart or the "**Chart Design**" tab on the Excel ribbon.



and/or



6. For the current data, one can obtain a Pie chart as shown below after some formatting.



7. One can paste the pie chart on to any other document or save it as a picture by right-clicking on the chart and select "**Copy**" or "**Save as Picture**".

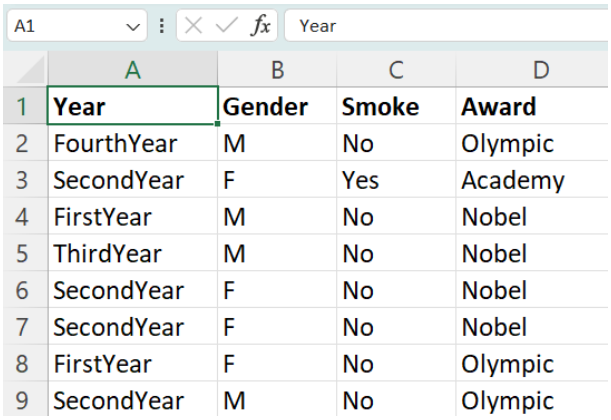
BIVARIATE ANALYSIS OF CATEGORICAL DATA

Question: Using the CleanedStudentSurvey Excel file, create a contingency table to show the joint distribution of *Award* and *Gender* based on frequency and percentage. Also, produce a side-by-side bar chart to display the *Award* by *Gender*.

Tabular Summaries: Crosstabulation or Contingency or Two-way Table

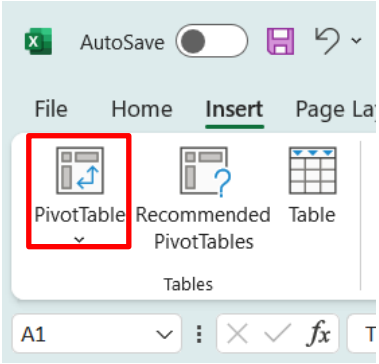
The PivotTable feature of Excel can be used to easily create a Contingency or Two-way table. The steps to create the contingency table using PivotTable feature of Excel are as follows.

1. Make sure that the data is well organised with clear column headings for the PivotTable feature to work well.
2. Click anywhere within the data range or select the whole data in the worksheet.



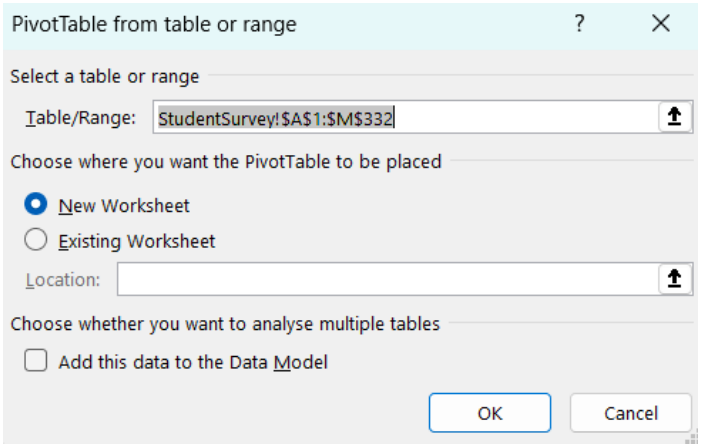
	A	B	C	D
1	**Year**	**Gender**	**Smoke**	**Award**
2	FourthYear	M	No	Olympic
3	SecondYear	F	Yes	Academy
4	FirstYear	M	No	Nobel
5	ThirdYear	M	No	Nobel
6	SecondYear	F	No	Nobel
7	SecondYear	F	No	Nobel
8	FirstYear	F	No	Olympic
9	SecondYear	M	No	Olympic

3. Navigate to the Insert menu on the Excel ribbon and click on the **PivotTable** option located in the "**Tables**" group. This will open the "**Create PivotTable**" dialog box.

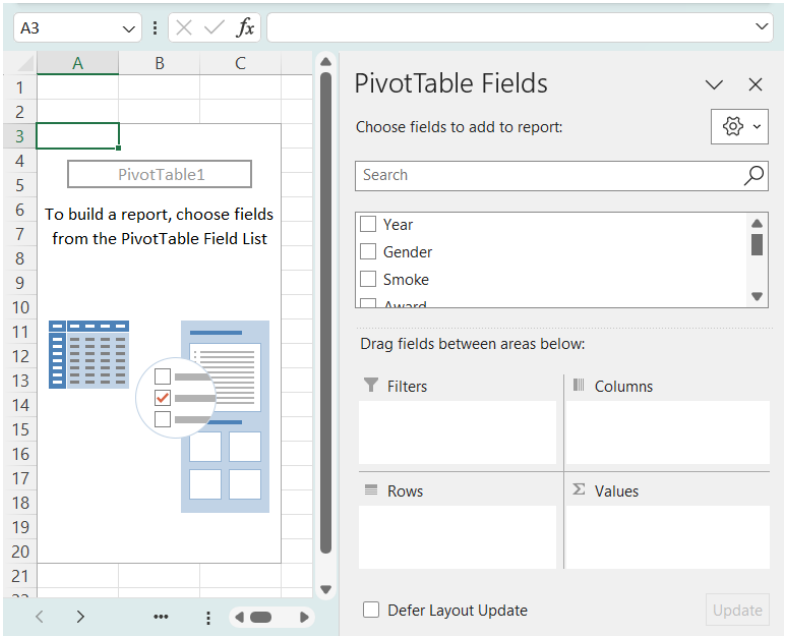


4. In the window that appears, check that the "**Table/Range**" field is displaying the range of data that you are going to analyse. If not, one can manually select the range

at this point. The user then has the option to choose the place to insert the PivotTable. By default, Excel will suggest keeping the PivotTable on a new worksheet, but the user can insert the PivotTable on the same worksheet. For the demonstration, the PivotTable will be placed on a new worksheet. Click the "**OK**" button.

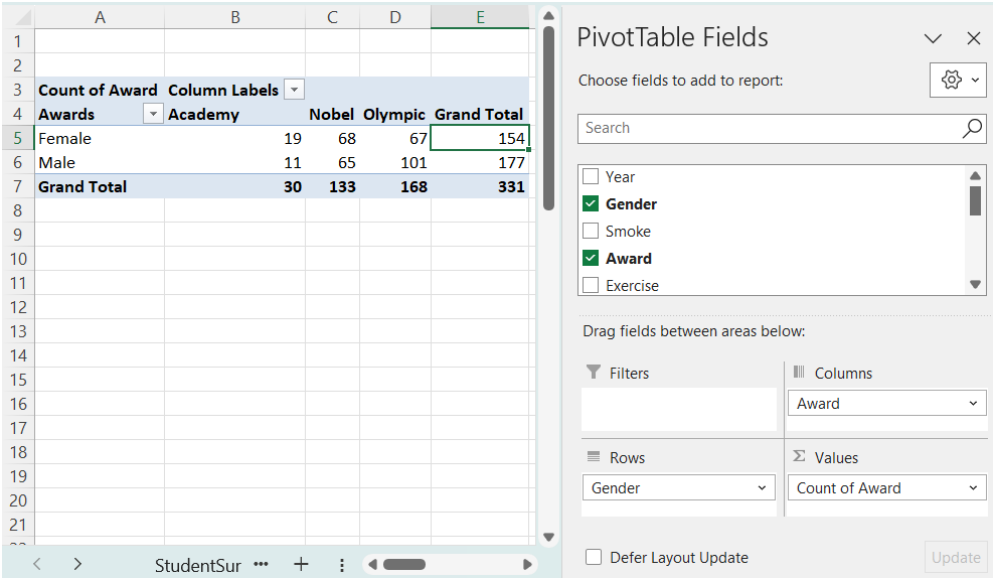


5. A PivotTable will be inserted on a new worksheet named as '**Sheet1**' to the left of the current worksheet. There are mainly two sections in this worksheet. On the left, within the range **A3:C20**, a blank PivotTable will be placed and, on the right, outside of the worksheet, the "**PivotTable Field List**" window can be seen. Note that the name of the worksheet can be changed as a user wish at any time.

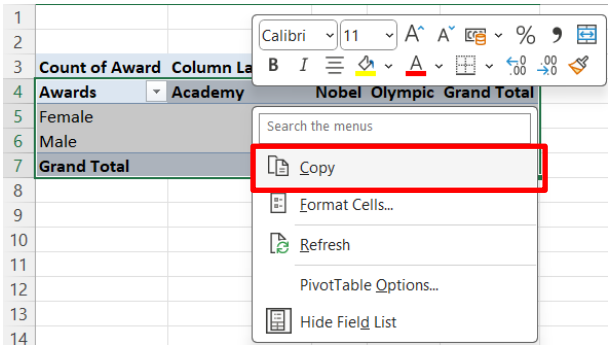


6. In the PivotTable Fields window, all the column headings can be seen as a list. Below that, there are four areas, namely, "**Filters**", "**Columns**", "**Rows**" and "**Values**". By dragging and dropping the column headings to the right areas, one can build a meaningful PivotTable.
7. Now, to create a frequency table to show the joint distribution of *Award* and *Gender*, one may drag the *Gender* field first to the "**Rows**" on the "**PivotTable Field List**"

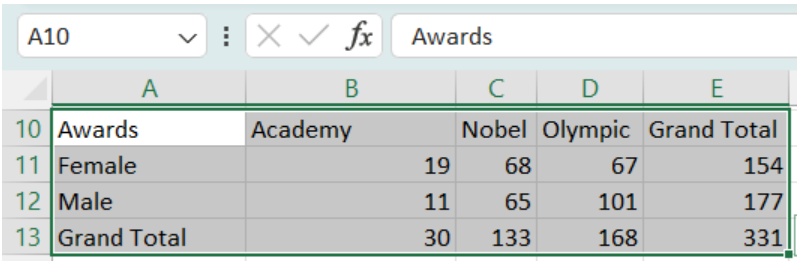
window. Drag and drop *Award* field to the "**Columns**" and then again to "**Values**" areas. The user will then obtain the two-way table that shows the different *Awards* in the first column and the *Gender* in the following columns in the PivotTable.



8. This can be copied to another location or document and formatted. To do this, select the values. Right click on the selection and click on the "**Copy**" option.



9. Navigate to a location where the data needs to be placed. Click on the cell and then on the "**Paste Values**" option in the Excel ribbon under the "**Home**" tab.

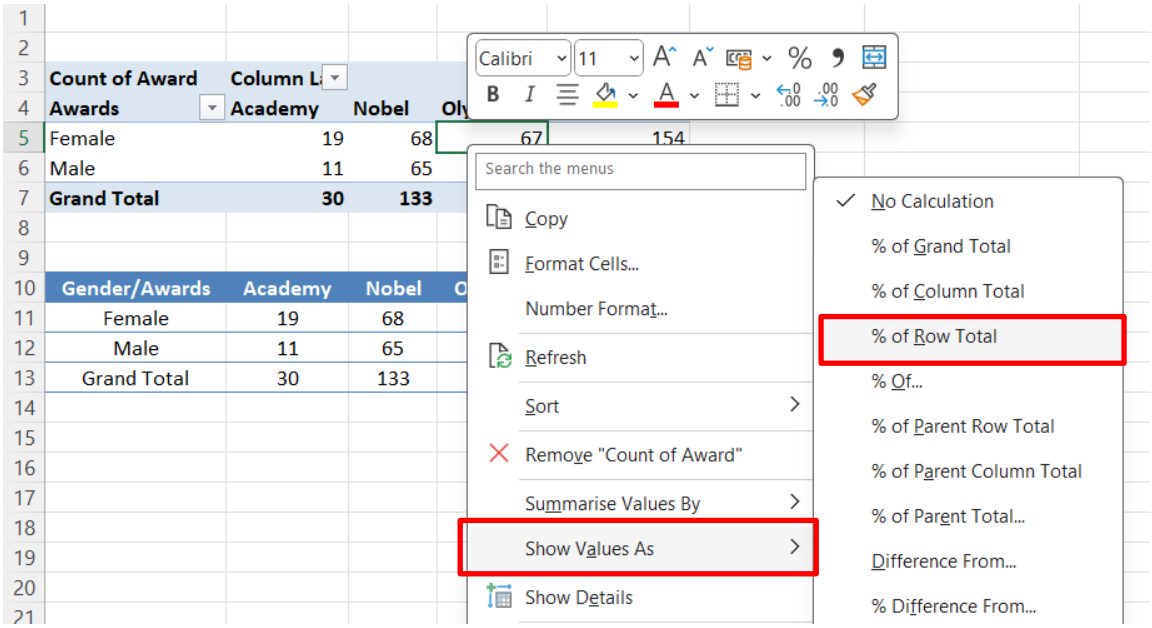


10. After some formatting, one can obtain the contingency table as below.

Gender/Awards	Academy	Nobel	Olympic	Grand Total
Female	19	68	67	154
Male	11	65	101	177
Grand Total	30	133	168	331

11. Furthermore, one can convert these numbers into percentages for further analysis. To do this, click on one of the values in the PivotTable, right-click and click on

"**Show Values As**" and then on "**% of Row Total**". Remember to keep the dependent variables in rows and the independent variables in columns as in the contingency table shown above.



12. The following table will be generated.

Count of Award	Column Labels			
Awards	Academy	Nobel	Olympic	Grand Total
Female	12.34%	44.16%	43.51%	100.00%
Male	6.21%	36.72%	57.06%	100.00%
Grand Total	**9.06%**	**40.18%**	**50.76%**	**100.00%**

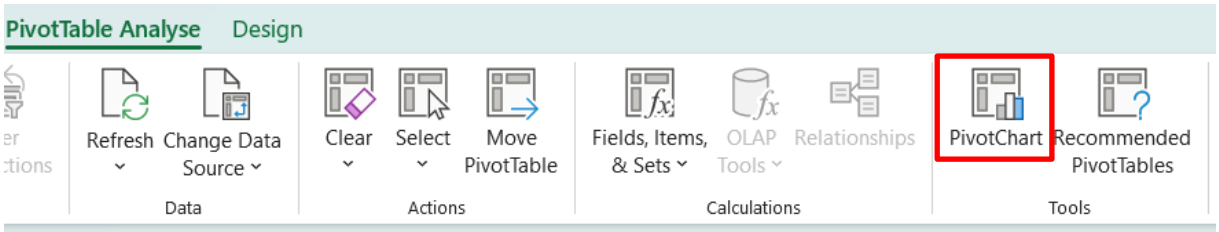
13. Upon further formatting, the contingency table based on percentage values can be obtained as,

Gender/Awards	Academy	Nobel	Olympic	Grand Total
Female	12.34%	44.16%	43.51%	100.00%
Male	6.21%	36.72%	57.06%	100.00%
Grand Total	9.06%	40.18%	50.76%	100.00%

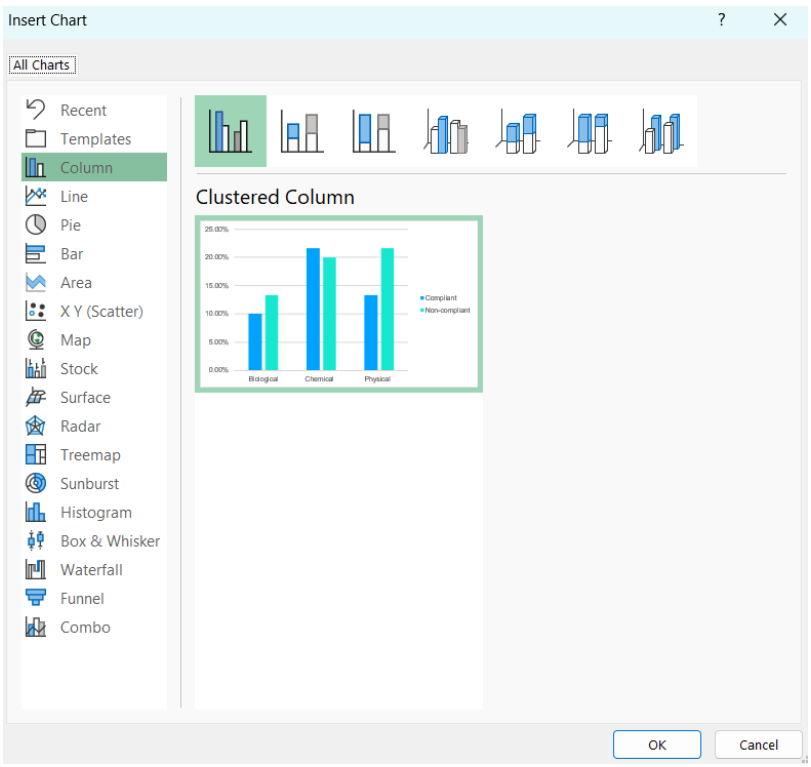
Graphical Summaries: Side by Side Bar Chart

To generate side-by-side bar chart, the PivotTable feature of Excel can be used. To do this make sure that a PivotTable is already generated. The data and PivotTable used in the previous section will be considered for demonstration. The contingency table with percentage values works the best for side-by-side bar charts. Hence, one may follow the steps 1-12 of the previous section to obtain the following PivotTable. To create the side-by-side bar chart, follow the steps below.

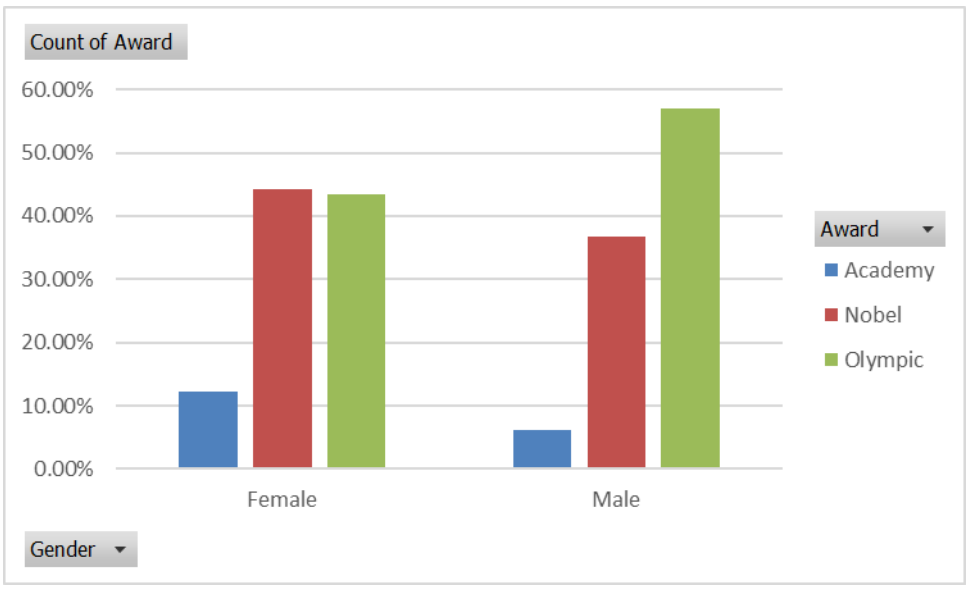
1. Click on any cell inside the PivotTable.
2. Navigate to the "**PivotTable Analyse**" tab in the Excel ribbon and click on the "**PivotChart**" option under the "**Tools**" group.



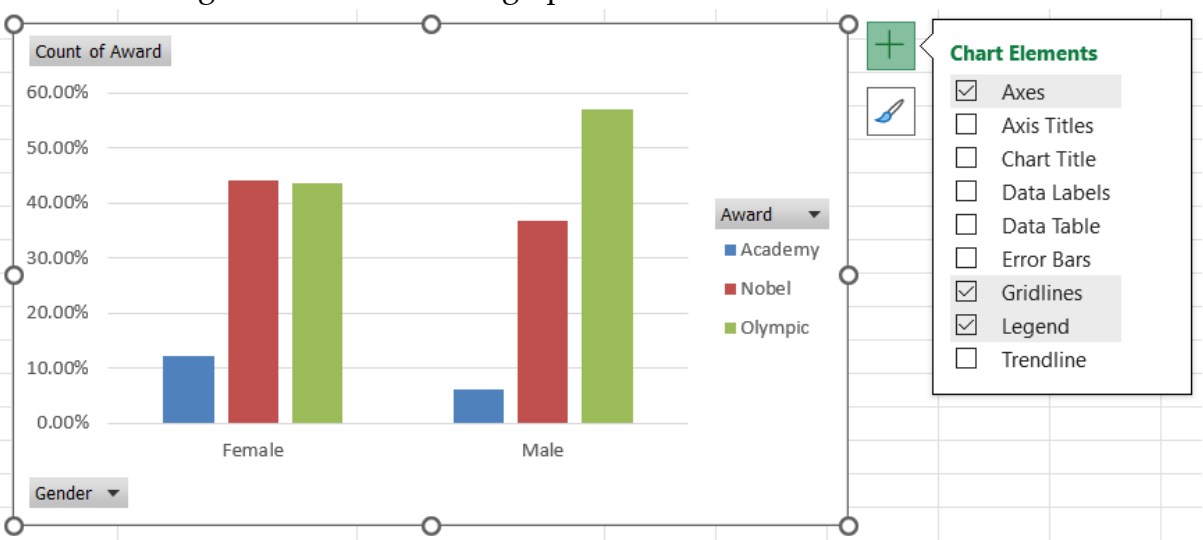
3. "**Insert Chart**" dialog box will be opened. Click on the "**Column**" dropdown button under the "**All Charts**" group and then click on "**Clustered Column**". Then click on the "**OK**" button.



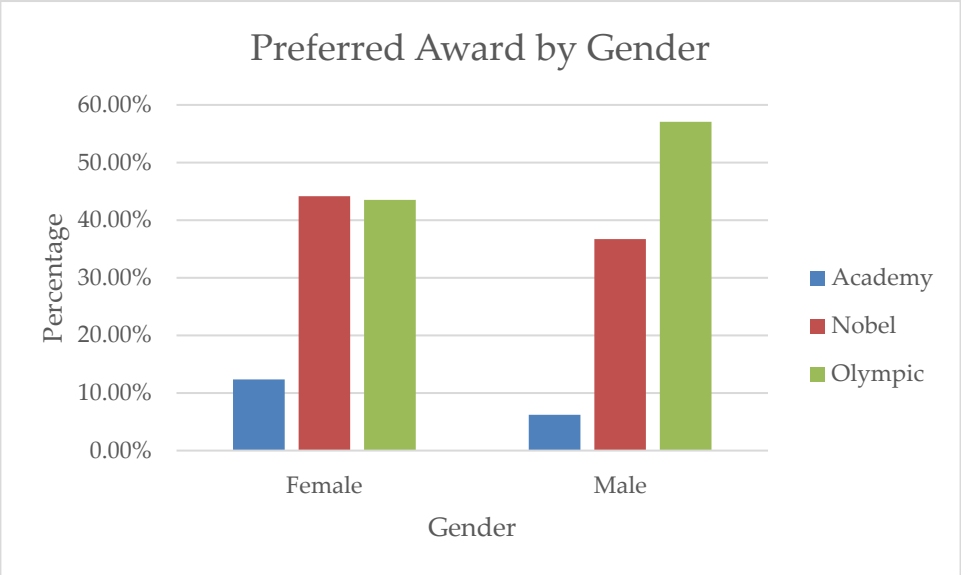
4. A side-by-side bar chart for the data in the PivotTable will be generated as follows.



5. Using the "**Chart Elements**" button near the top right of the side-by-side bar chart or the "**Chart Design"** tab on the Excel ribbon, one can customise and format the chart according to the data and design preference.



6. Upon formatting and customizing, a side-by-side bar chart can be obtained as follows for the data showing the joint distribution of the *Award* by *Gender*.

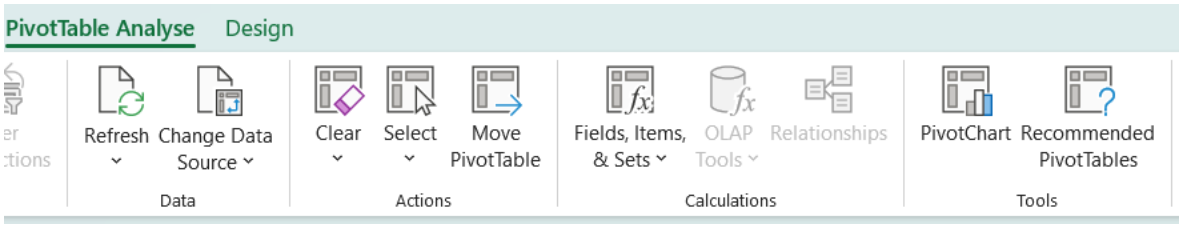


7. If one wants to use the chart in another document, then right-click on the chart and select "**Copy**" or "**Save as Picture**". The chart can then be pasted on to any other document or saved as a picture.

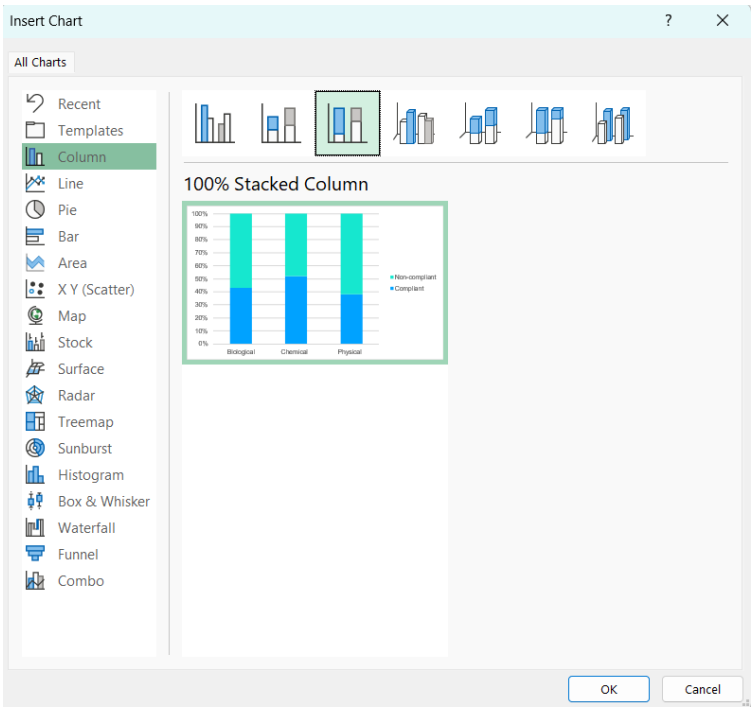
Graphical Summaries: Segmented Bar Chart

Segmented bar charts can also be created using PivotTable in the same way as side-by-side charts. Once again, the visualization needs a PivotTable to begin with. To insert a PivotTable with percentage values, follow steps 1-12 of the "*Tabular Summaries: Crosstabulation or Contingency or Two-way Table*". Once the PivotTable is generated, following are the steps to create segmented bar chart.

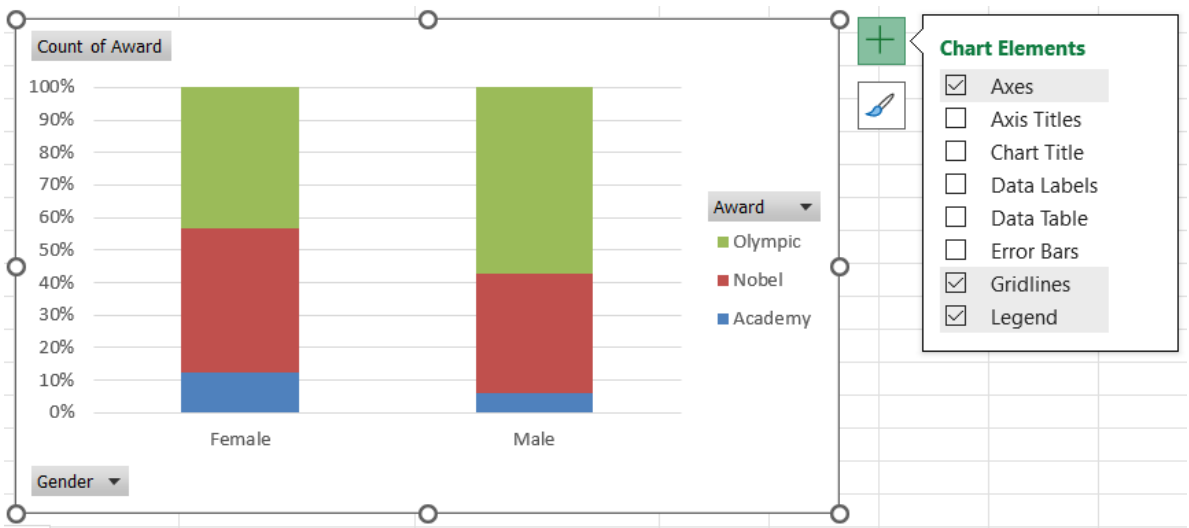
1. Click on any cell inside the PivotTable.
2. Navigate to the "**PivotTable Analyse**" tab in the Excel ribbon and click on the "**PivotChart**" option under the "**Tools**" group.



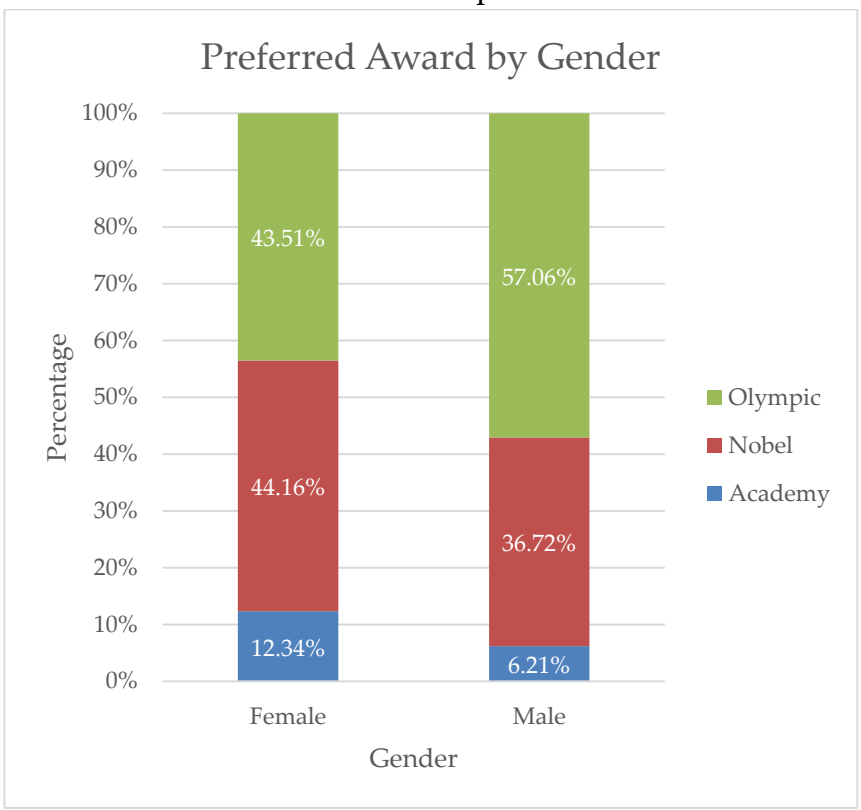
3. "**Insert Chart**" dialog box will be opened. Click on the "**Column**" dropdown button under the "**All Charts**" group and then click on "**100% Stacked Column**". Click on the "**OK**" button.



4. A segmented (100% stacked column) chart will be created on the worksheet based on the active PivotTable. This can be customised and formatted using the "**Chart Elements**" button on the top right of the chart or the "**Chart Design**" tab on the Excel ribbon.



5. After customizing and formatting, one can obtain a segmented bar graph to show the compliance status of different treatment plants as below.



6. The chart thus created can be copied or saved as a picture by right-clicking on the chart and clicking on "**Copy**" or "**Save as Picture**".

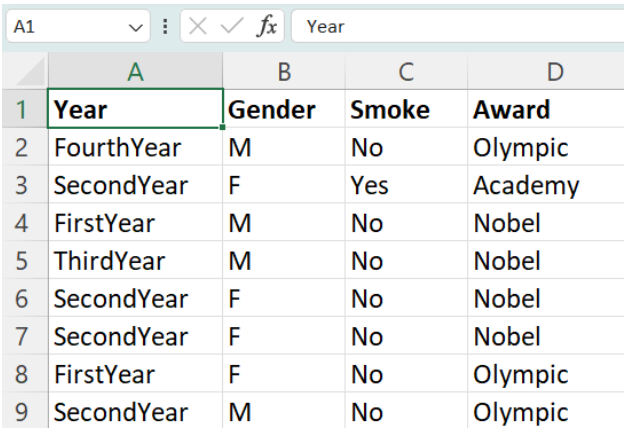
UNIVARIATE ANALYSIS OF QUANTITATIVE DATA

Question: Using the CleanedStudentSurvey Excel file, create a frequency table showing the frequency and relative frequency of the number of hours students exercised per week. Also, produce a histogram and boxplot to display the number of hours students exercised per week.

Tabular Summaries: Frequency Distribution

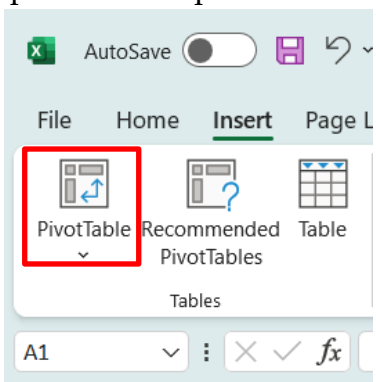
The easiest way to create the frequency distribution in Excel is to use the PivotTable feature. Following are the steps to insert PivotTable in Excel. Make sure the data is well-organised in a tabular form with meaningful column headings.

1. Click anywhere within the data range or select the whole data in the worksheet.



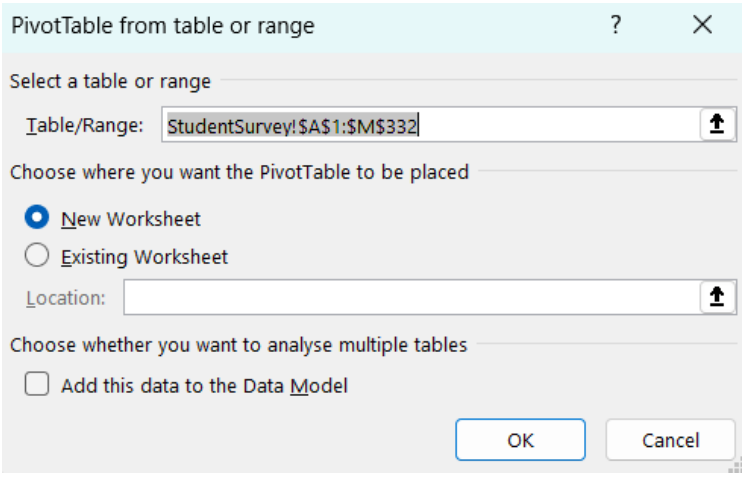
	A	B	C	D
1	**Year**	**Gender**	**Smoke**	**Award**
2	FourthYear	M	No	Olympic
3	SecondYear	F	Yes	Academy
4	FirstYear	M	No	Nobel
5	ThirdYear	M	No	Nobel
6	SecondYear	F	No	Nobel
7	SecondYear	F	No	Nobel
8	FirstYear	F	No	Olympic
9	SecondYear	M	No	Olympic

2. Navigate to the "**Insert**" tab on the Excel ribbon and click on the PivotTable option located in the "**Tables**" group. This will open the "**Create PivotTable**" dialog box.

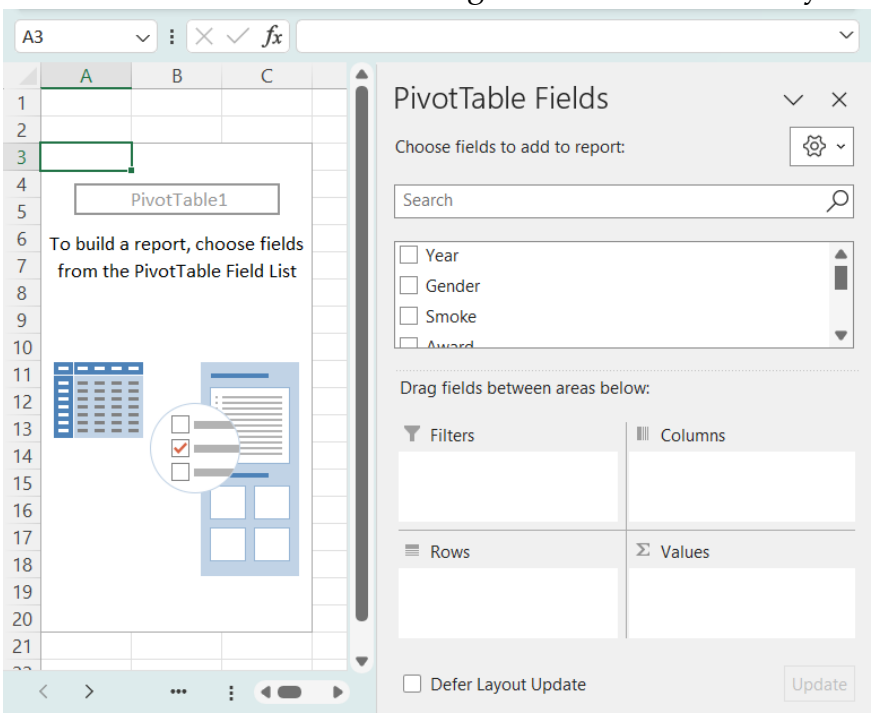


3. In the window that appears, check that the "**Table/Range**" field is displaying the range of data that you are going to analyse. If not, you can manually select the range at this point. The user then has the option to choose the place to insert the

PivotTable. By default, Excel will suggest keeping the PivotTable on a new worksheet, but the user can insert the PivotTable on the same worksheet also. For the demonstration, the PivotTable will be placed on a new worksheet. Next, click the "**OK**" button.

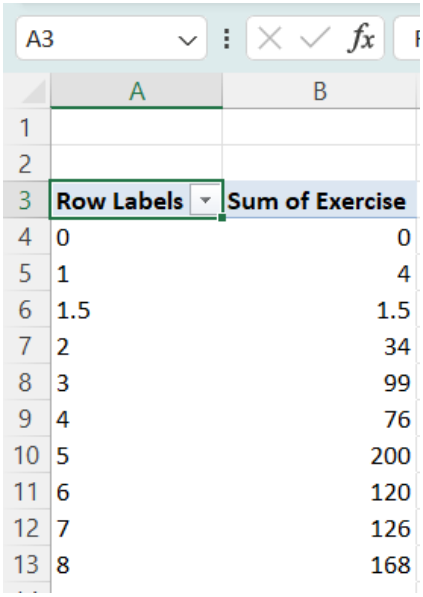


4. A PivotTable will be inserted on a new worksheet named as 'Sheet1' to the left of the current worksheet. There are mainly two sections in this worksheet. On the left, within the range **A3:C20**, a blank PivotTable will be placed and, on the right, outside of the worksheet, the "**PivotTable Field List**" window can be seen. Note that the name of the worksheet can be changed as a user wish at any time.



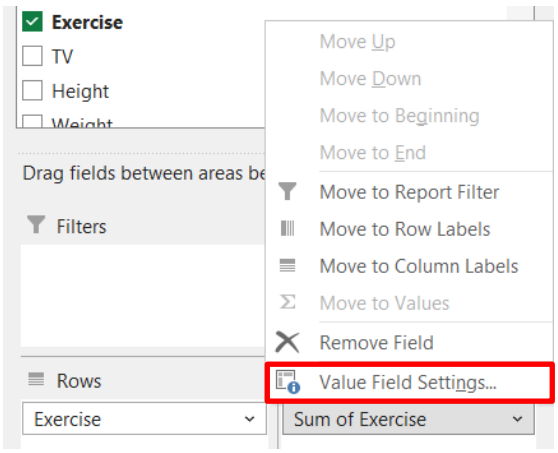
20. In the PivotTable Fields window, all the column headings can be seen as a list. Below that, there are four areas, namely, "**Filters**", "**Columns**", "**Rows**" and "**Values**". By dragging and dropping the column headings to the right areas, one can build a meaningful PivotTable.

21. To create a frequency table of the number of hours students' exercised per week, one may drag the *Exercise* field first to the "**Rows**" and then again to "**Values**" areas on the "**PivotTable Field List**" window. Upon this, on the worksheet, user will obtain a table with different *Exercise* times in the first column and the count of each of them in the second column respectively. This is nothing but the Frequency Table.

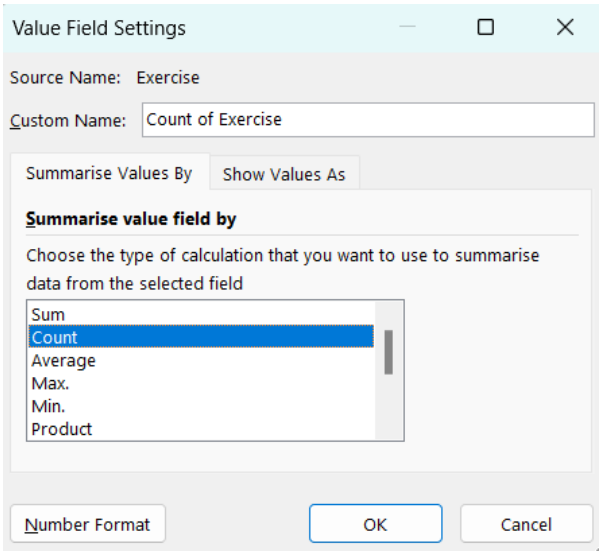


	A	B
1		
2		
3	**Row Labels**	**Sum of Exercise**
4	0	0
5	1	4
6	1.5	1.5
7	2	34
8	3	99
9	4	76
10	5	200
11	6	120
12	7	126
13	8	168

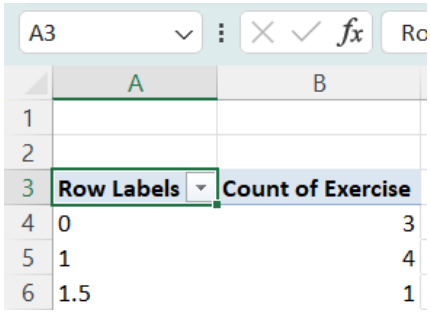
5. However, this is not yet in the form of a frequency table. For this, one needs to change the 'Sum' that appears in Column B to 'Count'. To do this, navigate to the "**Values**" area, click on the dropdown menu placed to the right of "**Sum of Exercise**" and then click on "**Value Field Settings**".



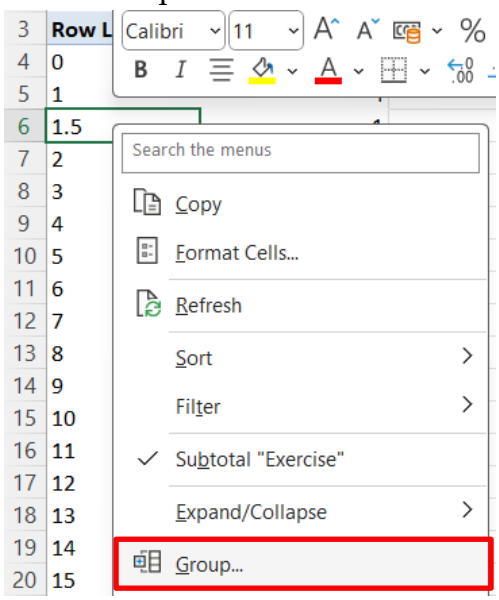
6. In the window that opens, change the "**Summarise Value field by**" 'Sum' to 'Count' and click **OK**.



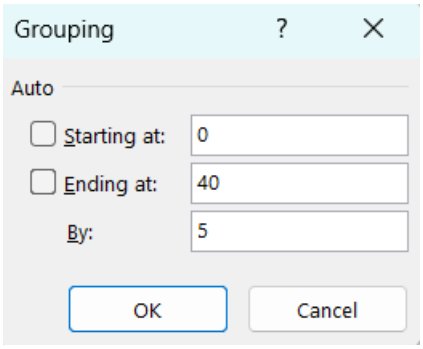
7. This will change the sum shown in Column B of the PivotTable to count.



8. Now, to generate intervals, click one of the cells under Row Labels (in Column A). Right click, and then click on Group.



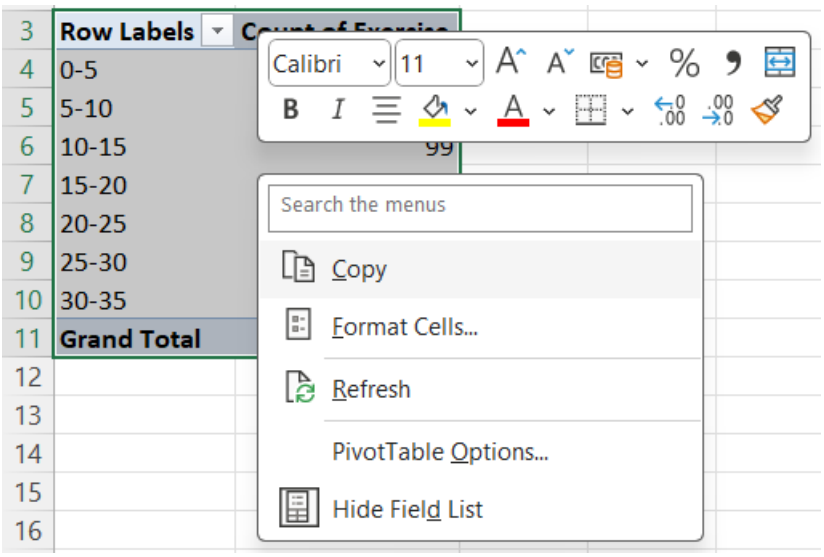
9. In the window that appears, input the start and end points of the data range and the class width. In this case, the start and end points are chosen to be 0 and 40 while the class width is chosen as 5. Click "**OK**".



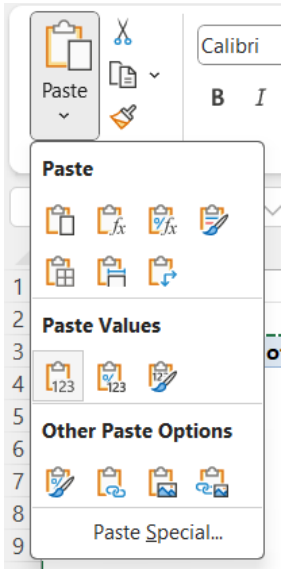
10. This will generate the frequency distribution of the number of hours students exercised as shown below.

	Row Labels	Count of Exercise
4	0-5	77
5	5-10	102
6	10-15	99
7	15-20	33
8	20-25	14
9	25-30	5
10	30-35	1
11	**Grand Total**	**331**

11. Copy the table values alone to a different location and then the relative frequency distribution can be manually calculated. To do this, select the values. Right click on the selection and click on the "**Copy**" option.



12. Navigate to a location where the data needs to be placed. Click on the cell and then on the "Paste Values" option in the Excel ribbon under the "Home" tab.



13. One can customise the frequency distribution table as shown below.

	A	B	C
14	Intervals	Frequency	Relative Frequency
15	0-5	77	
16	5-10	102	
17	10-15	99	
18	15-20	33	
19	20-25	14	
20	25-30	5	
21	30-35	1	
22	Grand Total	331	

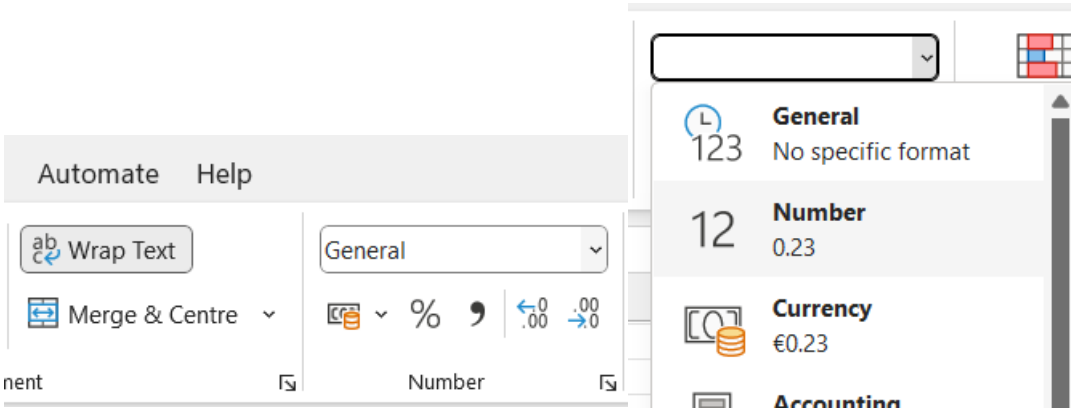
14. To calculate the relative frequency of the number of hours students exercised in cell **C15**, the following formula can be used.

=B15/$B$22

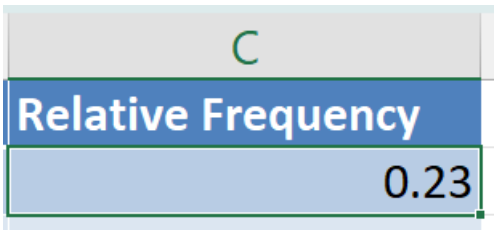
15. This formula will return the relative frequency as shown below.

	A	B	C
14	Intervals	Frequency	Relative Frequency
15	0-5	77	0.232628399
16	5-10	102	
17	10-15	99	
18	15-20	33	
19	20-25	14	
20	25-30	5	
21	30-35	1	
22	Grand Total	331	

16. The data type can be converted to Number by using the "**Number**" option under the "**Number**" formatting group, under the "**Home**" tab and keep the desired number of decimal places.



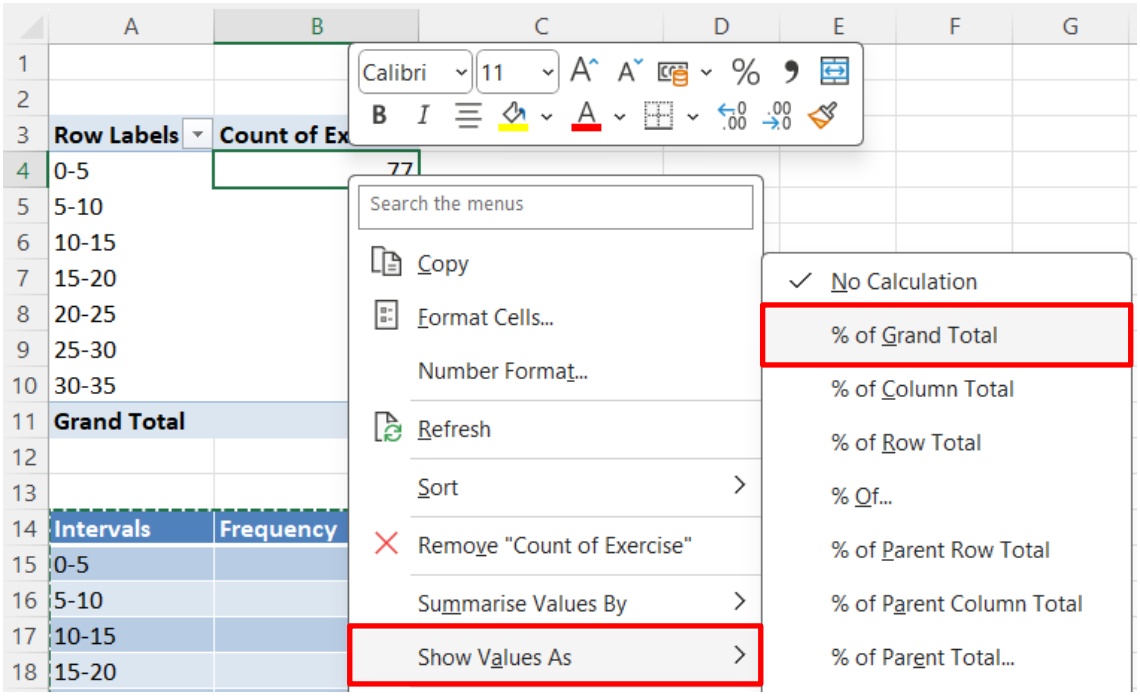
17. Next, the equation can be extended to all the rows by using the AutoFill option. There are many ways to do this. The easiest way to do this is to double click on the small square that appears in the cell selection.



18. The frequency table showing the frequency and the relative frequency of the number of hours students exercised can be obtained as shown below,

Intervals	Frequency	Relative Frequency
0-5	77	0.233
5-10	102	0.308
10-15	99	0.299
15-20	33	0.100
20-25	14	0.042
25-30	5	0.015
30-35	1	0.003
Grand Total	331	1.000

19. To include the percentage of exercise hours falling into each category, right click on one of the cells in the 'Count of Exercise' column in the PivotTable obtained in step 10. Click on "**Show Value as**" and then "**% of Grand Total**".



20. The values in the 'Count of Exercise' column will be converted to percentages.

Row Labels	Count of Exercise
0-5	23.26%
5-10	30.82%
10-15	29.91%
15-20	9.97%
20-25	4.23%
25-30	1.51%
30-35	0.30%
Grand Total	**100.00%**

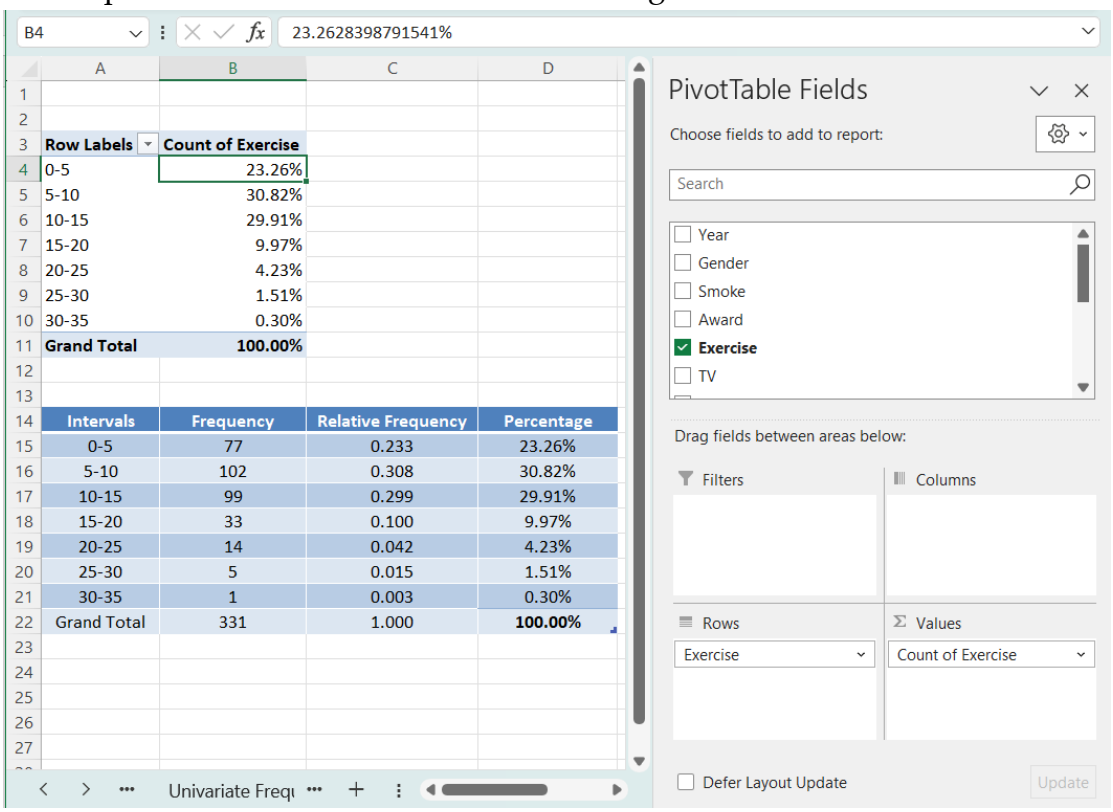
21. The percentage values can be added to the frequency to table to obtain the final frequency table as,

Intervals	Frequency	Relative Frequency	Percentage
0-5	77	0.233	23.26%
5-10	102	0.308	30.82%
10-15	99	0.299	29.91%
15-20	33	0.100	9.97%
20-25	14	0.042	4.23%
25-30	5	0.015	1.51%
30-35	1	0.003	0.30%
Grand Total	331	1.000	**100.00%**

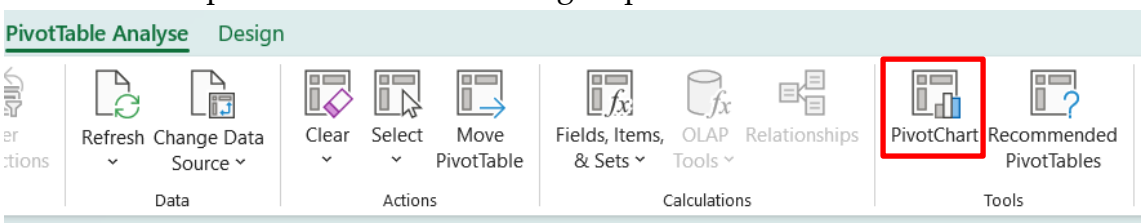
Graphical Summaries: Histogram

Histogram showing the frequency distribution of the number of hours students exercised can be created using the "**PivotTable Analyse**" option once the frequency distribution is created using PivotTable feature of Excel. To create the histogram from a PivotTable in Excel, one can follow the steps given below.

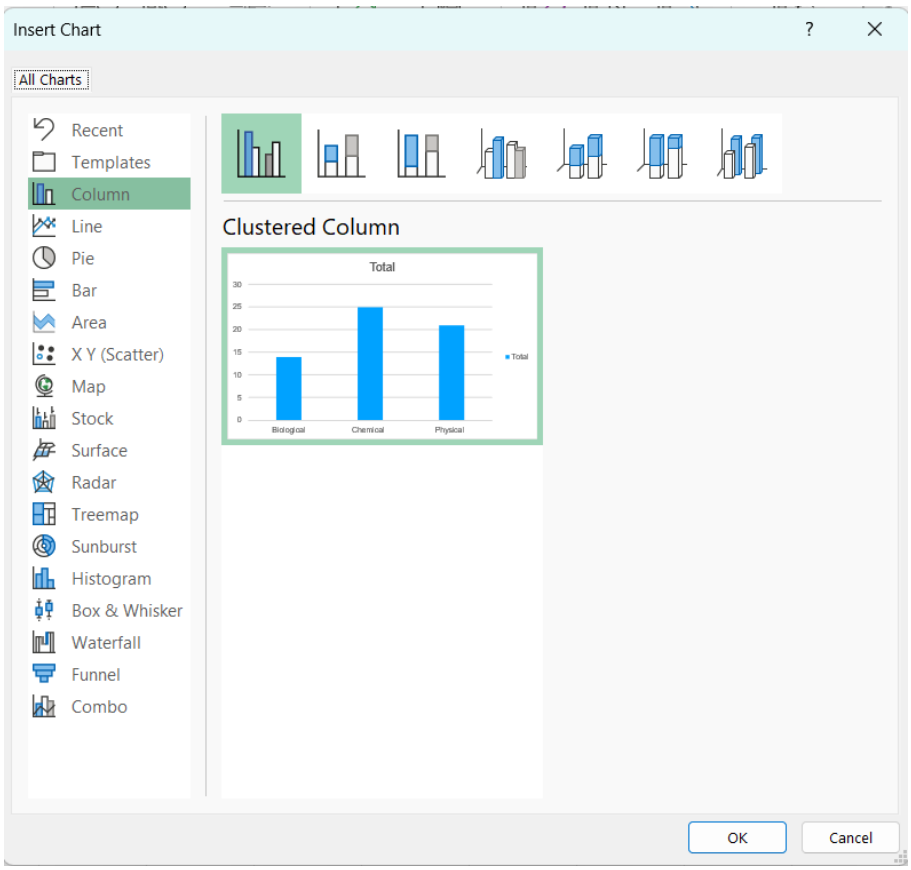
1. Ensure that a PivotTable is already created with the data that needs to be visualised. The data and PivotTable used in the previous section will be considered for demonstration. If the PivotTable is not created before, one may follow the steps 1-20 of the previous section to obtain the following PivotTable.



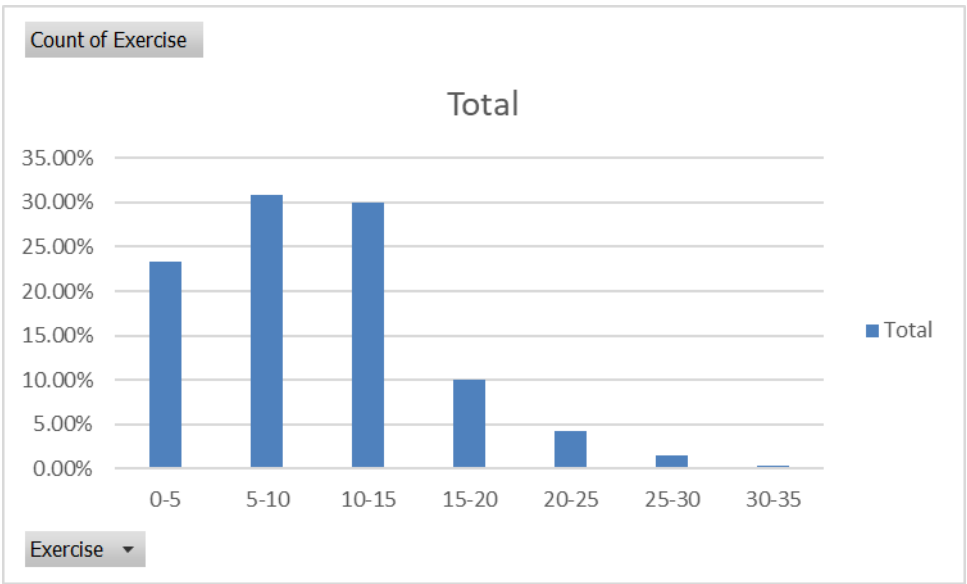
2. Click one of cells in the PivotTable.
3. Navigate to the "**PivotTable Analyse**" tab in the Excel ribbon and click on the "**PivotChart**" option under the "**Tools**" group.



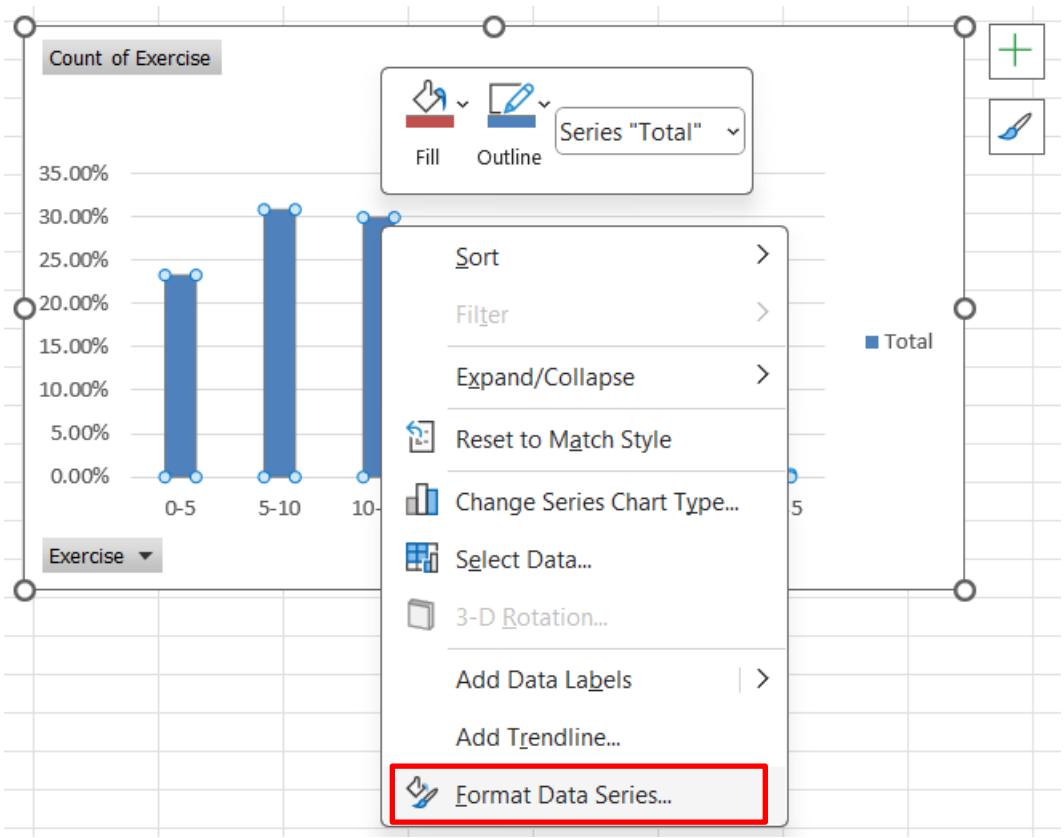
4. "**Insert Chart**" dialog box will open. Click on the "**Column**" dropdown button under the "**All Charts**" group and then click on "**Clustered Column**". Click on "**OK**" button.



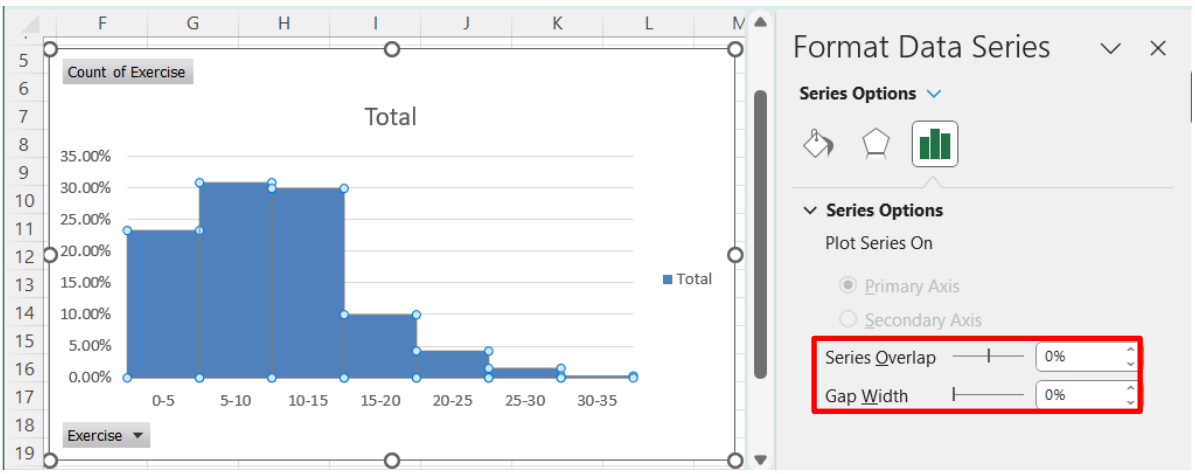
5. A bar chart based on the data currently displayed in the PivotTable will be created in the current worksheet.



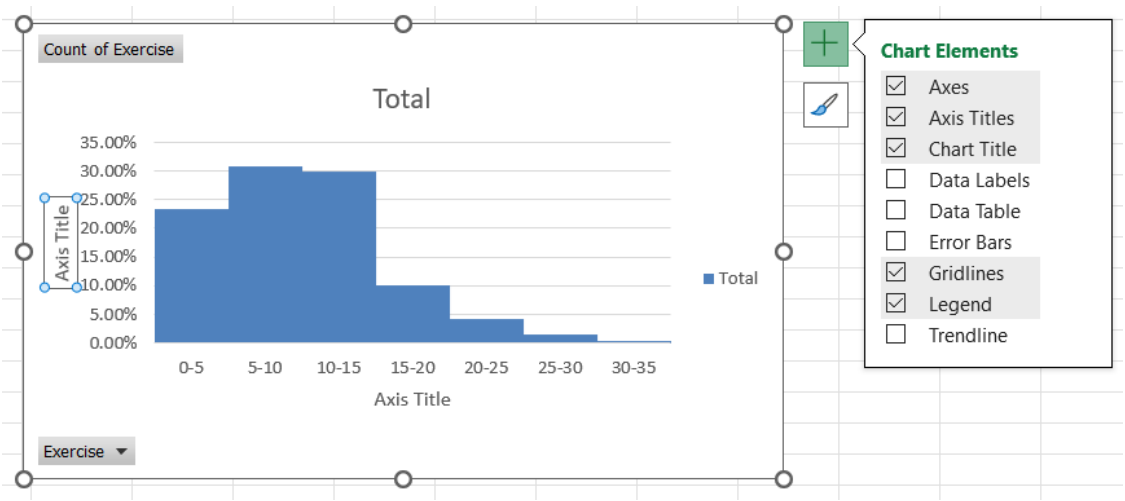
6. Right click on one of the bars in the graph and click on the "**Format Data Series**" option.



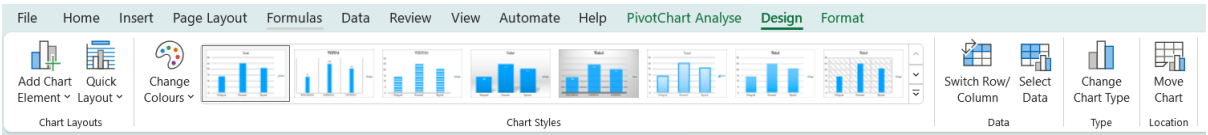
7. On the window that appears on the right side, keep the "**Series Overlap**" and "**Gap Width**" to zero.



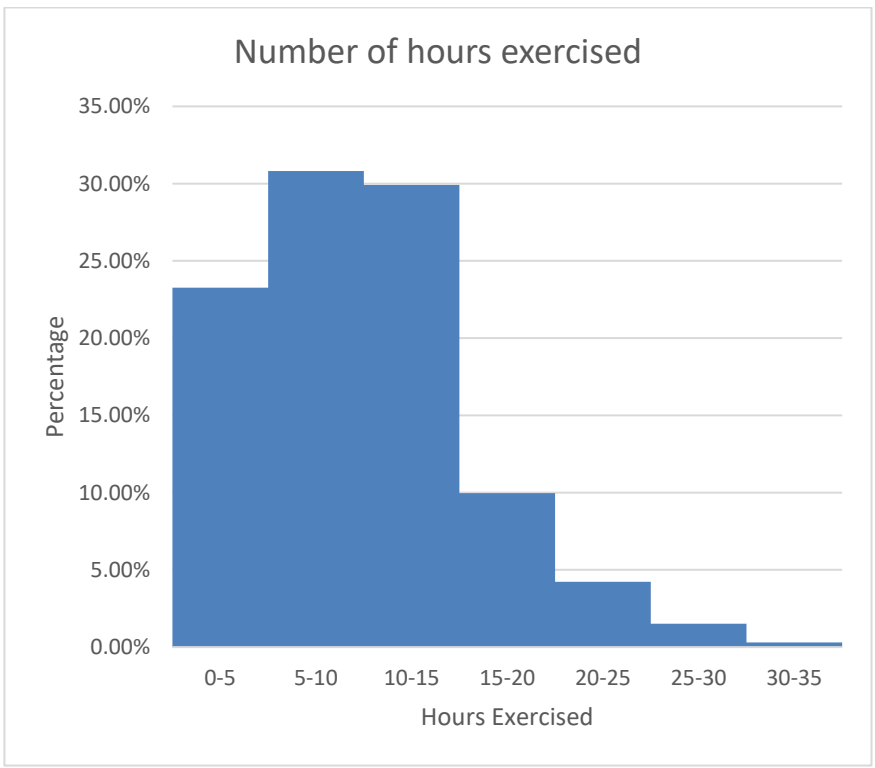
8. The user can now customise or format the histogram using the "**Chart Elements**" button on the top right to,
    a. Add data labels,
    b. Show or hide chart title,
    c. Insert Axis labels, or
    d. Add or remove legend.



9. Furthermore, the user can make changes to the design using the tools and templates provided under the "**Chart Design**" tab in the Excel ribbon.



10. After some formatting, one can obtain the histogram that shows the distribution of the number of hours students exercised as,



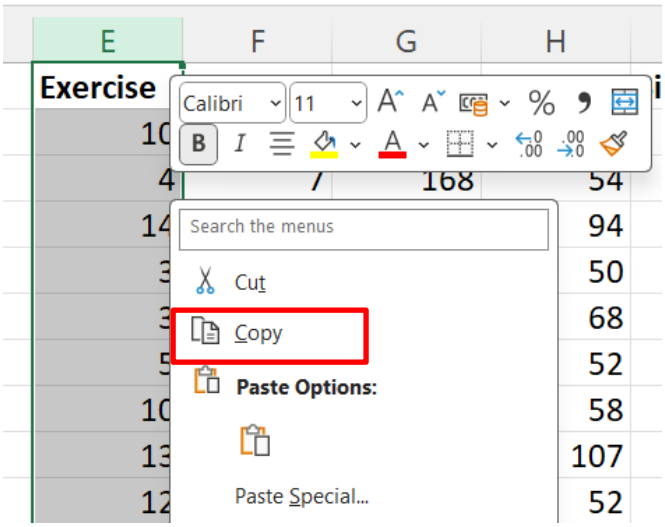
Graphical Summaries: Boxplot

Boxplot for a clean dataset can be created in Excel using the Charts function directly. The boxplot can be created for the distribution of the number of hours students exercised by using the following steps.

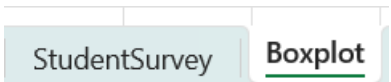
1. Select the column(s) with the data that needs to be plotted.

	A	B	C	D	E	F	G	H	I
1	**Year**	**Gender**	**Smoke**	**Award**	**Exercise**	**TV**	**Height**	**Weight**	**Siblings**
2	FourthYear	M	No	Olympic	10	1	180	82	4
3	SecondYear	F	Yes	Academy	4	7	168	54	2
4	FirstYear	M	No	Nobel	14	5	183	94	2
5	ThirdYear	M	No	Nobel	3	1	160	50	1
6	SecondYear	F	No	Nobel	3	3	165	68	1
7	SecondYear	F	No	Nobel	5	4	165	52	2
8	FirstYear	F	No	Olympic	10	10	168	58	1
9	SecondYear	M	No	Olympic	13	8	188	107	1
10	FirstYear	F	No	Nobel	12	1	152	52	7
11	SecondYear	F	No	Olympic	12	6	165	64	1
12	ThirdYear	F	No	Nobel	6	1	173	61	2
13	FirstYear	F	No	Olympic	10	2	160	50	1

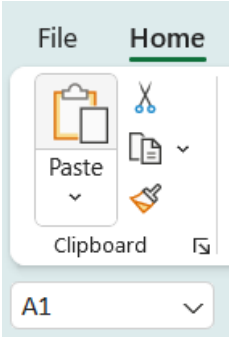
2. Right click on the selected range and click on "**Copy**".



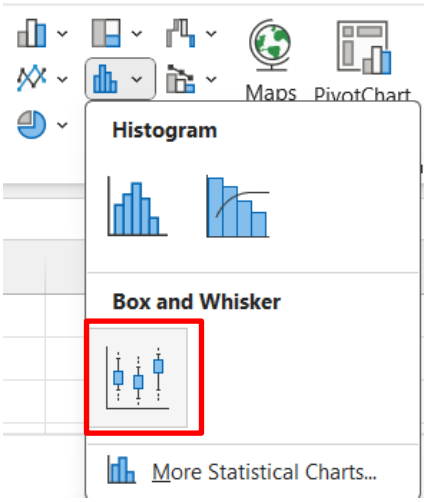
3. Create a new worksheet on the workbook with a suitable name. Click on the plus button on the right side of the current worksheet.



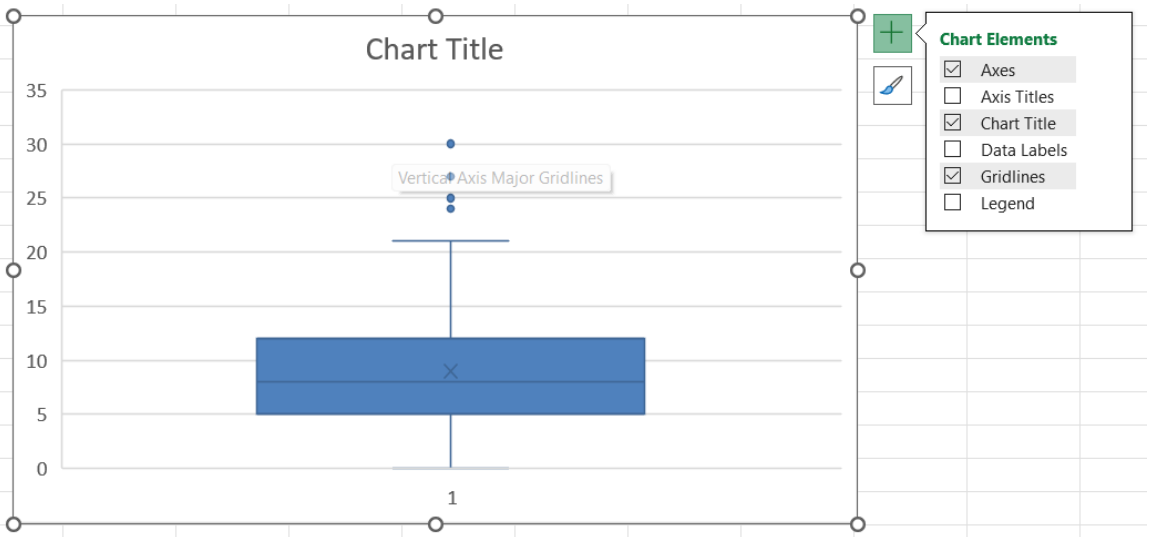
4. On the new worksheet, paste the previously copied data (as in step 2). To do this, click on a cell where the data needs to be pasted and click on the "**Paste**" button in the Home tab.



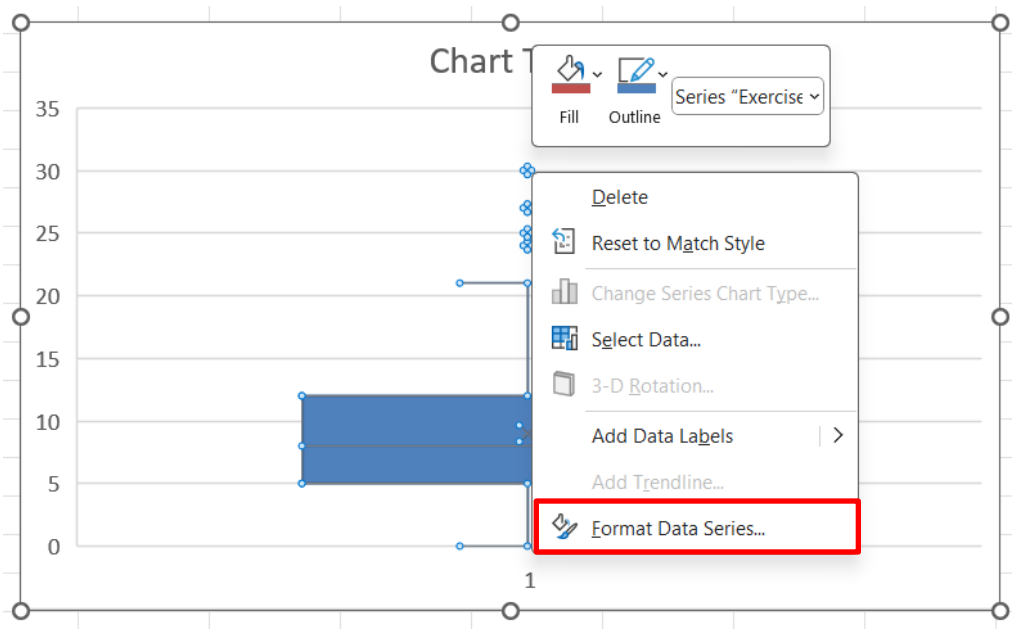
5. Select the data that needs to be plotted in the new worksheet. Navigate to the "**Insert**" tab in the Excel ribbon and click on the "**Box and Whisker**" charts under the Statistical charts group.



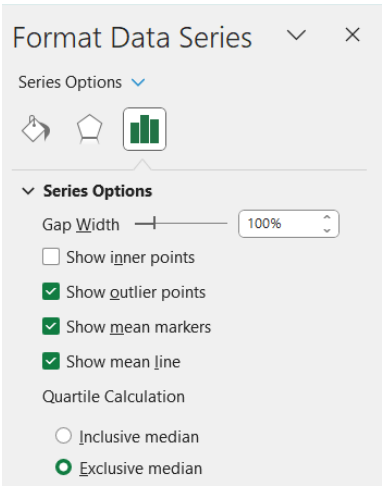
6. A Box and Whisker chart will be created as follows.



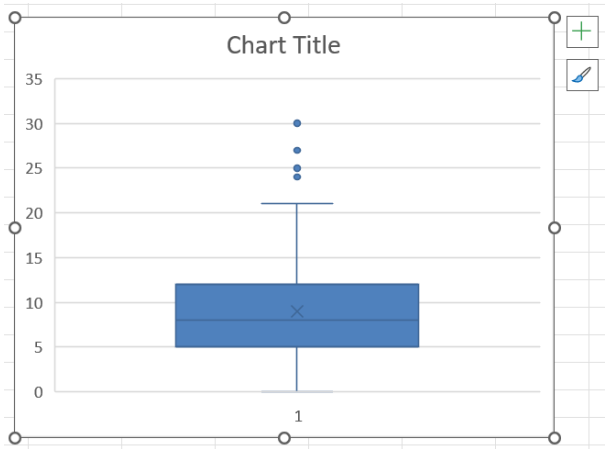
7. Right click on the box inside the chart and click on the "**Format Data Series**" option.



8. On the window that opens on the right side, choose the preferred options. Here, the "**Show mean line**" option is enabled, and all the other options are left as default.



9. Previous step returns the boxplot as below.

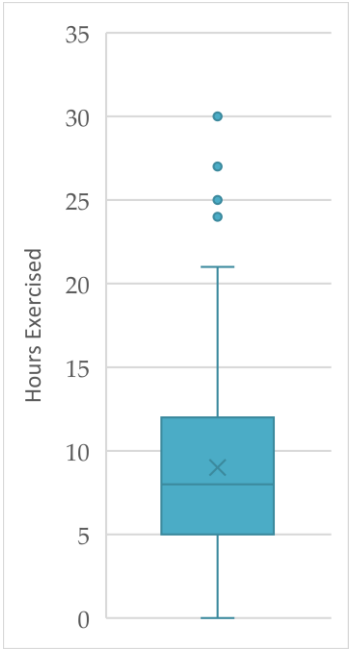


10. The user can now customise or format the boxplot using the "**Chart Elements**" button on the top right to,

a. Show or hide chart title or
b. Insert Axis labels.

11. Furthermore, the user can make changes to the design using the tools and templates provided under the "**Chart Design**" tab in the Excel ribbon.

12. After inserting appropriate axis titles and formatting, the boxplot showing the number of hours students exercised can be obtained as below.

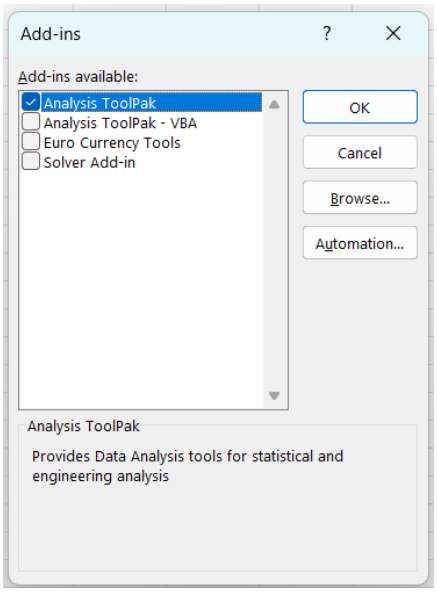


QUANTITATIVE DATA: NUMERICAL SUMMARIES

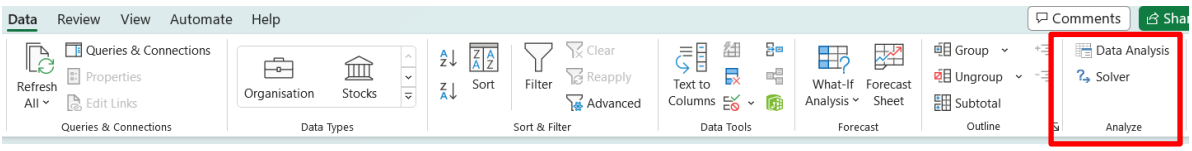
Numerical Measures

Most of the numerical analysis of the data can be easily done using the "**Data Analysis ToolPak**" in Excel. However, this is not a default feature available on Excel. It is, by default, a disabled Excel Add-in. To enable this Add-in,

1. Navigate to the "**File**" tab in the Excel ribbon, click "**Options**" and then click the "**Add-Ins**" category.
2. In the "**Manage**" box, select "**Excel Add-Ins**" and then click "**Go**".
3. In the "**Add-Ins**" box, check the "**Analysis ToolPak**" check box and then click "**OK**".



4. Once this is enabled, the "**Analyze**" group will appear under the "**Data**" tab in the Excel ribbon.

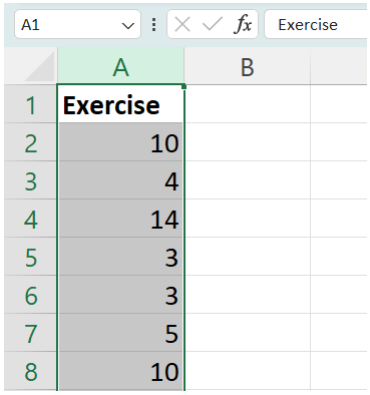


Now, that the "Data Analysis" feature is enabled, it can be used to carry out most of the numerical analysis.

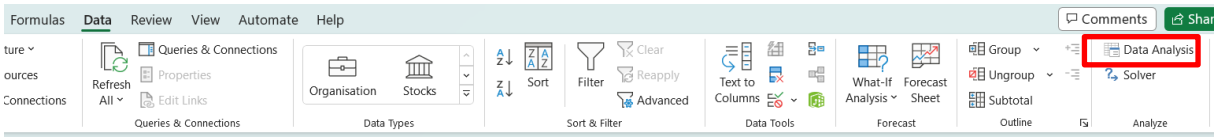
Question: Using the CleanedStudentSurvey Excel file, produce numerical summaries (measures of center, spread, location and shape) of the number of hours students surveyed exercised per week.

To start the analysis,

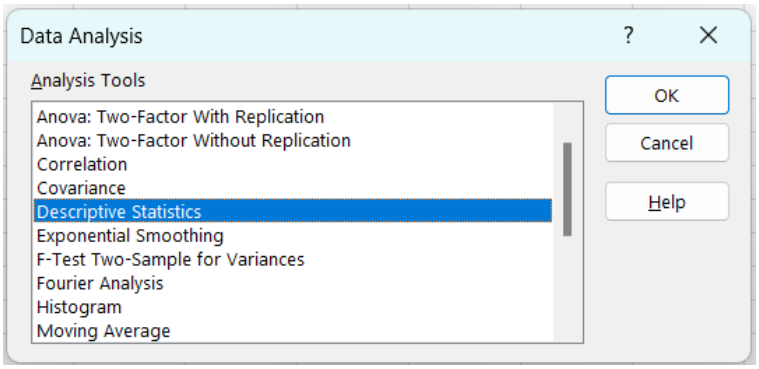
1. Select the data that needs to be analysed. In this demonstration, the number of hours students exercised per week data is used. So, the Column E that contains the data is copied and pasted on to a new worksheet. The steps to copy and paste data are provided in the previous section, Graphical Summaries: Boxplot, in steps 1-4.



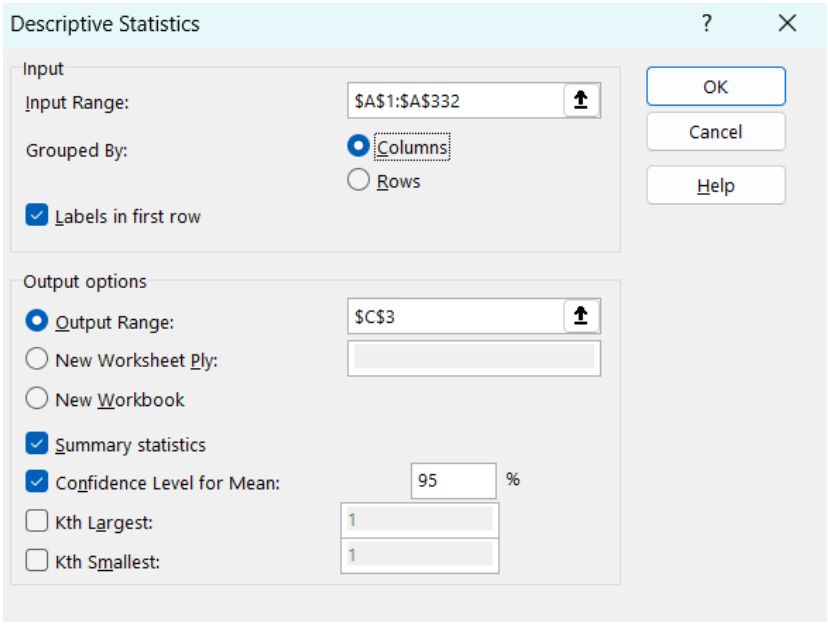
2. Once the data is pasted on to a new worksheet, navigate to the "**Data**" tab in the Excel ribbon. Click on the "**Data Analysis**" option under the "**Analyze**" group.



3. In the window that opens, select the "**Descriptive Statistics**" option and click "**OK**".



4. In the window that opens, select the input data range. In this case, it will be **$A$1:$A$332**. Enable the "**Labels in first row**" and "**Summary statistics**" options. Also, enable the "**Confidence Level for Mean**" and set the value to be 95%. Leave the other options as it is and click "**OK**".



5. The previous step will return the numerical summary that includes mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, sum, count, and the 95% confidence level of the data in the following form.

Exercise	
Mean	9.001510574
Standard Error	0.307021336
Median	8
Mode	5
Standard Deviation	5.585763626
Sample Variance	31.20075529
Kurtosis	0.602375095
Skewness	0.859014165
Range	30
Minimum	0
Maximum	30
Sum	2979.5
Count	331
Confidence Level(95.0%)	0.603965824

6. Note that the quartiles and interquartile range (IQR) are not included in the "**Data Analysis**" feature. One needs to calculate these parameters using some of the inbuilt functions of Excel.

7. To calculate the first quartile, click on a cell where the value needs to be placed. Enter the following formula in the cell to obtain the first quartile value.

=QUARTILE.INC(A2:A332,1)

8. To calculate the third quartile value, use the following formula in a nearby cell. As a good practice, remember to include the captions beside each of these quartile values.

=QUARTILE.INC(A2:A332,3)

9. Once both first and third quartiles are calculated, the interquartile range can be obtained by subtracting first quartile from the third quartile. In this demonstration, Q1 is placed in cell **D19** and Q3 is placed in cell **D20**. Hence, the formula to calculate the IQR is,

= D20-D19

10. The numerical summary obtained using the Data Analysis tool, Q1, Q3 and IQR can be placed in the following form.

	C	D
3	*Exercise*	
4		
5	Mean	9.001510574
6	Standard Error	0.307021336
7	Median	8
8	Mode	5
9	Standard Deviation	5.585763626
10	Sample Variance	31.20075529
11	Kurtosis	0.602375095
12	Skewness	0.859014165
13	Range	30
14	Minimum	0
15	Maximum	30
16	Sum	2979.5
17	Count	331
18	Confidence Level(95.0%)	0.603965824
19	Q1	5
20	Q3	12
21	IQR	7

11. To conclude, one can apply the "**Number**" format with 2 decimal places for all these numbers and obtain the final numerical summary of the number of hours students exercised as shown below.

Exercise	
Mean	9.00
Standard Error	0.31
Median	8.00
Mode	5.00
Standard Deviation	5.59
Sample Variance	31.20
Kurtosis	0.60
Skewness	0.86
Range	30.00
Minimum	0.00
Maximum	30.00
Sum	2979.50
Count	331.00
Confidence Level(95.0%)	0.60
Q1	5.00
Q3	12.00
IQR	7.00

BIVARIATE ANALYSIS OF QUANTITATIVE DATA

Question: Using the CleanedStudentSurvey Excel file, produce side-by-side boxplots illustrating the distribution of *Exercise* by *Award*.

One Categorical Variable and One Quantitative Variable

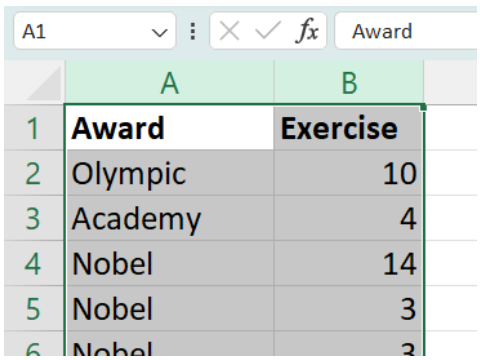
Side by Side Boxplot

To create the side-by-side boxplot, the following steps can be used.

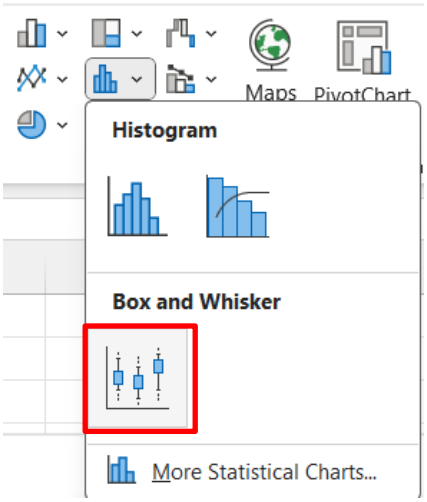
1. Select the data that is needed for the side-by-side boxplot. In this demonstration, the distribution of *Exercise* by *Award* is used. So, the Column D that contains the *Awards* data and the Column E that contains the *Exercise* data are selected.

C	D	E	F
Smoke	**Award**	**Exercise**	**TV**
No	Olympic	10	1
Yes	Academy	4	7
No	Nobel	14	5
No	Nobel	3	1
No	Nobel	3	3

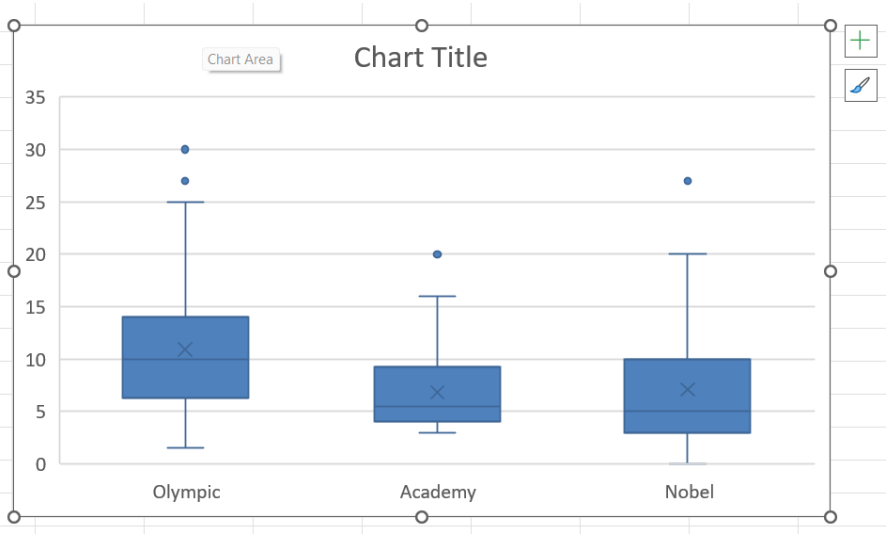
2. This data is copied and pasted on a new worksheet. The steps to copy and paste data are provided in steps 1-4 of the previous section, Graphical Summaries: Boxplot.
3. Select the pasted data.



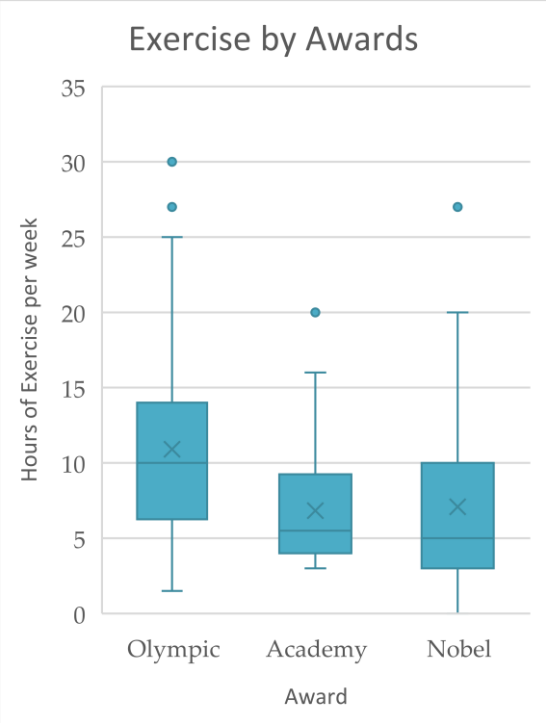
4. Navigate to the "**Insert**" tab in the Excel ribbon and click on the "**Box and Whisker**" charts under the Statistical charts group.



5. The following chart will be created on the current worksheet.



6. One can format the data series as mentioned in steps 7 and 8 in the Graphical Summaries: Boxplot section.
7. Furthermore, the user can customise or format the chart using the "**Chart Elements**" button on the top right to,
    a. Show or hide chart title or
    b. Insert Axis labels.
8. After inserting appropriate axis titles and formatting, the side-by-side boxplot showing the distribution of *Exercise* by *Decade* can be obtained as,

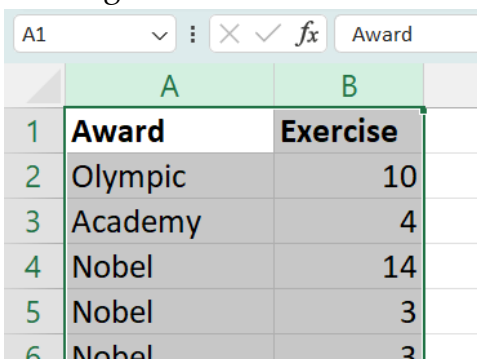

Exercise by Awards
Hours of Exercise per week
35
30
25
20
15
10
5
0
Olympic
Academy
Nobel
Award

Question: Using the CleanedStudentSurvey Excel file, produce bar chart of the means illustrating the distribution of Exercise by Award with error bars showing their standard deviation, standard error and 95% confidence intervals.

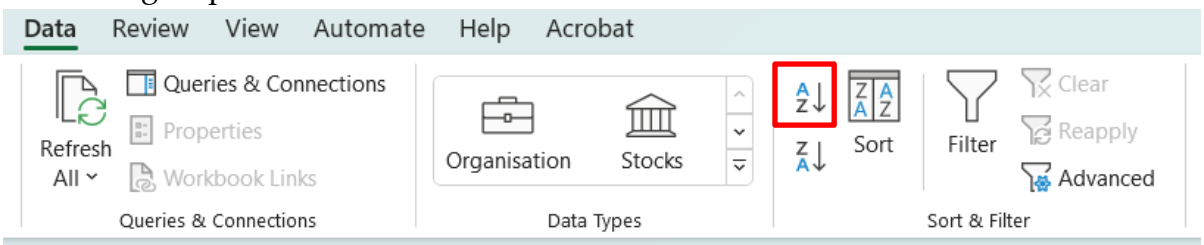
Bar Chart With Error Bars

To generate bar graph of mean with their standard deviation. standard error and 95% confidence intervals, one needs to carry out the numerical analysis on the data first. To do that,

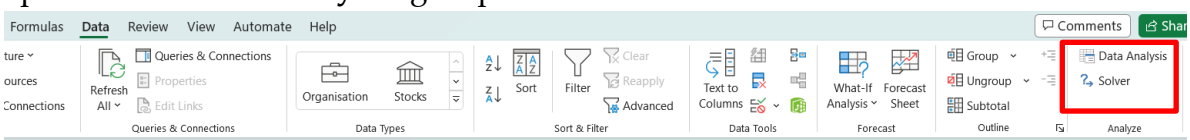
1. Copy and paste the *Exercise* and *Awards* data to a new worksheet. Follow steps 1-3 as provided in the previous section, Side by Side Boxplot). In this demonstration, the data is placed in the cell range **A1:B332**.



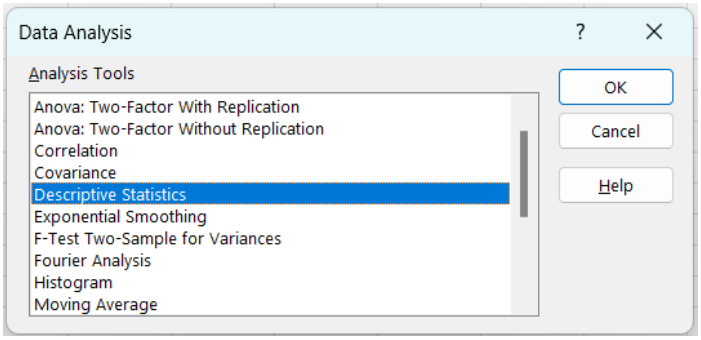
2. Sort the data based on the *Award* in the alphabetical order. To do this, navigate to the "**Data**" tab in the Excel ribbon and click on "**Sort A to Z**" icon under the "**Sort & Filter**" group.



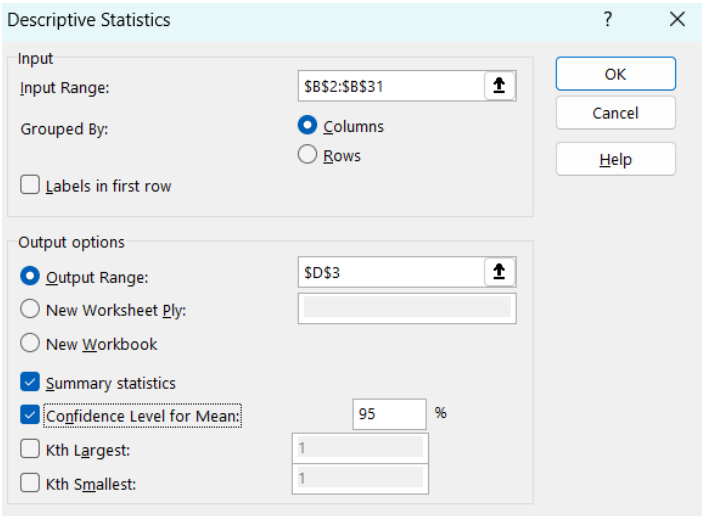
3. The data will be sorted in the order of *Academy Awards* followed by *Nobel* and *Olympic Awards*.
4. Now, navigate to the "**Data**" tab in the Excel ribbon. Click on the "**Data Analysis**" option under the "**Analyze**" group.



5. In the window that opens, select the "**Descriptive Statistics**" and click "**OK**".



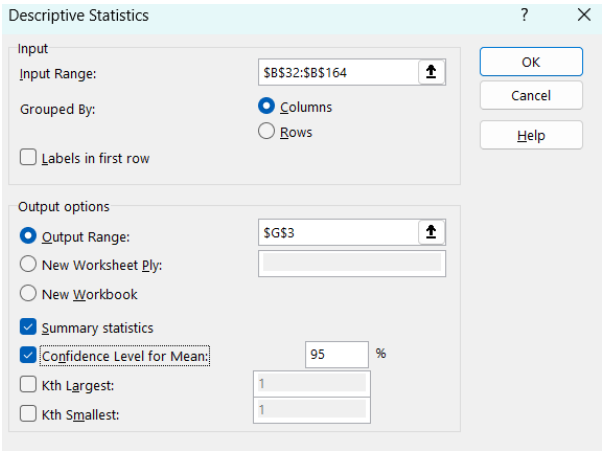
6. In the window that opens, select the input data range. In this case, one needs to carry out the analysis for individual awards separately. For *Academy Award*, select the input range **$B$2:$B$31**. Enable the "**Summary statistics**" option. Also, enable the "**Confidence Level for Mean**" and set the value to be 95%. Select the cell where the summary should be placed. In this example, **D3** is chosen as the preferred location. Leave the other options as it is and click "**OK**".



7. The output of the previous step will look like the following.

	D	E
3	*Column1*	
4		
5	Mean	6.833333333
6	Standard Error	0.758224383
7	Median	5.5
8	Mode	3
9	Standard Deviation	4.152965981
10	Sample Variance	17.24712644
11	Kurtosis	2.629948536
12	Skewness	1.590358126
13	Range	17
14	Minimum	3
15	Maximum	20
16	Sum	205
17	Count	30
18	Confidence Level(95.0%)	1.550742983

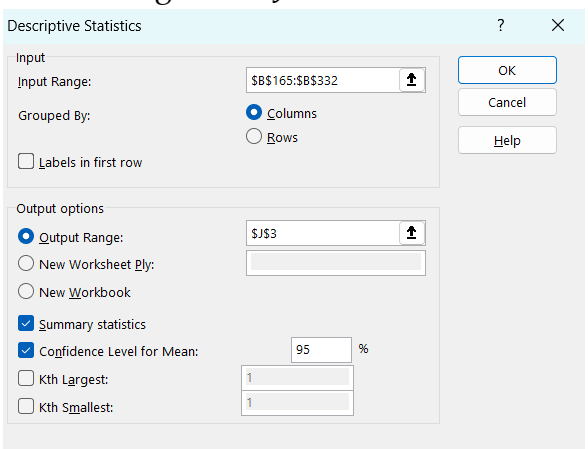
8. Similarly, one can obtain the summary for the *Nobel Award*. For *Nobel Award*, the input range will be **$B$32:$B$164**. In this example, the output range is set as **G3**. All other parameters remain the same as in step 4.



9. The output of the above step will be the following.

	G	H
3	*Column1*	
4		
5	Mean	7.082706767
6	Standard Error	0.425337456
7	Median	5
8	Mode	3
9	Standard Deviation	4.905230839
10	Sample Variance	24.06128959
11	Kurtosis	1.294643419
12	Skewness	1.089519593
13	Range	27
14	Minimum	0
15	Maximum	27
16	Sum	942
17	Count	133
18	Confidence Level(95.0%)	0.84135953

10. One can repeat the process as mentioned in step 6 or 8 to create the "**Summary statistics**" for Olympics Awards. The input range will be **$B$165:$B$332**. Also, one can place the table starting in cell **J3**.



11. The output for the above step will look like the following.

	J	K
3	*Column1*	
4		
5	Mean	10.9077381
6	Standard Error	0.437855316
7	Median	10
8	Mode	10
9	Standard Deviation	5.675253531
10	Sample Variance	32.20850264
11	Kurtosis	0.413651153
12	Skewness	0.668424819
13	Range	28.5
14	Minimum	1.5
15	Maximum	30
16	Sum	1832.5
17	Count	168
18	Confidence Level(95.0%)	0.864445033

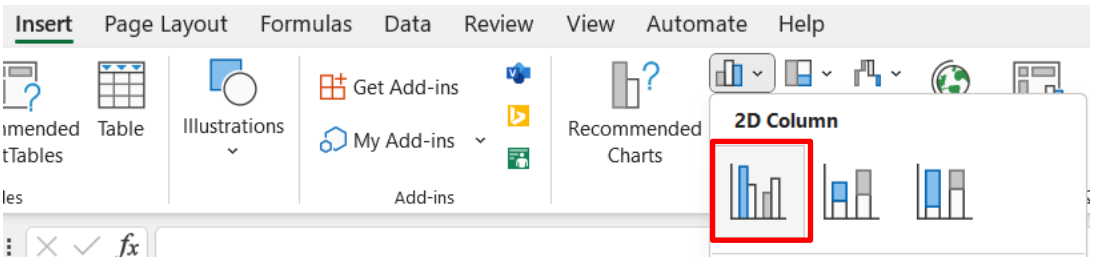
12. Now that the mean, standard deviation, standard error and 95% confidence intervals of the *Exercise* data are obtained for all the *Awards*, one can summarise these into the following form.

	M	N	O	P
4		Academy	Nobel	Olympic
5	Mean	6.83	7.08	10.91
6	Standard Deviation	4.15	4.91	5.68
7	Standard Error	0.76	0.43	0.44
8	95% Confidence Interval	1.55	0.84	0.86

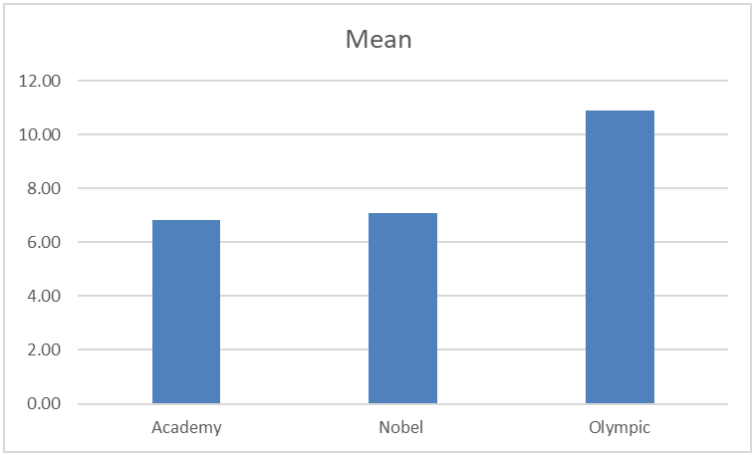
13. To insert the bar chart with of means, first select the following data.

	M	N	O	P
4		Academy	Nobel	Olympic
5	Mean	6.83	7.08	10.91
6	Standard Deviation	4.15	4.91	5.68
7	Standard Error	0.76	0.43	0.44
8	95% Confidence Interval	1.55	0.84	0.86

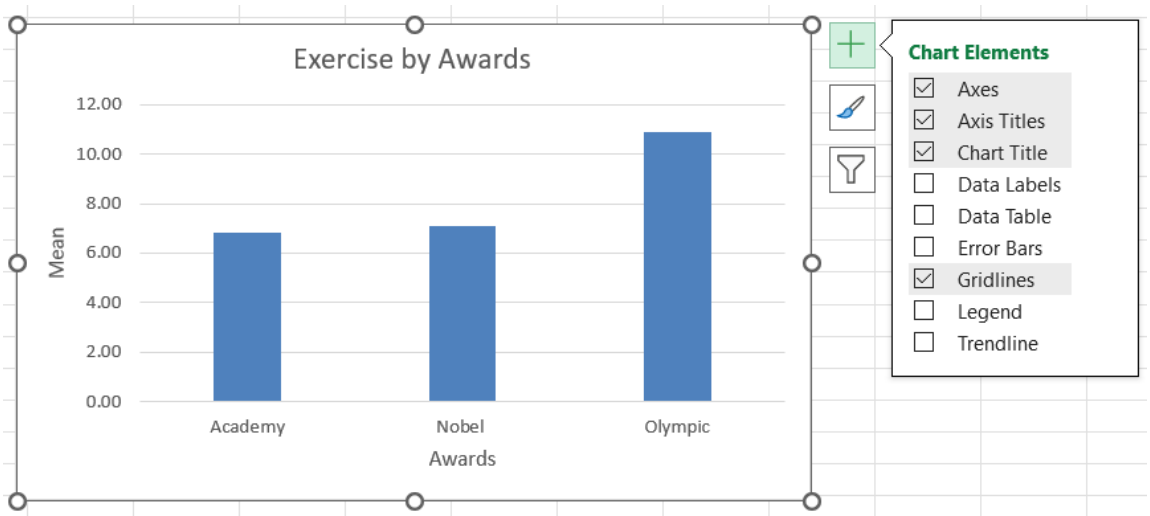
14. Navigate to the "**Insert**" tab in the Excel ribbon, then click on the "**2D Clustered Column**" chart under the "**Charts**" group.



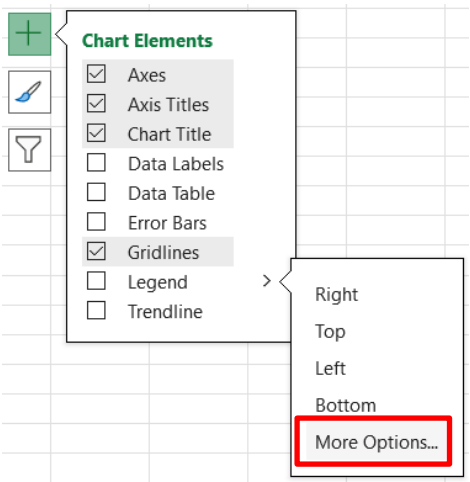
15. A simple 2D bar graph showing the mean values of *Exercise* by *Academic*, *Nobel* and *Olympic Awards* will be inserted as shown below.



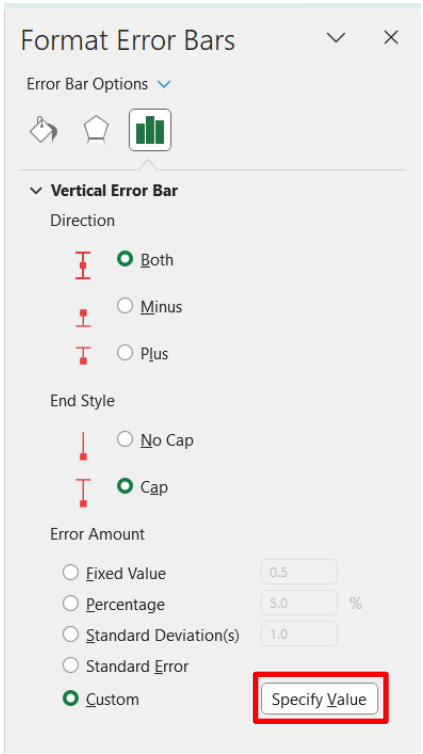
16. Insert appropriate chart title and axis titles to the chart as required using the "**Chart Elements**" icon on the top right of the chart.



17. To insert standard deviation as error bars in this chart, navigate to the "**Chart Elements**", click on the right arrow beside the "**Error Bars**" option and then click "**More Options**".

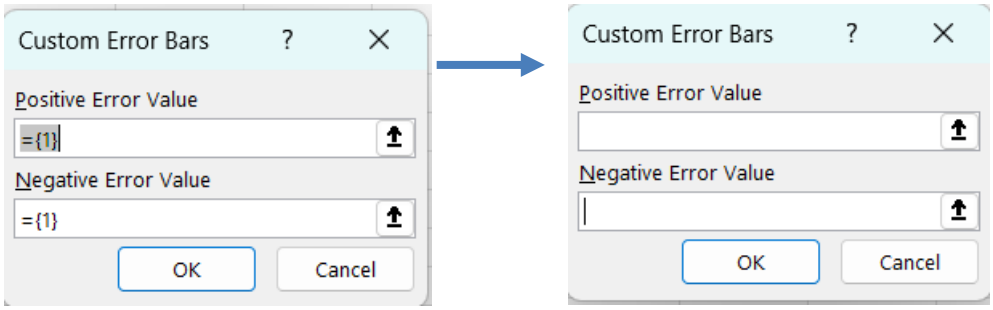


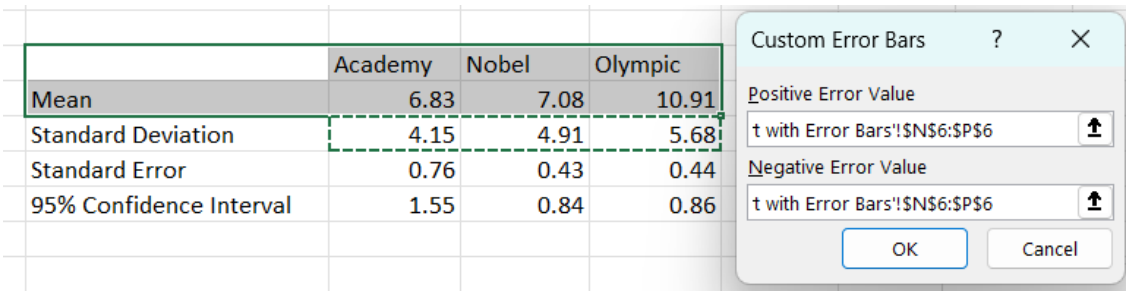
18. The previous step will open a window on the left with the title "**Format Error Bars**". Click on "**Custom**" option at the bottom and then click on "**Specify Value**".



19. In the new window that appears, clear the existing positive and negative error values.

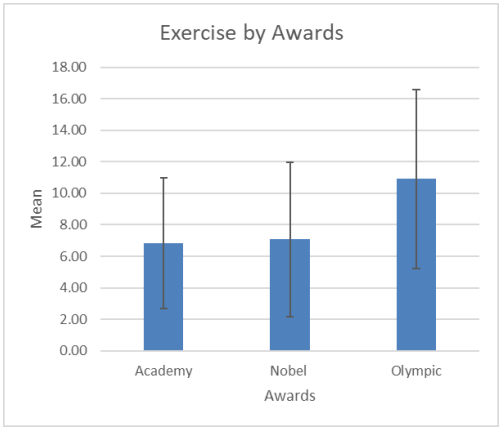
20. For both positive and negative error values, select the three columns with the standard deviation.



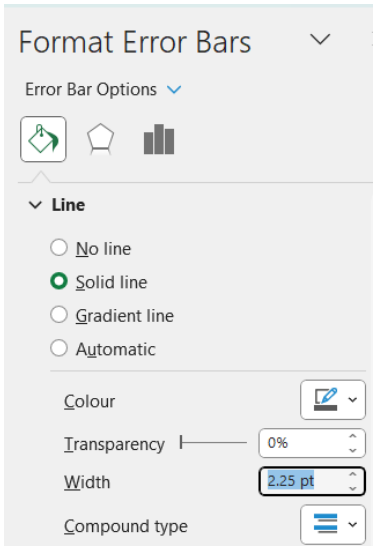


	Academy	Nobel	Olympic
Mean	6.83	7.08	10.91
Standard Deviation	4.15	4.91	5.68
Standard Error	0.76	0.43	0.44
95% Confidence Interval	1.55	0.84	0.86

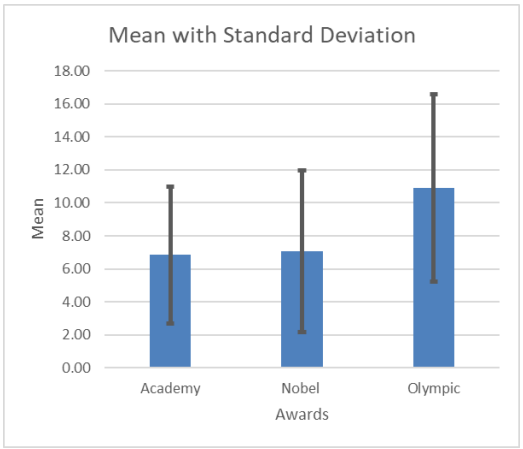
21. This will return the bar chart with error bars illustrating the standard deviation as shown below.



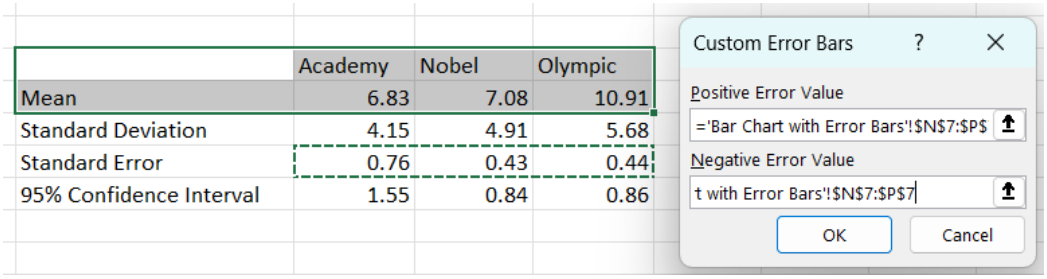
22. One can increase the width of the error bar to make it more visible using the option available in the "**Format Error Bars**" window on the right side.



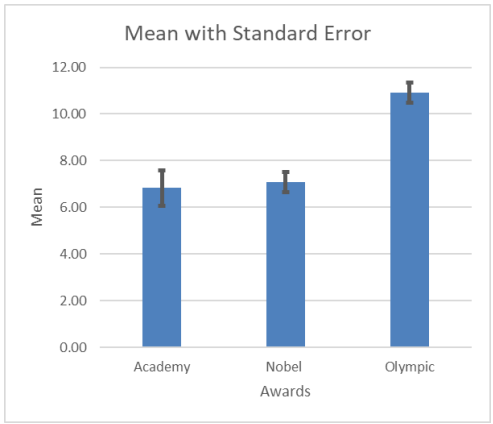
23. After setting the width to 2.25 pt, the bar chart will look like the following. The chart title can be changed to Mean with Standard Deviation.



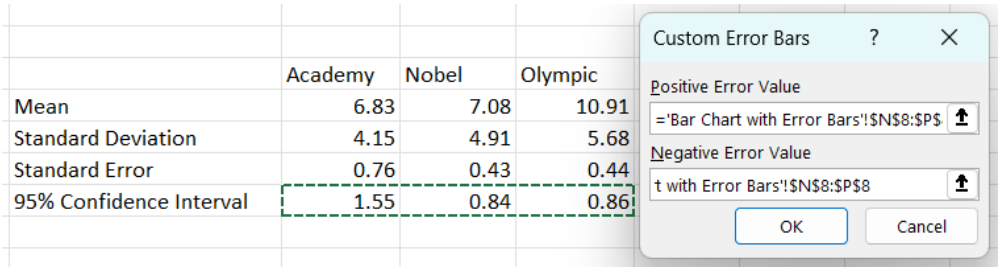
24. To generate the bar chart of mean values with error bars illustrating standard errors, repeat steps 13-19. In the "**Custom Error Bars**" window, select the standard error values instead of standard deviation as shown below.



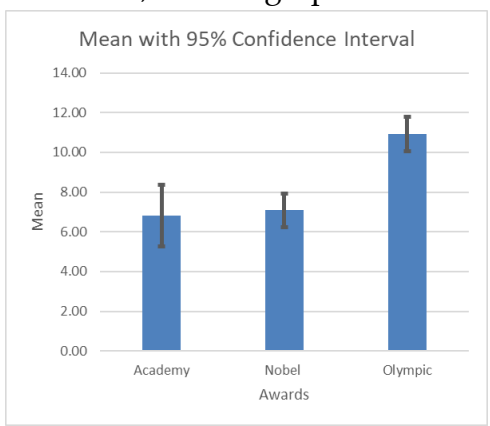
25. Click "**OK**" to return the Standard Error values as the error bars. After setting the error bar width and renaming the chart title, the bar graph of mean with standard error as error bars can be obtained as shown below.



26. To generate the bar chart of mean values with error bars illustrating 95% confidence intervals, repeat steps 13-19. In the "**Custom Error Bars**" window, select the 95% confidence interval values instead of standard deviation as shown below. Click "**OK**".



27. After setting the error bar width to 2.25 pt and renaming the chart to "**Mean with 95% Confidence Interval**", the bar graph will look like the following.



Two Quantitative Variables

Question: Using the CleanedStudentSurvey Excel file, create a scatterplot to display the relationship between *Height (in cm)* and *Weight in (kg)*. Also, obtain the correlation between *Height* and *Weight*.

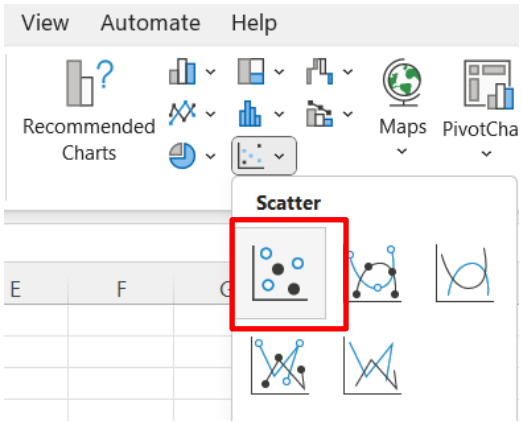
Scatterplot with Regression Line

To create the scatterplot between *Height (in cm)* and *Weight in (kg)*,

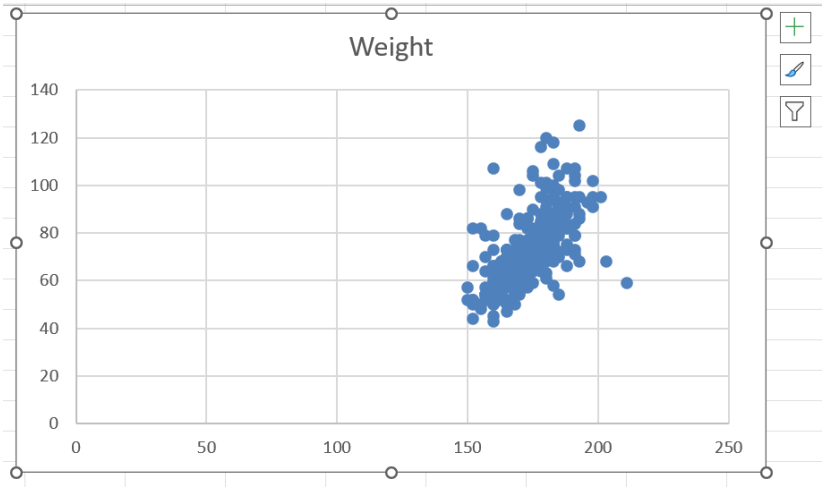
1. Select the data that needs to be plotted. In this demonstration, the *Height (in cm)* and *Weight in (kg)* data are used. Hence, Column G that contains the *Height* and Column H that contains the *Weight* data form the **StudentSurvey** worksheet are copied and pasted on to a new worksheet. The steps to copy and paste data are provided in the previous section, Graphical Summaries: Boxplot, in steps 1-4. Note that one must keep *Weight* in the first column and *Height* in the second column so that they will be plotted in the x and y-axis respectively.

	A	B
1	**Height**	**Weight**
2	180	82
3	168	54
4	183	94
5	160	50
6	165	68
7	165	52

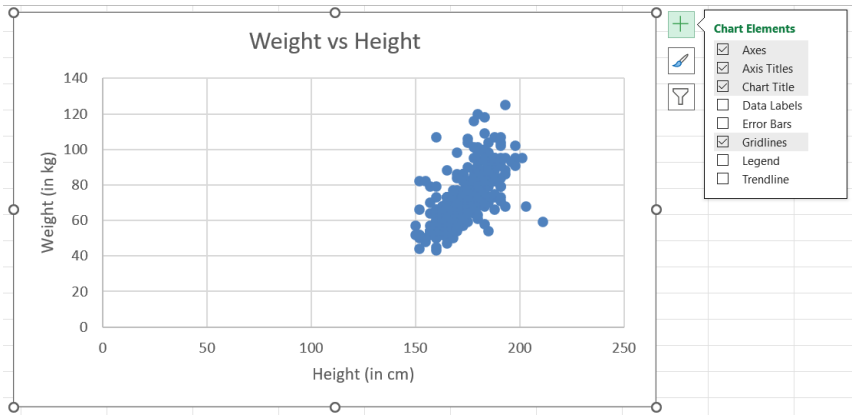
2. Select the data spread in the range **A1:B332** as shown above. Navigate to the "**Insert**" tab in the Excel ribbon and click on the "**Scatter**" chart.



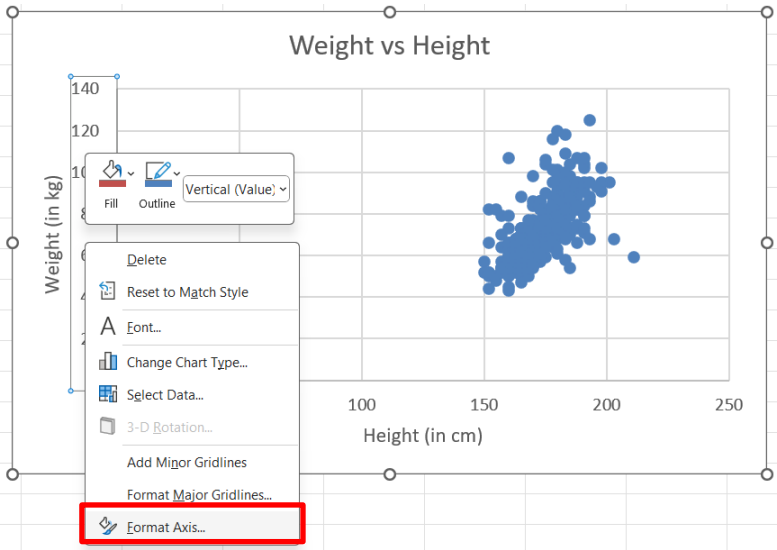
3. A scatter chart will be created as shown below.



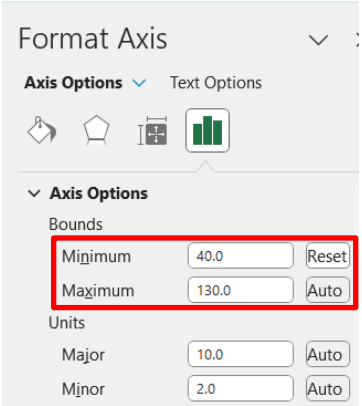
4. Using the "**Chart Elements**" icon on the top right of the chart, insert axis titles and rename the chart title as required. After these steps, the chart will look like the following.



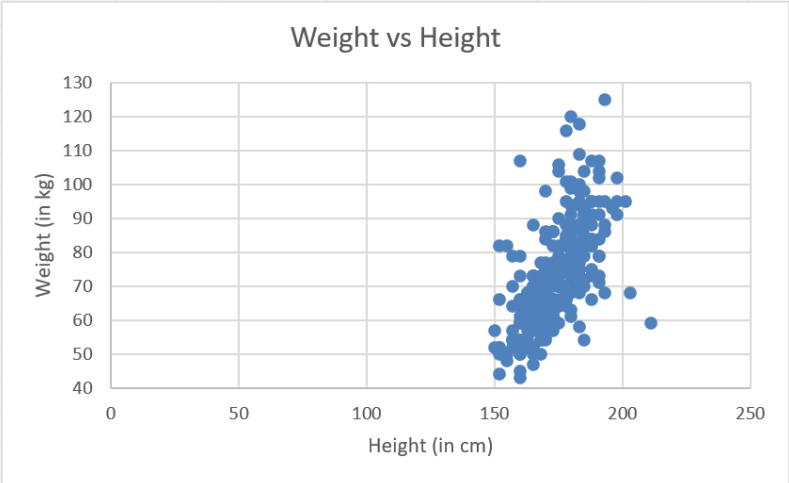
5. In the next step, one can rescale the axes to show only the relevant data points. To do this, right click on the Y-axis and then click on "**Format Axis**".



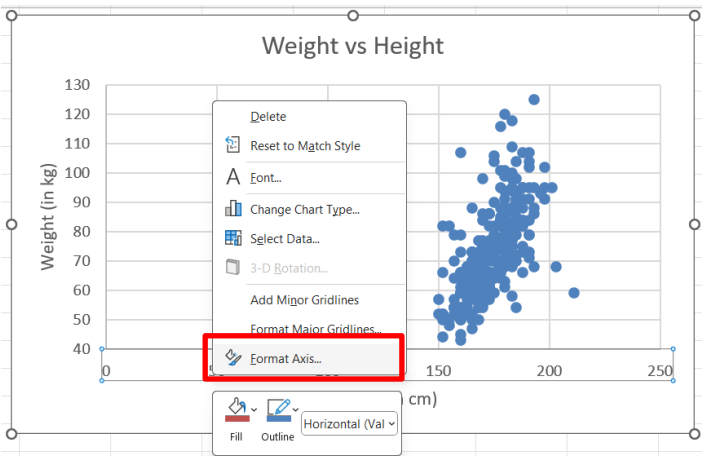
6. In the "**Format Axis**" window that appears on the right side, change the minimum and maximum values as required. For this example, the minimum can be set to 40 and the maximum to 130.



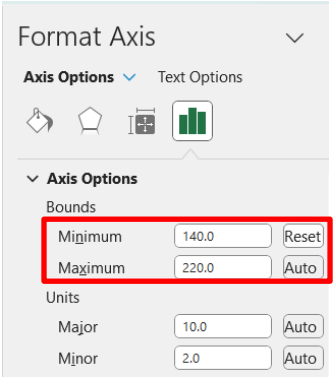
7. The above step will change the Y-axis values and the chart will be updated to the following form.



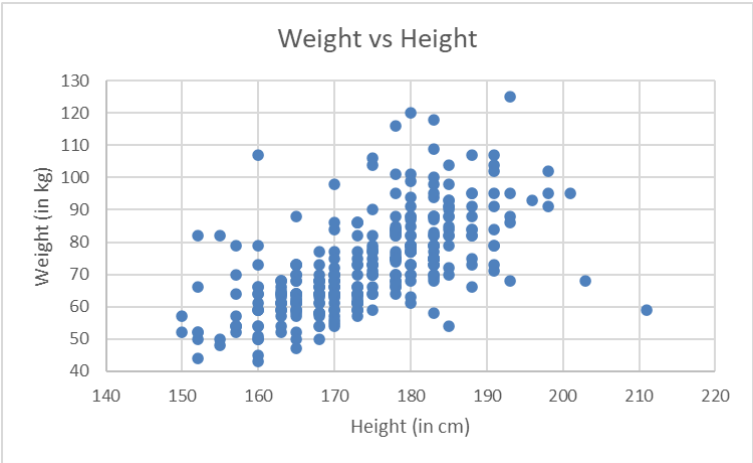
8. One can do the same with the X-axis scale as well. Right on the X-axis and click on "**Format Axis**".



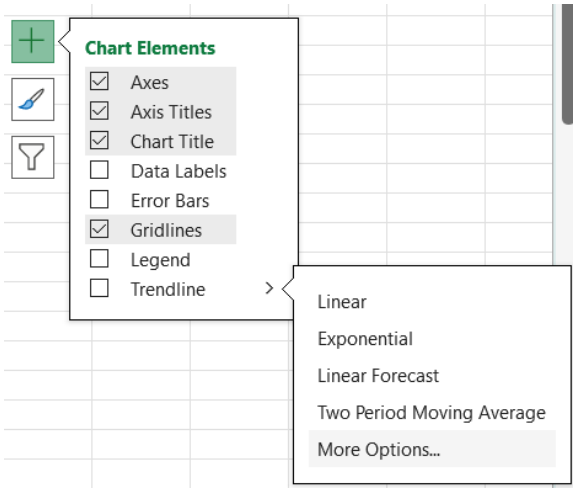
9. In the "**Format Axis**" window, the minimum and maximum values for the X-axis can be set. In this example, the minimum and maximum values are set to 140 and 220 respectively.



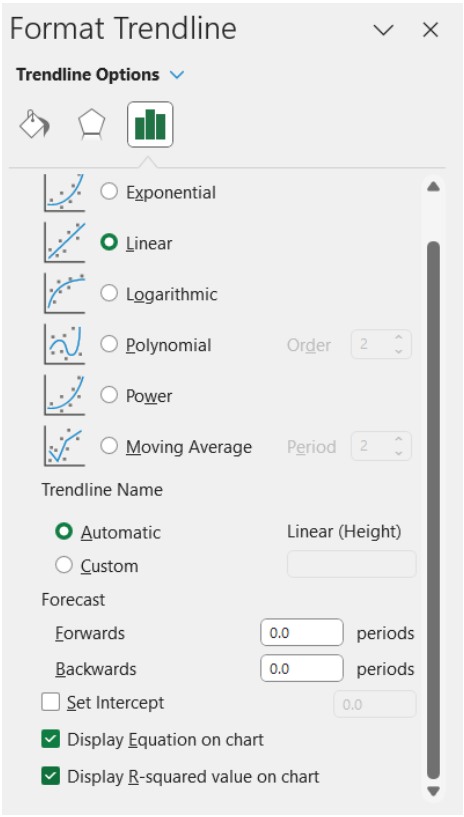
10. The above step will return the chart as shown below.



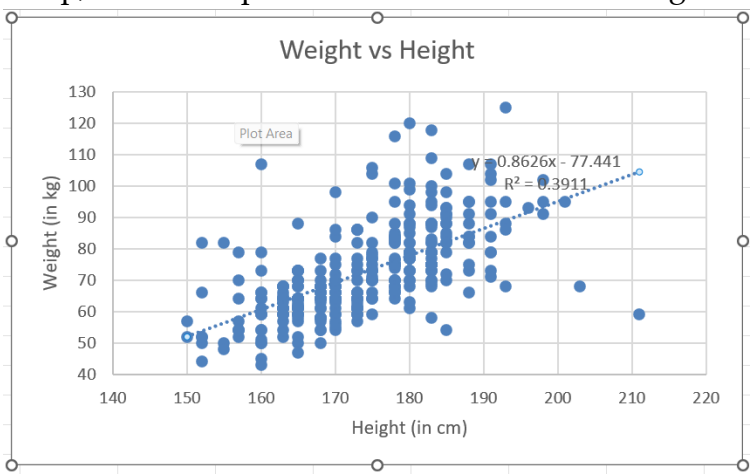
11. The '*line of best fit*' also known as "**Trendline**" in Excel can be inserted by using the "**Chart Elements**" icon on the top right of the chart. Click on the right arrow beside the "**Trendline**" option and then click on "**More Options**".



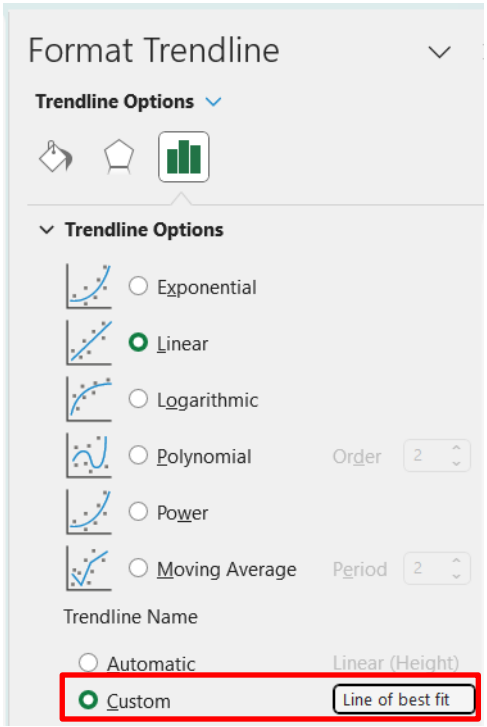
12. In the "**Format Trendline**" window that appears on the right side, select "**Linear**". Furthermore, enable the "**Display Equation on chart**" and the "**Display R-squared value on chart**" options. The chart will be updated with the "line of best fit", equation of that line and R^2 value.



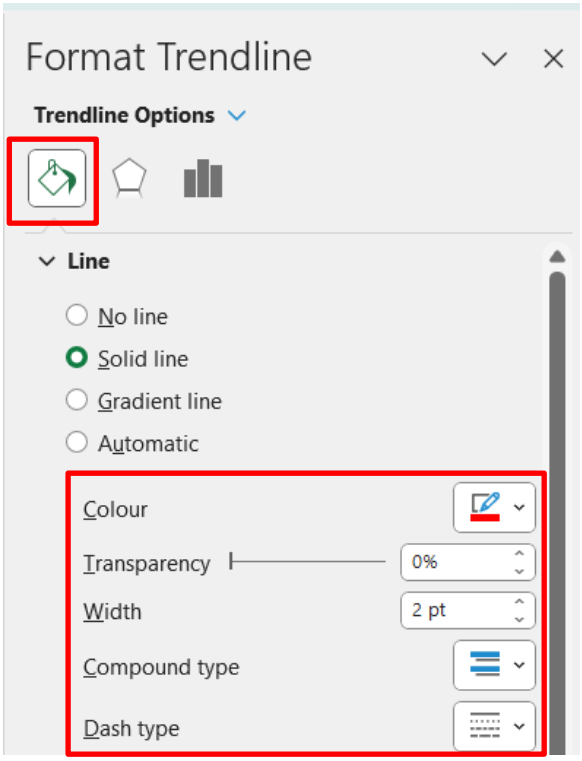
13. After the above step, the scatterplot will look like the following.



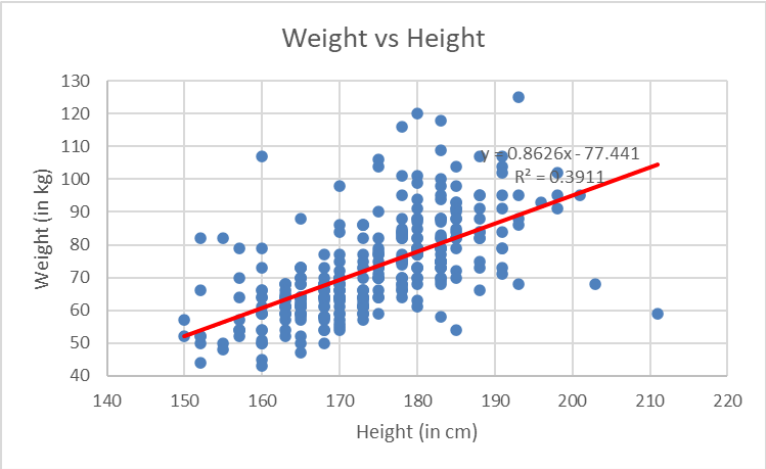
14. Furthermore, the trendline can be given a custom name by clicking on "**Custom**" button and filling in an appropriate name.



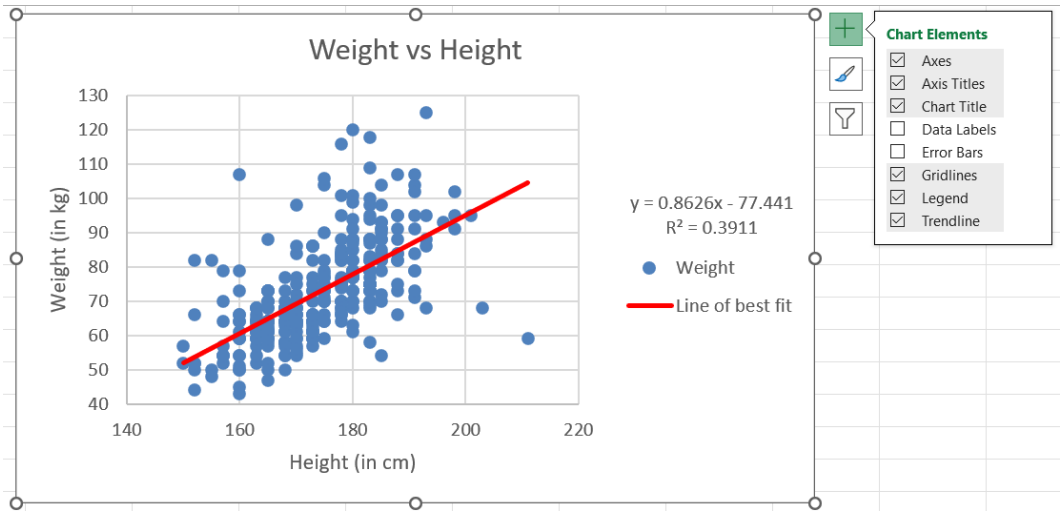
15. One can also change the appearance of the trendline by clicking on the paint bucket icon on the top of the "**Format Trendline**" window.



16. The above steps will return the trendline with equation and R^2 as shown below.



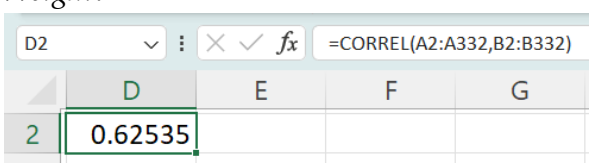
17. As the chart now contains both data points and the trendline, one can enable the legend to describe them. For this, click on the "**Chart Elements**" and enable the "**Legend**" option.



18. The final chart scatterplot with the "line of best fit" will look like the above.
19. To find the correlation between *Height* and *Weight*, one needs to use the inbuilt **CORREL** function in Excel. For this data, enter **=CORREL()** in one of the cells outside the data point (say D2). Then, select the first dataset (*Height*) and the second dataset (*Weight*) separated by a comma and then close the bracket. The function will look like,

=CORREL(A2:A332,B2:B332)

This function will return the correlation value as 0.62535 for the relationship between *Height* and *Weight*.



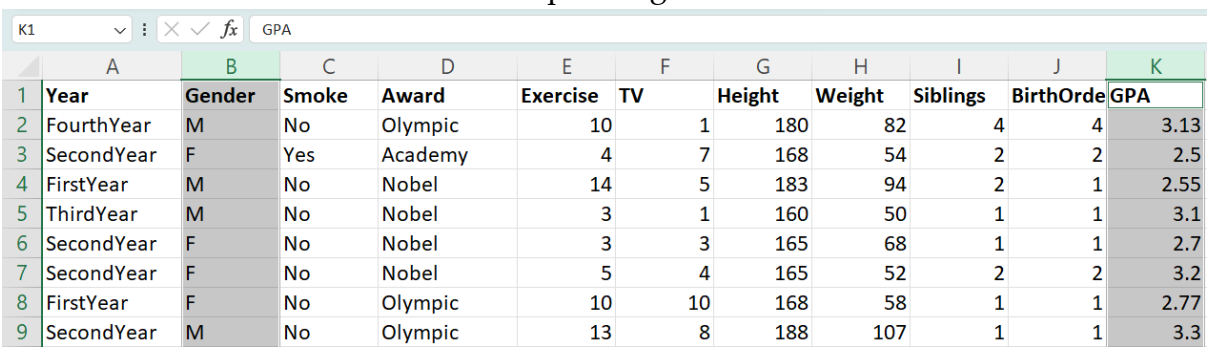
STATISTICAL INFERENCE

GENERATING RANDOM SAMPLES

Question: Select a simple random sample of 50 cases from Male and 50 cases from Female (Making a total of 100 cases).

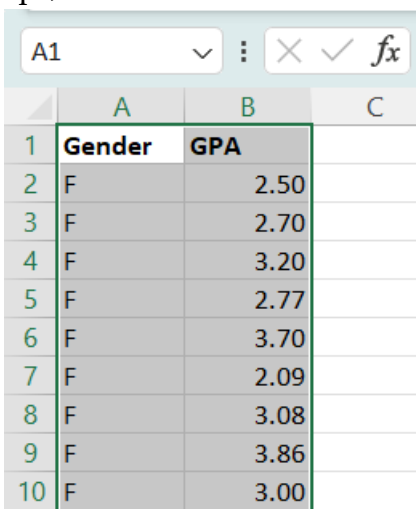
Using INDEX and RANDBETWEEN functions

1. Select the columns with the data that needs to be sampled. In this case, the columns that contains the *Gender* and the corresponding *GPA* data are selected.



	A	B	C	D	E	F	G	H	I	J	K
1	Year	Gender	Smoke	Award	Exercise	TV	Height	Weight	Siblings	BirthOrde	GPA
2	FourthYear	M	No	Olympic	10	1	180	82	4	4	3.13
3	SecondYear	F	Yes	Academy	4	7	168	54	2	2	2.5
4	FirstYear	M	No	Nobel	14	5	183	94	2	1	2.55
5	ThirdYear	M	No	Nobel	3	1	160	50	1	1	3.1
6	SecondYear	F	No	Nobel	3	3	165	68	1	1	2.7
7	SecondYear	F	No	Nobel	5	4	165	52	2	2	3.2
8	FirstYear	F	No	Olympic	10	10	168	58	1	1	2.77
9	SecondYear	M	No	Olympic	13	8	188	107	1	1	3.3

2. Copy and paste the data on to a new worksheet and rename the worksheet with a suitable name. Data is placed in the range **A1:B332** in this case. Also, the data is sorted in the alphabetical order to collect all the *GPA*s that correspond female students and then male students. Important: To reproduce the functions that follows in the upcoming steps, make sure that the data is placed in the same range.



	A	B	C
1	Gender	GPA	
2	F	2.50	
3	F	2.70	
4	F	3.20	
5	F	2.77	
6	F	3.70	
7	F	2.09	
8	F	3.08	
9	F	3.86	
10	F	3.00	

3. In column **D**, insert serial numbers from 1 to 50. Preferably, one may give a title as Serial Number and the numbers can be inserted in the rows that follow. In this case, in cell **D2**, the title is included and, in the range, **D3:D52**, the serial numbers are inserted.

	D
1	
2	Sl. No
3	1
4	2
5	3
6	4
7	5
8	6

4. One can also include meaningful titles to include the 25 selections each for *Female* and *Male* students as follows. In this demonstration, a title *Female* is inserted in cell **E1**. Below that 25 selections are inserted in the range **E2:AC2**. Following this, a title *Male* is inserted in cell **AD1** and the 25 selections corresponding to *Male* students' GPA are listed in the range **AD2:BB2**.

	D	E	F	G	H	I	J
1		Female					
2	Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5	Selection 6
3	1						
4	2						
5	3						

5. In cell **C3**, insert the following function.

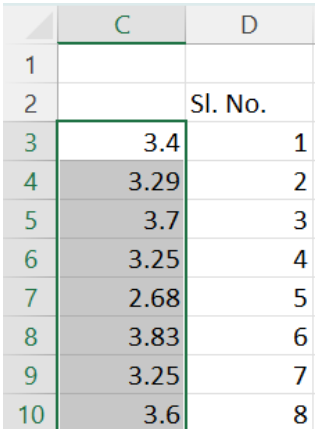
   =INDEX($B$2:$B$155,RANDBETWEEN(1,COUNTA($B$2:$B$155)),1)

6. The above function will insert a number randomly from the *Female* students *GPA* data listed in the range **B2:B155**.

C3 =INDEX($B$2:$B$155,RANDBETWEEN(1,COUNTA($B$2:$B$155)),1)

	A	B	C	D	E	F	G	H
1	**Gender**	**GPA**			Female			
2	F	2.50		Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4
3	F	2.70	3.5	1				

7. Use the AutoFill option to copy this function to the following 49 rows. This will return 50 random numbers, including the first number you picked in the previous step, from the set of 154 data points of *GPAs* that corresponds to *Female* students.



	C	D
1		
2		Sl. No.
3	3.4	1
4	3.29	2
5	3.7	3
6	3.25	4
7	2.68	5
8	3.83	6
9	3.25	7
10	3.6	8

The INDEX function in Excel returns a cell value based on the intersection of the row and column number of the range. In Excel's definition, the INDEX function has the following syntax.

=INDEX(array, row_num, [column_num])

In simple words, the first argument inside the bracket tells Excel where to look for the value. The second and third arguments gives the row and column numbers at which the intersection occurs.

=INDEX(where to look for, row number, [column number])

The RANDBETWEEN function generates a random number between a starting number and an ending number. For example, we need to select a random number between 1 and 48 that contains the set of all data points that corresponds to *Decade 1*.

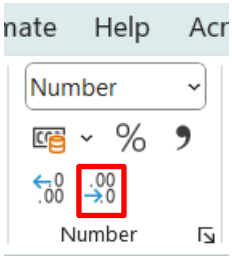
=RANDBETWEEN(bottom, top)

The COUNTA function counts the number of cells that are not empty in a range. So, if the number of data points for Decade 1 is not known, one may use the COUNTA function to count the number of points in the range and input that as the second argument for the RANDBETWEEN function. The number returned by the COUNTA function will act as the ending number for the RANDBETWEEN function.

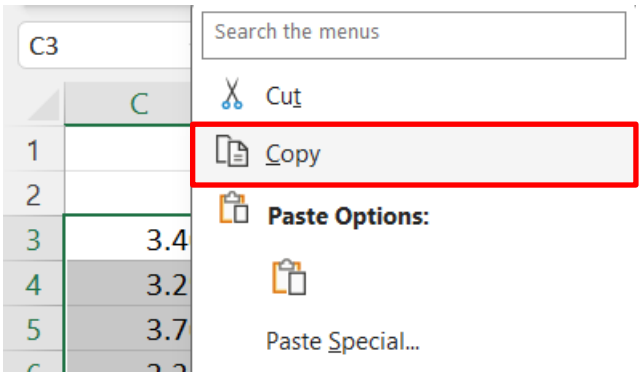
One may then combine these three functions to generate a random number from a set of data points as,

=INDEX(data range to look for, RANDBETWEEN(1, COUNTA(data range to look for)), 1)

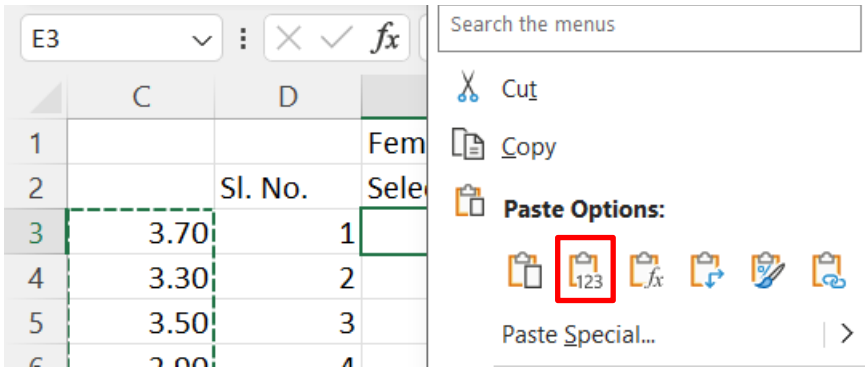
8. The number of decimal places can be reduced to 2 by changing the format of these cells to Number and then clicking on the Decrease decimal places button.



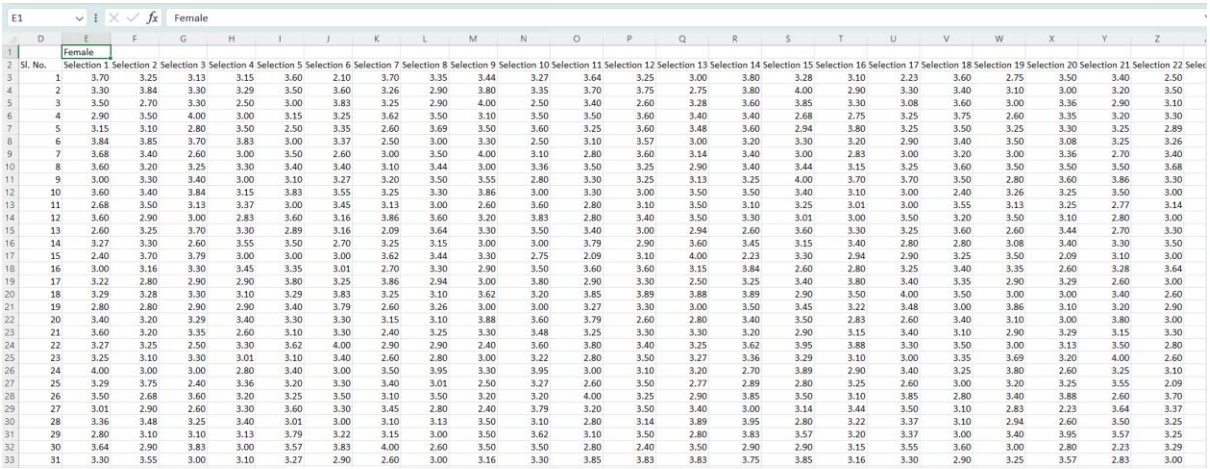
9. Once 50 random selections have been made, you may copy and paste these values on to 25 columns that were already created to keep the *GPA* data of *Female* Students. However, one should always remember to copy and paste the values alone, as the random selection changes after every operation done in that worksheet.
10. To copy and paste the values, select the 30 randomly selected data points. Right click and click on copy.



11. After copying the date, navigate to cell **E3**. Right click and click on Paste values.



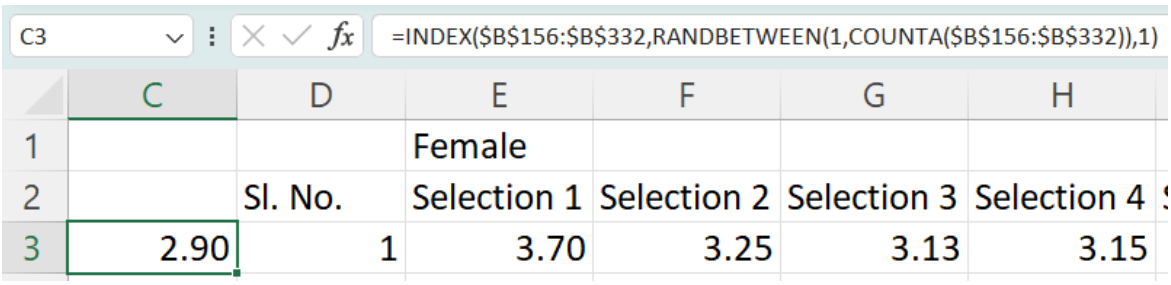
12. It can be noted that, after pasting the values, the data randomly selected and inserted in the range **C3:C52** changes. One can keep pasting the values the values under Selection 2 to Selection 25 columns by just right clicking on the first cell in the corresponding column (just below the title) and repeating the above step.
13. Once completed, one may obtain the sample data as follows:



		Female																						
Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5	Selection 6	Selection 7	Selection 8	Selection 9	Selection 10	Selection 11	Selection 12	Selection 13	Selection 14	Selection 15	Selection 16	Selection 17	Selection 18	Selection 19	Selection 20	Selection 21	Selection 22	Selec
1	3.70	3.25	3.13	3.15	3.60	2.10	3.70	3.35	3.44	3.27	3.64	3.25	3.00	3.80	3.28	3.10	2.23	3.60	2.75	3.50	3.40	2.50	
2	3.30	3.84	3.30	3.29	3.50	3.60	3.26	2.90	3.80	3.35	3.70	3.75	2.75	3.80	4.00	2.90	3.30	3.40	3.10	3.00	3.20	3.50	
3	3.50	2.70	3.30	2.50	3.00	3.83	3.25	2.90	4.00	2.50	3.40	2.60	3.28	3.60	3.85	3.30	3.08	3.60	3.00	3.36	2.90	3.10	
4	2.90	3.50	4.00	3.00	3.15	3.25	3.62	3.50	3.10	3.50	3.50	3.60	3.40	3.40	2.68	2.75	3.25	3.75	2.60	3.35	3.20	3.30	
5	3.15	3.10	2.80	3.50	2.50	3.35	2.60	3.69	3.50	3.60	3.25	3.60	3.48	3.60	2.94	3.80	3.25	3.50	3.25	3.30	3.25	2.89	
6	3.84	3.85	3.70	3.83	3.00	3.37	2.50	3.00	3.30	2.50	3.10	3.57	3.00	3.20	3.30	3.20	2.90	3.40	3.50	3.08	3.25	3.26	
7	3.68	3.40	2.60	3.00	3.50	2.60	3.00	3.50	4.00	3.10	2.80	3.60	3.14	3.40	3.00	2.83	3.00	3.20	3.00	3.36	2.70	3.40	
8	3.60	3.20	3.25	3.30	3.40	3.40	3.10	3.44	3.00	3.36	3.50	3.25	2.90	3.40	3.44	3.15	3.25	3.60	3.50	3.50	3.50	3.68	
9	3.00	3.30	3.40	3.00	3.10	3.27	3.20	3.50	3.55	2.80	3.30	3.25	3.13	3.25	4.00	3.70	3.70	3.50	2.80	3.60	3.86	3.30	
10	3.60	3.40	3.84	3.15	3.83	3.55	3.25	3.30	3.86	3.00	3.30	3.00	3.50	3.50	3.40	3.10	3.00	2.40	3.26	3.25	3.50	3.00	
11	2.68	3.50	3.13	3.37	3.00	3.45	3.13	3.00	2.60	3.60	2.80	3.10	3.50	3.10	3.25	3.01	3.00	3.55	3.13	3.25	2.77	3.14	
12	3.60	2.90	3.00	2.83	3.60	3.16	3.86	3.60	3.20	3.83	2.80	3.40	3.50	3.30	3.01	3.00	3.50	3.20	3.50	3.10	2.80	3.00	
13	2.60	3.25	3.70	3.30	2.89	3.16	2.09	3.64	3.30	3.50	3.40	3.00	2.94	2.60	3.60	3.30	3.25	3.60	2.60	3.44	2.70	3.30	
14	3.27	3.30	2.60	3.55	3.50	2.70	3.25	3.15	3.00	3.00	3.79	2.90	3.60	3.45	3.15	3.40	2.80	2.80	3.08	3.40	3.30	3.50	
15	2.40	3.70	3.79	3.00	3.00	3.00	3.62	3.44	3.30	2.75	2.09	3.10	4.00	2.23	3.30	2.94	2.90	3.25	3.50	2.09	3.10	3.00	
16	3.00	3.16	3.30	3.45	3.35	3.01	2.70	3.30	2.90	3.50	3.60	3.60	3.15	3.84	2.60	2.80	3.25	3.40	3.35	2.60	3.28	3.64	
17	3.22	2.80	2.90	2.90	3.80	3.25	3.86	2.94	3.00	3.80	2.90	3.30	2.50	3.25	3.40	3.80	3.40	3.35	2.90	3.29	2.60	3.00	
18	3.29	3.28	3.30	3.10	3.29	3.83	3.25	3.10	3.62	3.20	3.85	3.89	3.88	3.89	2.90	3.50	4.00	3.50	3.00	3.00	3.40	2.60	
19	2.80	2.80	2.90	2.90	3.40	3.79	2.60	3.26	3.00	3.00	3.27	3.30	3.00	3.50	3.45	3.22	3.48	3.00	3.86	3.10	3.20	2.90	
20	3.40	3.20	3.29	3.40	3.30	3.30	3.15	3.10	3.88	3.60	3.79	2.60	2.80	3.40	3.50	2.83	2.60	3.40	3.10	3.00	3.80	3.00	
21	3.60	3.20	3.35	2.60	3.10	3.30	2.40	3.25	3.30	3.48	3.25	3.30	3.30	3.20	2.90	3.15	3.40	3.10	2.90	3.29	3.15	3.30	
22	3.27	3.25	2.50	3.30	3.62	4.00	2.90	2.90	2.40	3.60	3.80	3.40	3.25	3.62	3.95	3.88	3.30	3.50	3.00	3.13	3.50	2.80	
23	3.25	3.10	3.30	3.01	3.10	3.40	2.60	2.80	3.00	3.22	2.80	3.50	3.27	3.36	3.29	3.10	3.00	3.35	3.69	3.20	4.00	2.60	
24	4.00	3.00	3.00	2.80	3.40	3.00	3.50	3.95	3.30	3.95	3.00	3.10	3.20	2.70	3.89	2.90	3.40	3.25	3.80	2.60	3.25	3.10	
25	3.29	3.75	2.40	3.36	3.20	3.30	3.40	3.01	2.50	3.27	2.60	3.50	2.77	2.89	2.80	3.25	2.60	3.00	3.20	3.25	3.55	2.09	
26	3.50	2.68	3.60	3.20	3.25	3.50	3.10	3.50	3.20	3.20	4.00	3.25	2.90	3.85	3.50	3.10	3.85	2.80	3.40	3.88	2.60	3.70	
27	3.01	2.90	2.60	3.30	3.60	3.30	3.45	2.80	2.40	3.79	3.20	3.50	3.40	3.00	3.14	3.44	3.50	3.10	2.83	2.23	3.64	3.37	
28	3.36	3.48	3.25	3.40	3.01	3.00	3.10	3.13	3.50	3.10	2.80	3.14	3.89	3.95	2.80	3.22	3.37	3.10	2.94	2.60	3.50	3.25	
29	2.80	3.10	3.10	3.13	3.79	3.22	3.15	3.00	3.50	3.62	3.10	3.50	2.80	3.83	3.57	3.20	3.37	3.00	3.40	3.95	3.57	3.25	
30	3.64	2.90	3.83	3.00	3.57	3.83	4.00	2.60	3.50	3.50	2.80	2.40	3.50	2.90	2.90	3.15	3.55	3.60	3.00	2.80	2.23	3.29	
31	3.30	3.55	3.00	3.10	3.27	2.90	2.60	3.00	3.16	3.30	3.85	3.83	3.83	3.75	3.85	3.16	3.30	2.90	3.25	3.57	2.83	3.00	

14. Note that the numbers selected may differ between the users. However, the table will look like the above.
15. Once the random selections for *Female* students' *GPA* is made successfully, one may proceed to the random selection of data points for *Male* students.
16. The easiest way to do this will be to delete all the data points in the range **C3:C52** and insert the following function in cell **C3**.

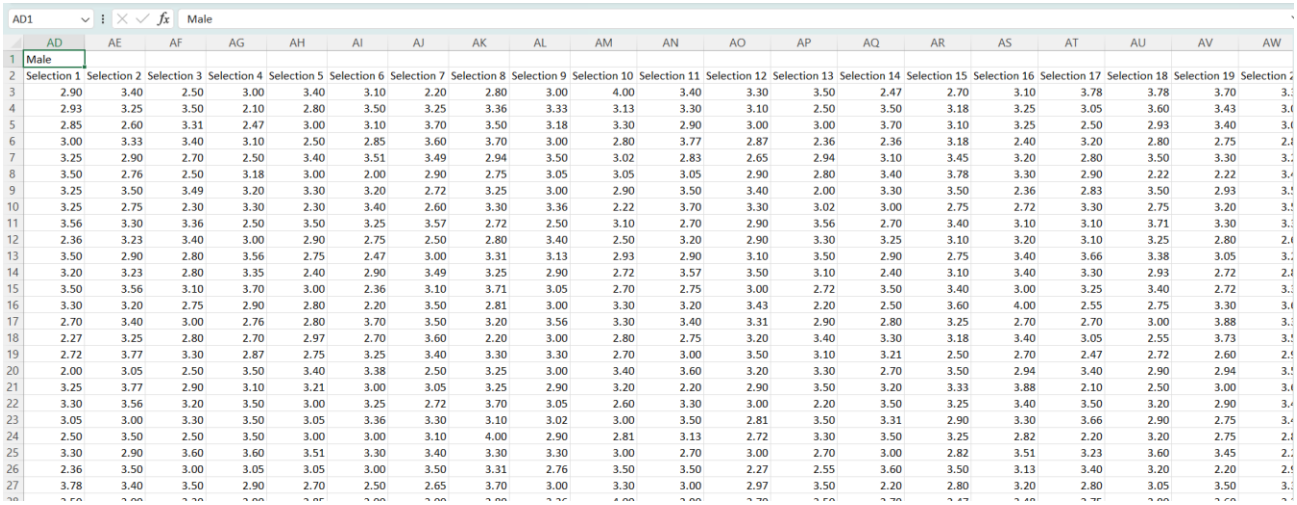
    =INDEX($B$156:$B$332,RANDBETWEEN(1,COUNTA($B$156:$B$332)),1)



C3 =INDEX($B$156:$B$332,RANDBETWEEN(1,COUNTA($B$156:$B$332)),1)

	C	D	E	F	G	H
1			Female			
2		Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4
3	2.90	1	3.70	3.25	3.13	3.15

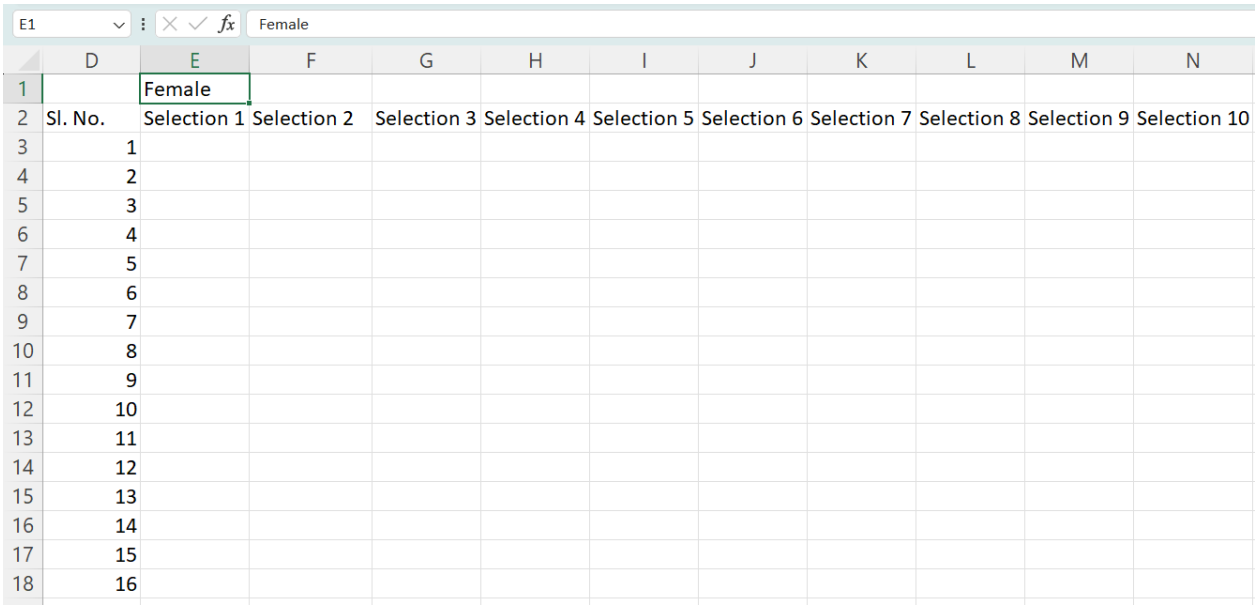
17. The above function will return a randomly selected data point from the *GPAs* corresponding to the set of data points that belong to the *Male* students.
18. Repeat steps 7 to 10.
19. To paste the set of 50 data points belonging to *Male* students, click on cell **AD3**. Paste Value as mentioned in step 11. Follow this step with step 12 to generate 25 selections of 50 randomly selected values from the *Male* students *GPA* data.
20. Once completed, one may obtain the sample data for *Male* students as,



	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	Male																			
2	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5	Selection 6	Selection 7	Selection 8	Selection 9	Selection 10	Selection 11	Selection 12	Selection 13	Selection 14	Selection 15	Selection 16	Selection 17	Selection 18	Selection 19	Selection 2
3	2.90	3.40	2.50	3.00	3.40	3.10	2.20	2.80	3.00	4.00	3.40	3.30	3.50	2.47	2.70	3.10	3.78	3.78	3.70	3.1
4	2.93	3.25	3.50	2.10	2.80	3.50	3.25	3.36	3.33	3.13	3.30	3.10	2.50	3.50	3.18	3.25	3.05	3.60	3.43	3.0
5	2.85	2.60	3.31	2.47	3.00	3.10	3.70	3.50	3.18	3.30	2.90	3.00	3.00	3.70	3.10	3.25	2.50	2.93	3.40	3.0
6	3.00	3.33	3.40	3.10	2.50	2.85	3.60	3.70	3.00	2.80	3.77	2.87	2.36	2.36	3.18	2.40	3.20	2.80	2.75	2.8
7	3.25	2.90	2.70	2.50	3.40	3.51	3.49	2.94	3.50	3.02	2.83	2.65	2.94	3.10	3.45	3.20	2.80	3.50	3.30	3.2
8	3.50	2.76	2.50	3.18	3.00	2.00	2.90	2.75	3.05	3.05	3.05	2.90	2.80	3.40	3.78	3.30	2.90	2.22	2.22	3.4
9	3.25	3.50	3.49	3.20	3.30	3.20	2.72	3.25	3.00	2.90	3.50	3.40	2.00	3.30	3.50	2.36	2.83	3.50	2.93	3.5
10	3.25	2.75	2.30	3.30	2.30	3.40	2.60	3.30	3.36	2.22	3.70	3.30	3.02	3.00	2.75	2.72	3.30	2.75	3.20	3.5
11	3.56	3.30	3.36	2.50	3.50	3.25	3.57	2.72	2.50	3.10	2.70	2.90	3.56	2.70	3.40	3.10	3.10	3.71	3.30	3.3
12	2.36	3.23	3.40	3.00	2.90	2.75	2.50	2.80	3.40	2.50	3.20	2.90	3.30	3.25	3.10	3.20	3.10	3.25	2.80	2.6
13	3.50	2.90	2.80	3.56	2.75	2.47	3.00	3.31	3.13	2.93	2.90	3.10	3.50	2.90	2.75	3.40	3.66	3.38	3.05	3.2
14	3.20	3.23	2.80	3.35	2.40	2.90	3.49	3.25	2.90	2.72	3.57	3.50	3.10	2.40	3.10	3.40	3.30	2.93	2.72	2.8
15	3.50	3.56	3.10	3.70	3.00	2.36	3.10	3.71	3.05	2.70	2.75	3.00	2.72	3.50	3.40	3.00	3.25	3.40	2.72	3.1
16	3.30	3.20	2.75	2.90	2.80	2.20	3.50	2.81	3.00	3.30	3.20	3.43	2.20	2.50	3.60	4.00	2.55	2.75	3.30	3.6
17	2.70	3.40	3.00	2.76	2.80	3.70	3.50	3.20	3.56	3.30	3.40	3.31	2.90	2.80	3.25	2.70	2.70	3.00	3.88	3.1
18	2.27	3.25	2.80	2.70	2.97	2.70	3.60	2.20	3.00	2.80	2.75	3.20	3.40	3.30	3.18	3.40	3.05	2.55	3.73	3.5
19	2.72	3.77	3.30	2.87	2.75	3.25	3.40	3.30	3.30	2.70	3.00	3.50	3.10	3.21	2.50	2.70	2.47	2.72	2.60	2.9
20	2.00	3.05	2.50	3.50	3.40	3.38	2.50	3.25	3.00	3.40	3.60	3.20	3.30	2.70	3.50	2.94	3.40	2.90	2.94	3.5
21	3.25	3.77	2.90	3.10	3.21	3.00	3.05	3.25	2.90	3.20	2.20	2.90	3.50	3.20	3.33	3.88	2.10	2.50	3.00	3.0
22	3.30	3.56	3.20	3.50	3.00	3.25	2.72	3.70	3.05	2.60	3.30	3.00	2.20	3.50	3.25	3.40	3.50	3.20	2.90	3.4
23	3.05	3.00	3.30	3.50	3.05	3.36	3.30	3.10	3.02	3.00	3.50	2.81	3.50	3.31	2.90	3.30	3.66	2.90	2.75	3.4
24	2.50	3.50	2.50	3.50	3.00	3.00	3.10	4.00	2.90	2.81	3.13	2.72	3.30	3.50	3.25	2.82	2.20	3.20	2.75	2.8
25	3.30	2.90	3.60	3.60	3.51	3.30	3.40	3.30	3.30	3.00	2.70	3.00	2.70	3.00	2.82	3.51	3.23	3.60	3.45	2.2
26	2.36	3.50	3.00	3.05	3.05	3.00	3.50	3.31	2.76	3.50	3.50	2.27	2.55	3.60	3.50	3.13	3.40	3.20	2.20	2.5
27	3.78	3.40	3.50	2.90	2.70	2.50	2.65	3.70	3.00	3.30	3.00	2.97	3.50	2.20	2.80	3.20	2.80	3.05	3.50	3.1

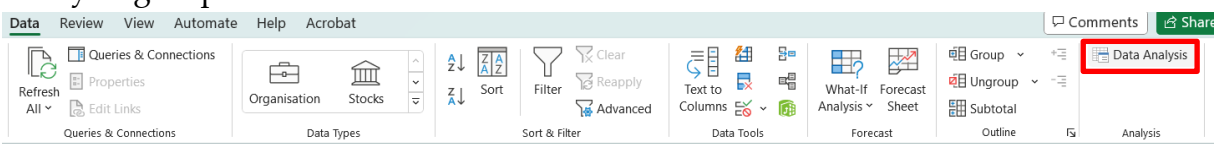
Using Data Analysis Toolpak

To select data randomly using the **Data Analysis** feature, '**Analysis Toolpak**' add-in should be installed in Excel. Follow steps 1-4 as mentioned in the previous section to have a worksheet with the *GPA* by *Gender* data, and a template to insert the random selections. At the end of step 4, one should ideally be having a layout as shown below.

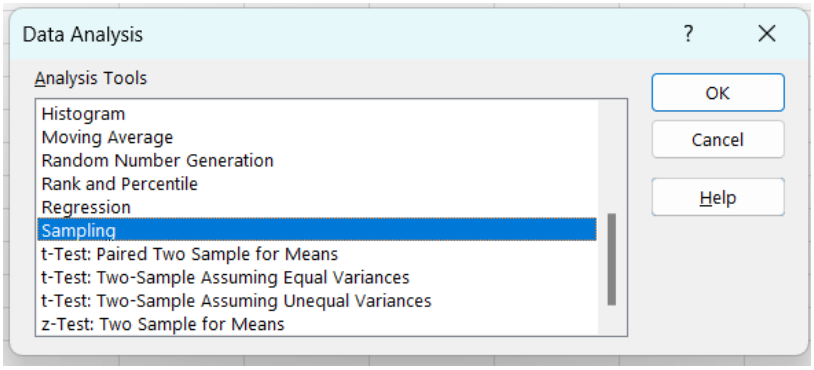


	D	E	F	G	H	I	J	K	L	M	N
1		Female									
2	Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5	Selection 6	Selection 7	Selection 8	Selection 9	Selection 10
3	1										
4	2										
5	3										
6	4										
7	5										
8	6										
9	7										
10	8										
11	9										
12	10										
13	11										
14	12										
15	13										
16	14										
17	15										
18	16										

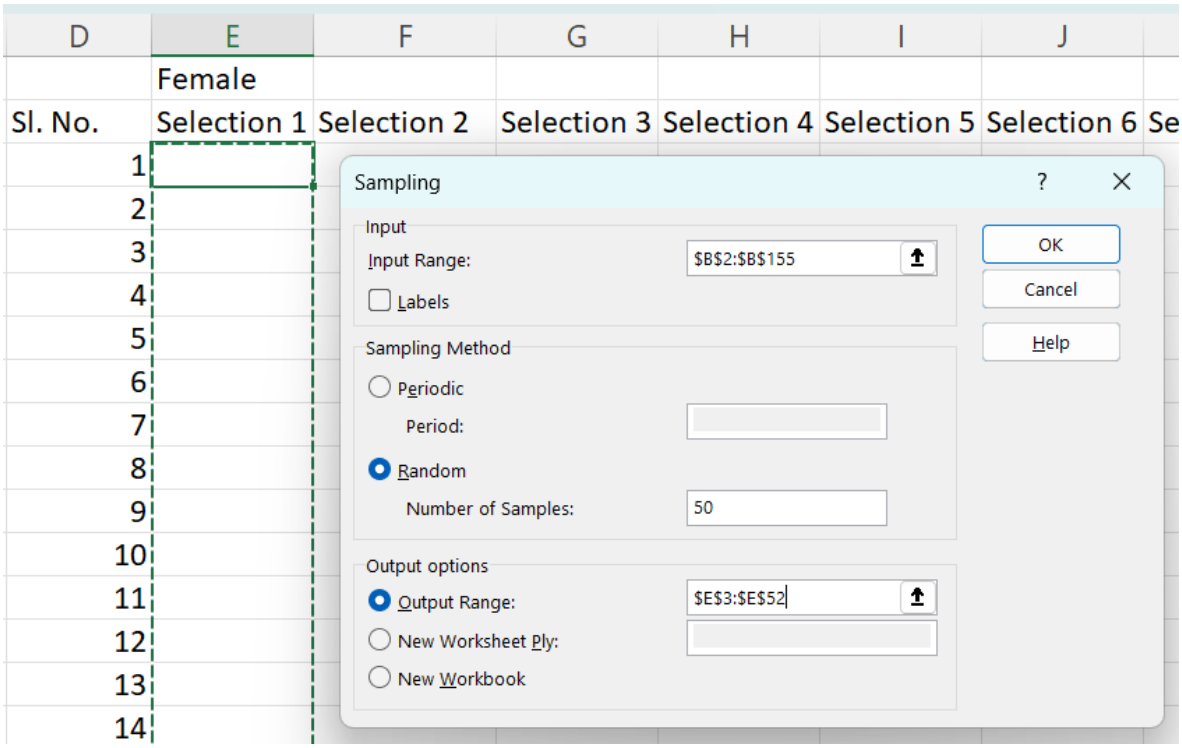
1. Navigate to the Data tab and click on the Data Analysis button placed in the Analysis group.



2. In the Data Analysis window that opens, select Sampling and click **OK**.



3. In the next window, select the input range. In this case, you need to select the *GPA* values that correspond to *Female* students. Hence, the input range to include will be **B2:B155**. To do this, click on the white box beside the **Input Range** and then click on cell **B2** and drag the selection to cell **B155**. Set the **Number of Samples** to be 50. Insert cells in the range **E3:E52** as the location to return the output. This can be done by clicking on the white box beside the **Output Range** and on cell **E3** and drag the selection to cell **E52**. All the other options can be left unchanged. Finally, click on **OK**.



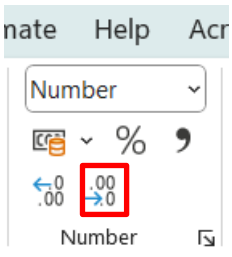
4. The above step will return 50 randomly selected values to under *Female*, Selection 1 column.

D	E
	Female
Sl. No.	Selection 1
1	3.30
2	3.30
3	2.75
4	3.57
5	3.60
6	3.10
7	3.50
8	2.90
9	2.23

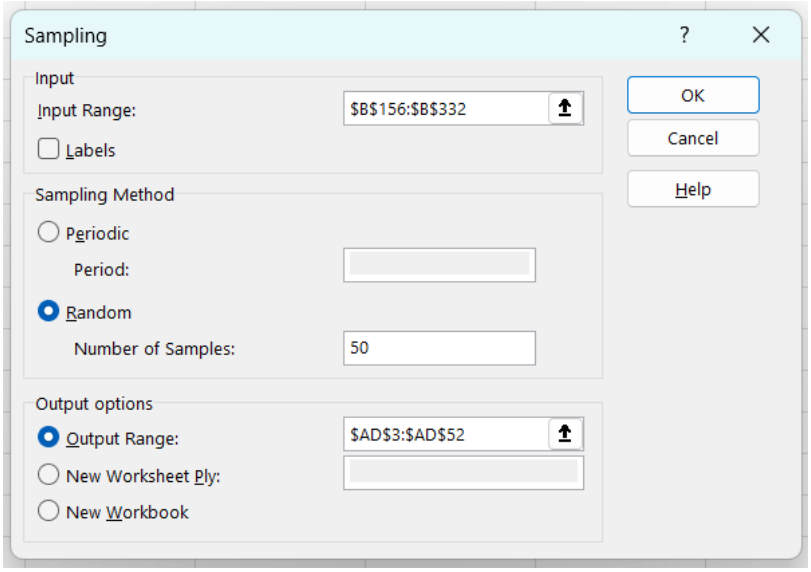
5. Now, the steps 1-3 can be repeated 24 times to fill in the rest of the 24 random selections.

D	E	F	G	H	I
	Female				
Sl. No.	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5
1	3.30	3.4	3.5	3.25	3.6
2	3.30	3.4	2.6	3	2.9
3	2.75	2.75	3.1	3.7	3.3
4	3.57	3.15	3.5	3	3.25
5	3.60	3.7	3.1	3.6	3.89
6	3.10	2.8	3.86	3.25	2.6
7	3.50	3.75	3.01	3.48	2.6
8	2.90	3.5	2.9	3.36	2.9

6. One can increase or decrease the decimal places of these randomly selected data by changing the format of cells to **Number** and clicking on the **Increase** or **Decrease decimal** points button.



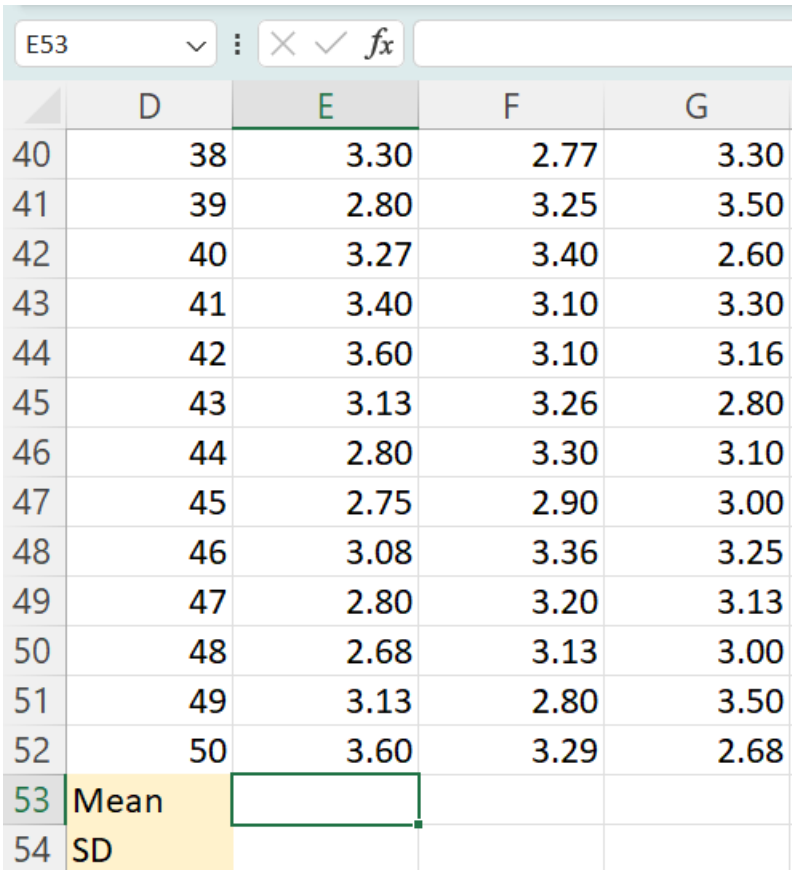
7. Similarly, the random selection of 50 data points can be made for *Male* students using the same method. The major difference will only be in the selection of input range and output locations in step 3.


Sampling
Input
Input Range:
$B$156:$B$332
Labels
Sampling Method
Periodic
Period:
Random
Number of Samples:
50
Output options
Output Range:
$AD$3:$AD$52
New Worksheet Ply:
New Workbook
OK
Cancel
Help

SAMPLING DISTRIBUTION

Calculating Average of Means and the Standard Error

Question: Report the measures of centre and measures of spread of GPA by Gender. Repeat the random selection process carried out in the previous section 25 times and record the means and standard deviations for Male and Female.



	D	E	F	G
40	38	3.30	2.77	3.30
41	39	2.80	3.25	3.50
42	40	3.27	3.40	2.60
43	41	3.40	3.10	3.30
44	42	3.60	3.10	3.16
45	43	3.13	3.26	2.80
46	44	2.80	3.30	3.10
47	45	2.75	2.90	3.00
48	46	3.08	3.36	3.25
49	47	2.80	3.20	3.13
50	48	2.68	3.13	3.00
51	49	3.13	2.80	3.50
52	50	3.60	3.29	2.68
53	Mean			
54	SD			

1. Click on cell **E53**.
2. Insert the following function to calculate the mean of 50 values in the first selection.

   =**AVERAGE**(E3:E52)

3. Use the AutoFill handle to reproduce the function for the remaining 24 selections corresponding to the randomly selected *GPAs* of *Female* and the 25 *Male* students. For demonstration purposes, only the results of *Female* students are shown below.

E53 =AVERAGE(E3:E52)

	D	E	F	G	H	I	J	K
47	45	2.75	2.90	3.00	3.10	3.40	3.00	2.80
48	46	3.08	3.36	3.25	3.70	3.85	3.50	3.55
49	47	2.80	3.20	3.13	3.00	3.10	2.23	3.00
50	48	2.68	3.13	3.00	2.90	2.94	3.55	3.20
51	49	3.13	2.80	3.50	3.40	3.10	3.10	3.68
52	50	3.60	3.29	2.68	3.88	2.23	2.70	2.60
53	Mean	3.2432	3.2128	3.2036	3.1956	3.2758	3.2724	3.1908

4. Now, click on cell **E54**.
5. Insert the following function to calculate the standard deviation of the selected values in the first selection. As we are dealing with sample data, the function to calculate the standard deviation based on the sample (**STDEV.S**) should be used.

=STDEV.S(E3:E52)

E54 =STDEV.S(E3:E52)

	D	E	F
50	48	2.68	3.13
51	49	3.13	2.80
52	50	3.60	3.29
53	Mean	3.2432	3.2128
54	SD	0.36	

6. Use the AutoFill handle to reproduce the function for the remaining 24 selections corresponding to the randomly selected *GPAs* of *Female* and the 25 *Male* students.

E54 =STDEV.S(E3:E52)

	D	E	F	G	H	I
52	50	3.60	3.29	2.68	3.88	2.23
53	Mean	3.2432	3.2128	3.2036	3.1956	3.2758
54	SD	0.36	0.29	0.39	0.34	0.35

7. Up next, one needs to calculate the average of means for *GPAs* of *Female* and *Male* students. Before doing this, just under the captions of Mean and Standard Deviation, one may insert small captions like the following.

	D	E	F	G
52	50	3.60	3.29	2.68
53	Mean	3.2432	3.2128	3.2036
54	SD	0.36	0.29	0.39
55	Average of Means			Female
56				Male
57	SD of Means = SE of the population			Female
58				Male

8. In cell **H55**, use the following function to calculate the average of the means corresponding to *Female* students.

=AVERAGE(E53:AC53)

9. The range **E53:AC53** corresponds to the means calculated for the 25 random selections for *Female* students using steps 2 and 3. If you have used a different location for the data or the means then this range needs to be changed.
10. In cell **H56**, use the AVERAGE function once again to calculate the average of the means of *Male* students. Remember to change the data range in the function to include the means of *Male* students as shown below.

=AVERAGE(AD53:BB53)

11. Steps 8 and 10 will return the Average of the means for *Female* and *Male* students as the following. Note that, these values will differ from user to user as the numbers are randomly selected.

	D	E	F	G	H
53	Mean	3.2432	3.2128	3.2036	3.1956
54	SD	0.36	0.29	0.39	0.34
55	Average of Means			Female	3.23
56				Male	3.10

12. Before calculating the standard error, one may add captions like in step 7 to keep the table laid out neatly. One may choose to follow the template as shown below.

	D	E	F	G	H
53	Mean	3.2432	3.2128	3.2036	3.1956
54	SD	0.36	0.29	0.39	0.34
55	Average of Means			Female	3.23
56				Male	3.10
57	SD of Means = SE of the population			Female	
58				Male	

13. Now, click on cell **H57**.

14. Insert the following function to calculate the standard error corresponding to *Female* students. Note that the standard deviation of the means of 25 selections for each decade will correspond to the standard error of the population.

=STDEV.S(E53:AC53)

15. The range **E53:AC53** in the above function corresponds to the means calculated for the 25 random selections for *Female* students using steps 2 and 3.
16. Click on cell **H58** and insert the following function to calculate the standard error corresponding to *Male* students.

=STDEV.S(AD53:BB53)

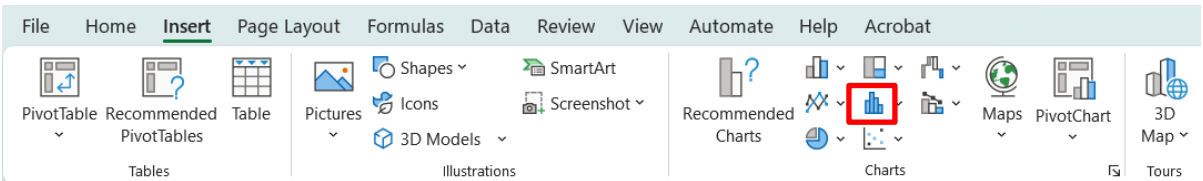
17. Steps 13 to 16 will return the standard errors for *Female* and *Male* students as the following.

	D	E	F	G	H
53	Mean	3.2432	3.2128	3.2036	3.1956
54	SD	0.36	0.29	0.39	0.34
55	Average of Means			Female	3.23
56				Male	3.10
57	SD of Means = SE of the population			Female	0.05
58				Male	0.05

Constructing Histograms of the distribution of Means

Once the means of all the 50 random selections have been made (refer to steps 1 to 4 in the previous section), the histogram can be used using the Statistic Chart option in Excel.

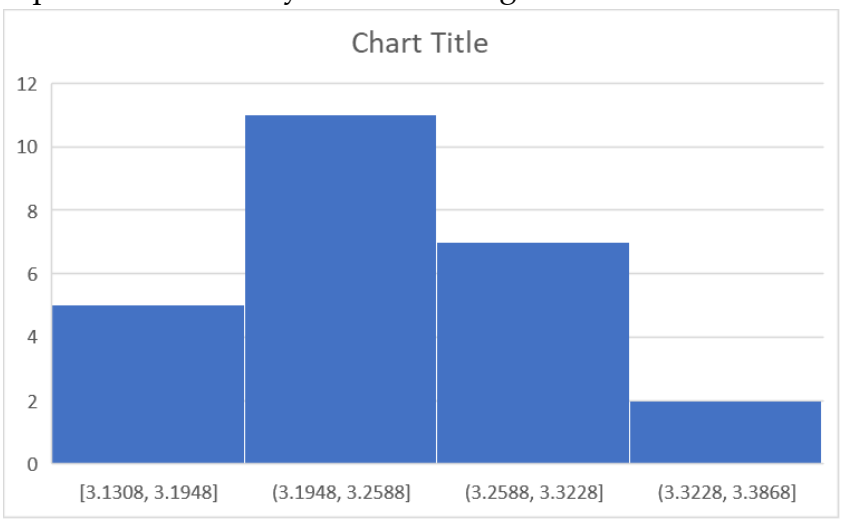
1. To do this select all the means corresponding to the selections made for *Female* students.
2. Navigate to the **Insert** tab in the Excel ribbon and click on the **Insert Statistic Chart** button under the **Charts** group.



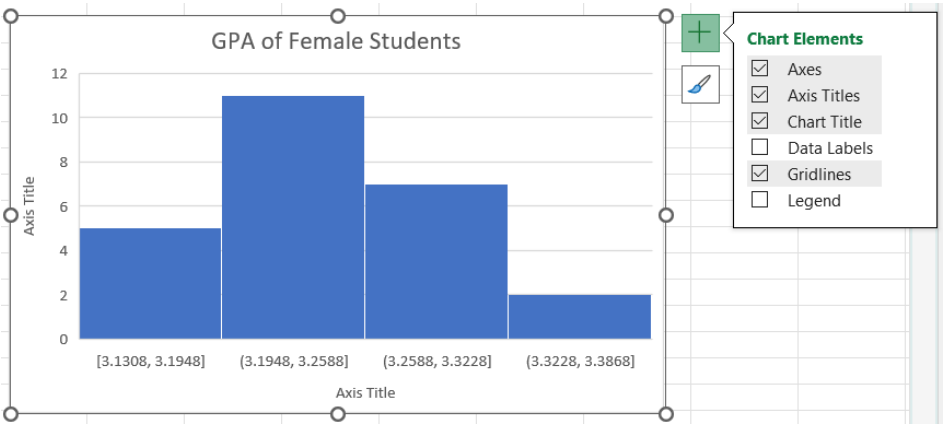
=AVERAGE(E3:E52)

	D	E	F	G	H	I	J
49	47	2.80	3.20	3.13	3.00	3.10	2.23
50	48	2.68	3.13	3.00	2.90	2.94	3.55
51	49	3.13	2.80	3.50	3.40	3.10	3.10
52	50	3.60	3.29	2.68	3.88	2.23	2.70
53	Mean	3.2432	3.2128	3.2036	3.1956	3.2758	3.2724
54	SD	0.36	0.29	0.39	0.34	0.35	0.40

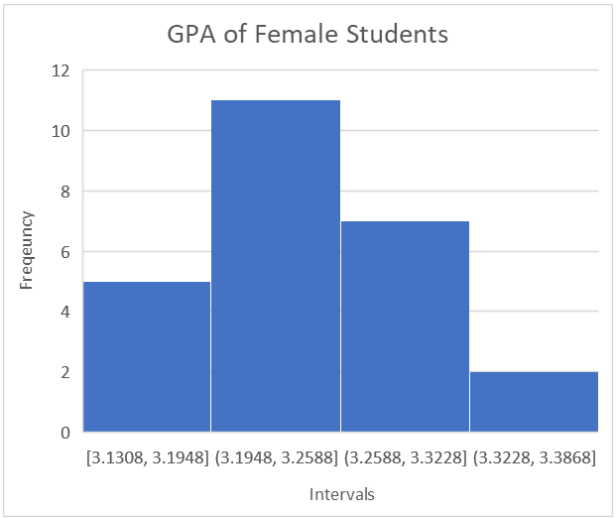
3. The above step will immediately return a histogram as shown below.



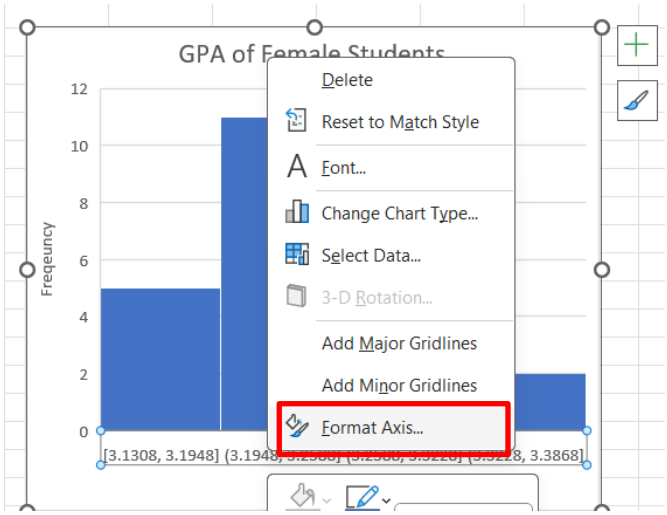
4. However, this histogram can be modified by using the Format options. One may insert appropriate chart title and axis titles to the histogram. To insert a chart title, double click on the current chart title and type in a suitable title. Secondly, to insert axis titles, click on the Chart Elements icon on the top right of the chart and enable the axis titles option.



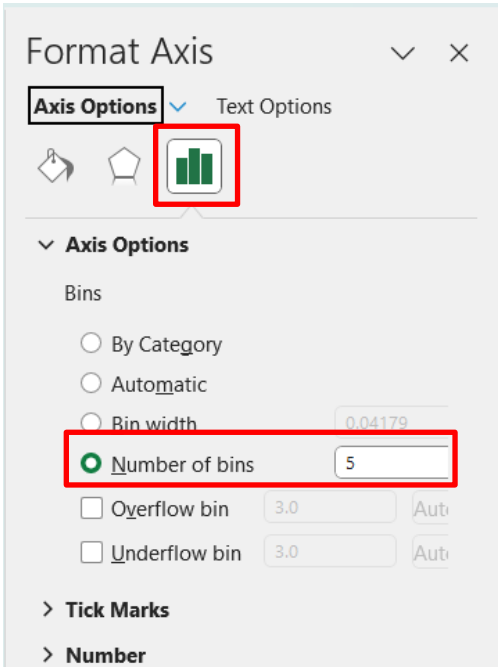
5. Double click on the horizontal and vertical axis titles and type in appropriate titles.



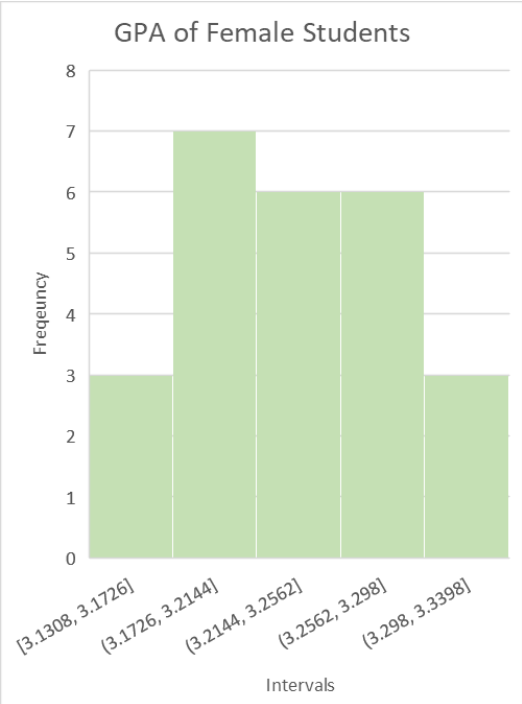
6. One may also change the number of bins if needed. To do this, right click on the X-axis and click on **Format Axis**.



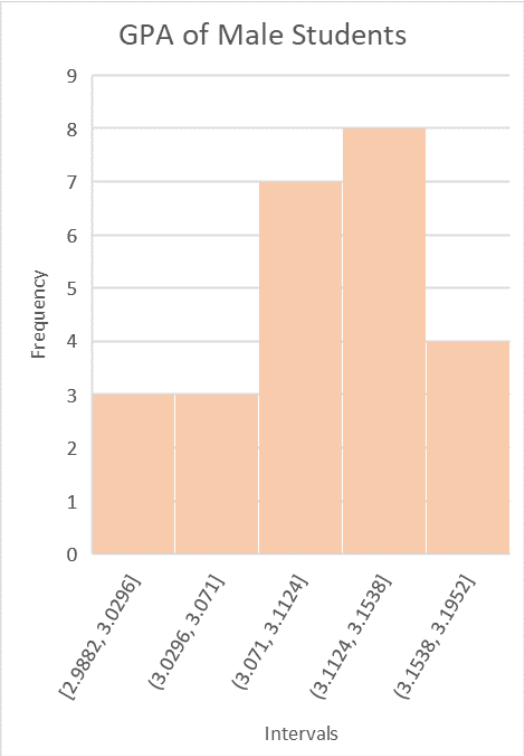
7. In the window that appears on the right side of the screen, under **Axis Options**, click on **Number of bins** and choose the number of bins required depending on the data.



8. After some formatting, one may obtain the histogram showing the distribution of the means of the *GPAs* of *Female* students as shown below.



9. Similarly, one can obtain the histogram for the distribution of means for *Male* students using steps 2-8 after selecting the means for *Male* students in the first step.


GPA of Male Students
Frequency
0
1
2
3
4
5
6
7
8
9
[2.9882, 3.0296]
(3.0296, 3.071]
(3.071, 3.1124]
(3.1124, 3.1538]
(3.1538, 3.1952]
Intervals

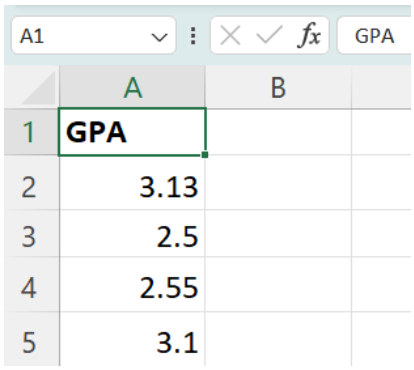
HYPOTHESIS TESTING

One-tailed hypothesis tests

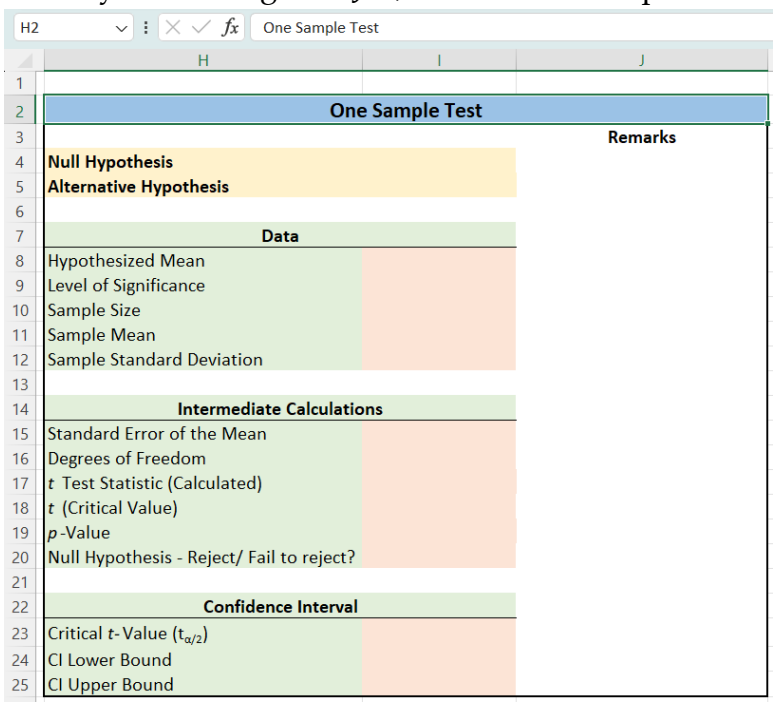
Question: Suppose the national average of *GPA* is 2.9. Is there evidence that the students in this college have a *GPA* more than the national average?

User needs to copy the *GPA* data from the 'StudentSurvey' worksheet to carry out the one-tailed hypothesis test.

1. Copy the *GPA* data to a new worksheet. Preferably, paste them in the range **A1:A332**.



2. To carry out the hypothesis testing, one can create a template in the current worksheet, preferably in the range **H2:J25**, as shown in the picture below.

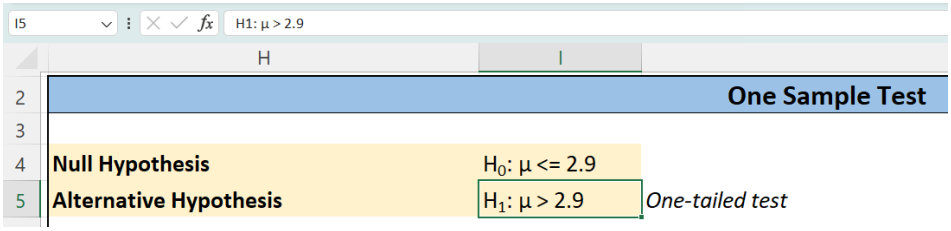


3. For this question, the null hypothesis and alternative hypothesis were identified as the following.

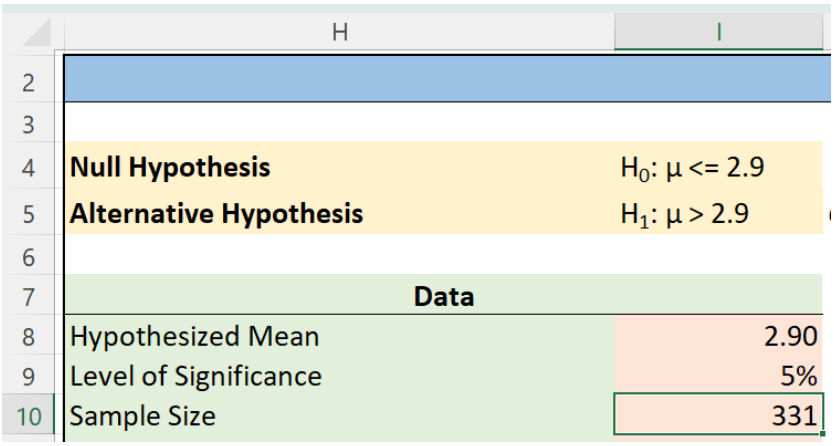
a. **Null hypothesis**: The mean GPA of students in the college is less than or equal to 2.9.
b. **Alternative hypothesis**: The mean GPA of students in the college is greater than 2.9.

As the alternative hypothesis includes a greater than condition, this falls under one-tailed hypothesis tests.

4. Accordingly, one may fill cells I4 and I5 with the null and alternative hypothesis.



5. We know that the hypothesized mean is 2.9, and the sample size is 331. Also, for this example, we choose the level of significance to be 5%. One may fill in the cells **I8:I10** with these values. Make sure that the level of significance is in the percentage format.



6. The sample mean can be calculated in cell **I11** using the following formula. Note that the formula works only if the sample of *GPA* is pasted in the cells **A2:A332**. If you have placed the sample elsewhere, then you may modify the formula to accommodate the range that contains the sample data.

=AVERAGE(A2:A332)

7. The sample standard deviation can be calculated in cell **I12** using the following formula. Just like the sample mean calculated in the previous step, the range included in the formula changes if the sample of *GPA* is pasted anywhere other than in the range **A2:A332**.

=STDEV.S(A2:A332)

8. Once calculated the table will look like the following.

	H	I
2		
3		
4	**Null Hypothesis**	H_0: μ <= 2.9
5	**Alternative Hypothesis**	H_1: μ > 2.9
6		
7	**Data**	
8	Hypothesized Mean	2.90
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	3.16
12	Sample Standard Deviation	0.40

9. The next parameter to calculate is the standard error of the mean. To calculate it, one needs to divide the sample standard deviation with the square root of sample size. Using the following formula in cell **I15**, one can calculate the standard error of the mean for this sample data.

=$I$12/SQRT($I$10)

10. The degrees of freedom can be calculated by subtracting 1 from the sample size. In this example, this can be calculated in cell **I16** using the following formula.

=$I$10-1

11. At the end of step 11, the table will look like the following.

	H	I
2		
3		
4	**Null Hypothesis**	H_0: μ <= 2.9
5	**Alternative Hypothesis**	H_1: μ > 2.9
6		
7	**Data**	
8	Hypothesized Mean	2.90
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	3.16
12	Sample Standard Deviation	0.40
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.02
16	Degrees of Freedom	330

12. In the next step, one needs to calculate the *t*-test statistic. This can be calculated in cell **I17** using the following formula.

=($I$11-$I$8)/$I$15

In terms of statistics, it is calculated as

(Sample mean – Hypothesized Mean)/(Standard error of the mean)

If the sample mean, hypothesized mean and standard error of the mean are not placed in cells **I11**, **I8** and **I15** respectively, then these cell references need to be changed.

13. As this is a right-tailed test, the critical value can be calculated in cell **I18** using the following formula.

=T.INV(1 - I9, I16)

where **I9** and **I16** corresponds to level of significance and degrees of freedom respectively. The **T.INV** function in Excel returns the left-tailed inverse of the student's t-distribution.

14. After calculating the t-test statistic and the critical value, the table will look like the following. Note that some of the values will be different as the sample created will be different each time.

	H	I
2		
3		
4	**Null Hypothesis**	H_0: μ <= 2.9
5	**Alternative Hypothesis**	H_1: μ > 2.9
6		
7	**Data**	
8	Hypothesized Mean	2.90
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	3.16
12	Sample Standard Deviation	0.40
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.02
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	11.81
18	*t* (Critical Value)	1.65

15. Before checking if the null hypothesis needs to be rejected, one may also find the *p*-value. One can make use of the **T.DIST** function in Excel that corresponds to the t-distribution to find the *p*-value. In cell **I19**, one can paste the following formula and obtain the *p*-value for the right-tailed test.

=1 - T.DIST(I17,I16,TRUE)

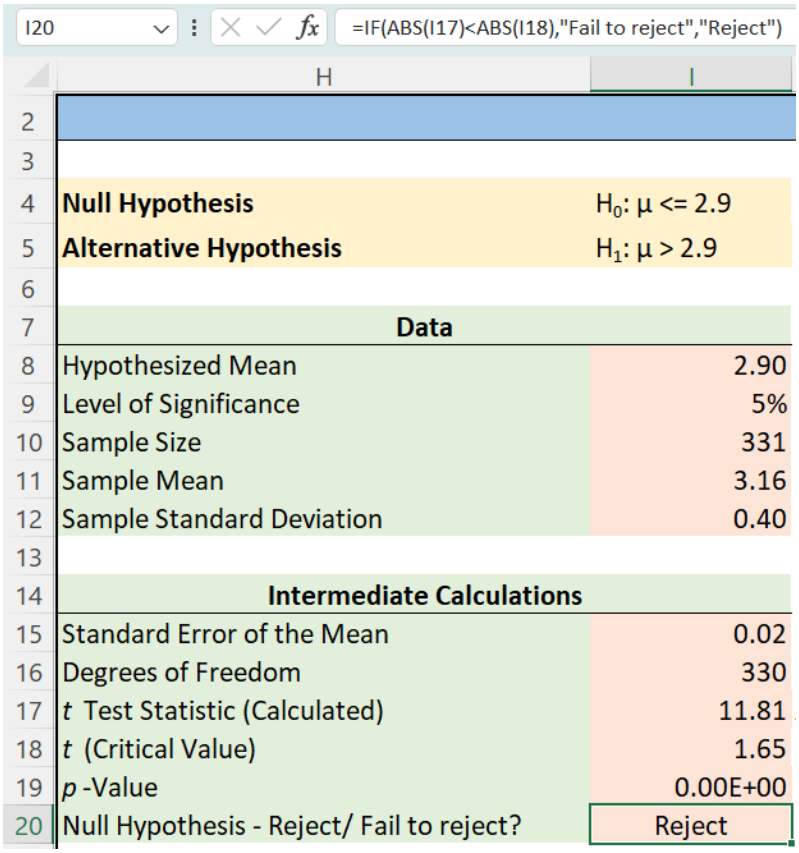
where the value in **I17** and **I16** corresponds to the t-statistic (calculated value) and the degrees of freedom respectively. The last argument corresponds to the the area under the curve to the left of *t* when Cumulative = *TRUE*.

16. If the absolute value of the calculated test statistic is less than the absolute value of the critical value, then we fail to reject the null hypothesis. Otherwise, we reject the hypothesis. This can be converted into an Excel formula and pasted in cell **I20** as below.

=IF(ABS(I17) < ABS(I18),"Fail to reject", "Reject")

I17 and **I18** correspond to the t-test statistic (calculated) and the critical value respectively.

17. The table updated with the results of steps 16 and 17 will look like the following.

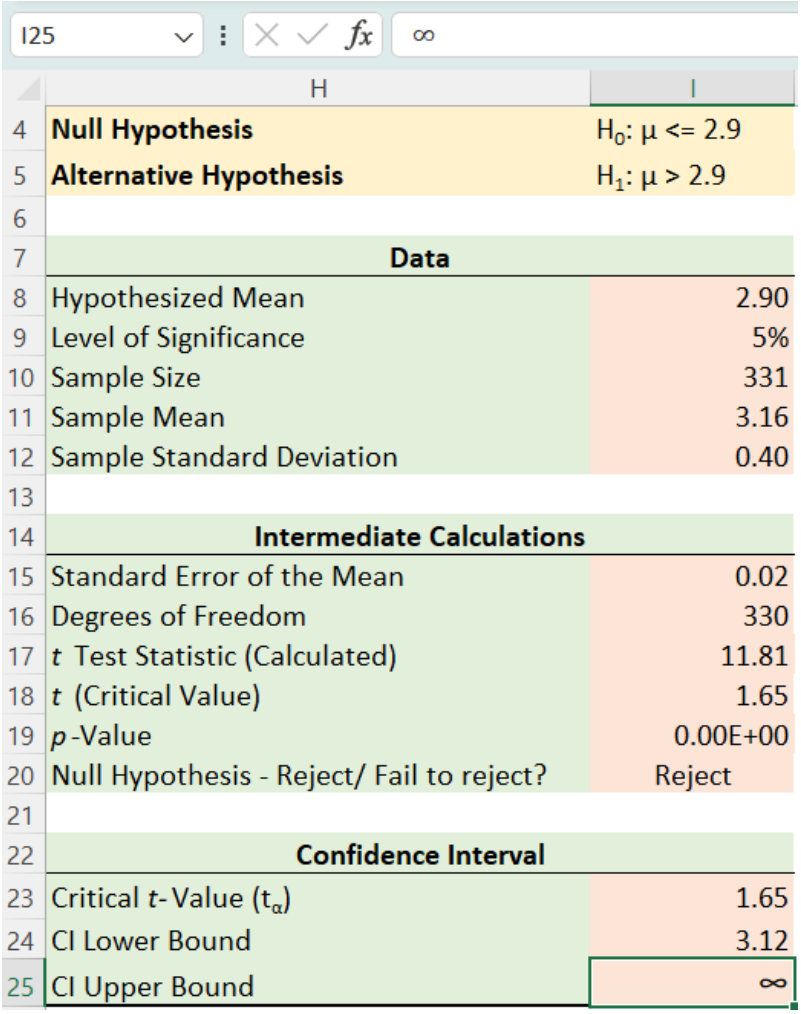


	H	I
2		
3		
4	**Null Hypothesis**	H_0: μ <= 2.9
5	**Alternative Hypothesis**	H_1: μ > 2.9
6		
7	**Data**	
8	Hypothesized Mean	2.90
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	3.16
12	Sample Standard Deviation	0.40
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.02
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	11.81
18	*t* (Critical Value)	1.65
19	*p* -Value	0.00E+00
20	Null Hypothesis - Reject/ Fail to reject?	Reject

18. To report the 95% confidence interval for the mean *GPA*, one can copy the *t*-statistic critical value that was generated in step 13. Insert **=I18** in cell **I23** to paste the t-statistic critical value in cell **I23**.
19. Once the critical *t*-value is calculated the lower critical value can be calculated in cell **I24** by using the following formula:

= I11 - I15*I23

20. As this is a right tailed distribution, the upper critical value can be set as ∞.
21. After the confidence interval calculations, the table will look like the following.



	H	I
4	**Null Hypothesis**	H_0: μ <= 2.9
5	**Alternative Hypothesis**	H_1: μ > 2.9
6		
7	**Data**	
8	Hypothesized Mean	2.90
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	3.16
12	Sample Standard Deviation	0.40
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.02
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	11.81
18	*t* (Critical Value)	1.65
19	*p*-Value	0.00E+00
20	Null Hypothesis - Reject/ Fail to reject?	Reject
21		
22	**Confidence Interval**	
23	Critical *t*-Value (t_α)	1.65
24	CI Lower Bound	3.12
25	CI Upper Bound	∞

Constructing one-sided confidence interval for a one-sided t-test

For a left tailed test

The one-sided confidence interval is therefore **$(-\infty$,95% Upper Bound)**, where the 95% Upper Bound = $\bar{x} + t_\propto \frac{s}{\sqrt{n}}$ and $t_\propto$ is the one tailed critical t-value.

For a right tailed test

The one-sided confidence interval is therefore **(95% Lower Bound, ∞)**, where the 95% Lower Bound = $\bar{x} - t_\propto \frac{s}{\sqrt{n}}$ and $t_\propto$ is the one tailed critical t-value.

NOTE

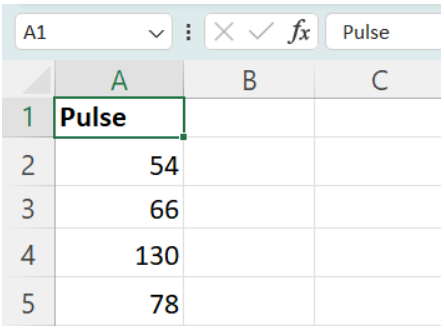
Critical value for right tailed tests, t_α = **T.INV(1 – α, df)**

Two-tailed hypothesis tests

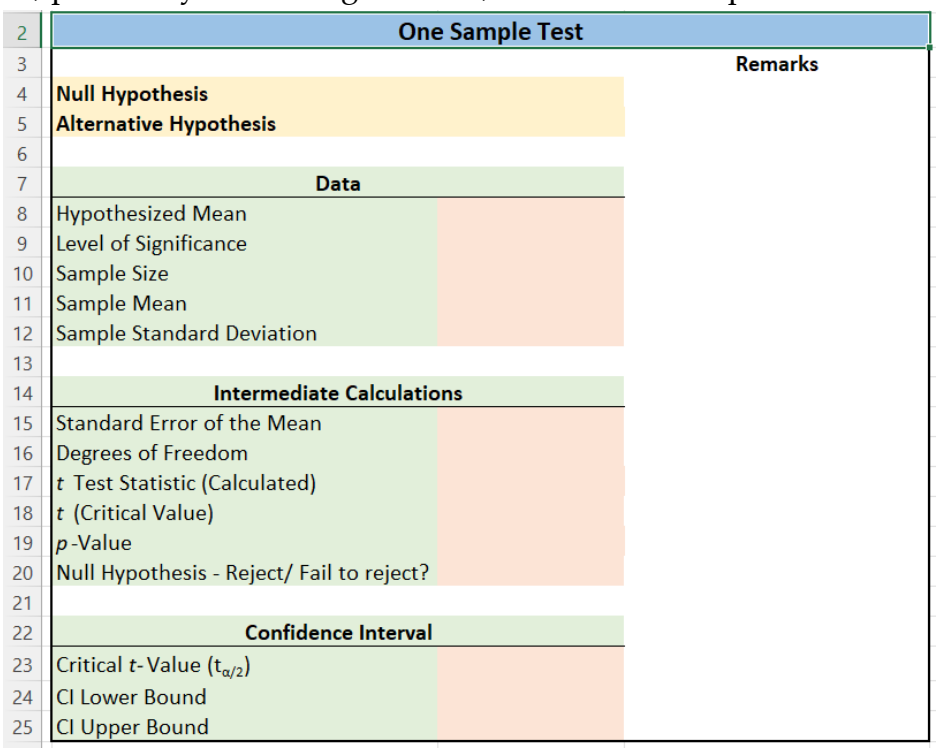
Question: Using the entire dataset, will the mean *pulse rate* of students surveyed be different from 72 beats per minute?

User needs to copy the *Pulse rate* data from the CleanedStudentSurvey Excel file to carry out the two-tailed hypothesis test.

1. Copy the *Pulse rate* data to a new worksheet. Preferably, paste them in the range **A1:A332**.



2. To carry out the hypothesis testing, one can create a template in the current worksheet, preferably in the range **F2:H25**, as shown in the picture below.

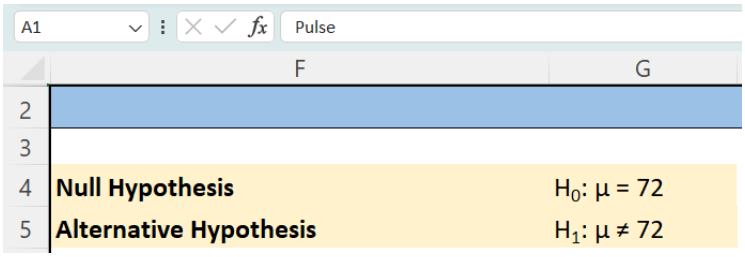


3. For this question, the null hypothesis and alternative hypothesis were identified as the following.
    a. **Null hypothesis**: Mean pulse rate of students surveyed is equal to 72 beats per minute.

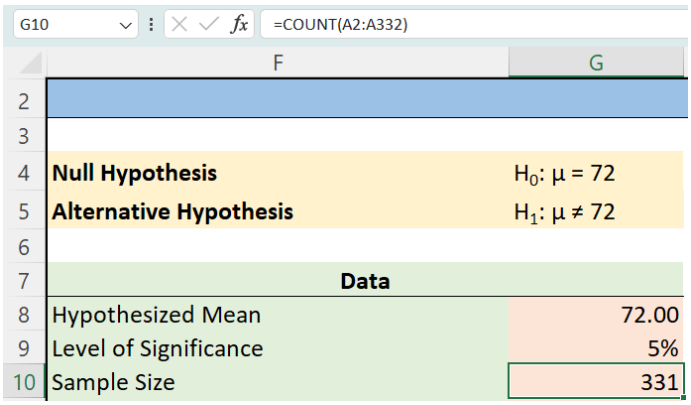
b. **Alternative hypothesis**: Mean pulse rate of students surveyed is different from 72 beats per minute.

As the alternative hypothesis includes a not equal to condition, this falls under two-tailed hypothesis tests.

4. Accordingly, one may fill cells **G4** and **G5** with the null and alternative hypothesis.



5. We know that the hypothesized mean is 72, and the sample size is 331. One thing a user needs to fix is the level of significance. For this example, we choose the level of significance to be 5%. One may fill in the cells **G8:G10** with these values. Make sure that the level of significance is in the percentage format.



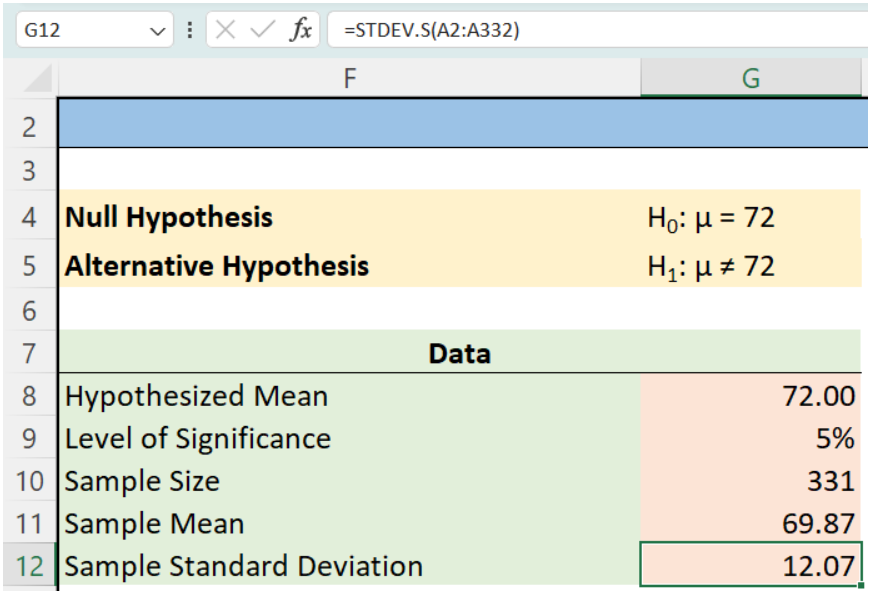
6. The sample mean can be calculated in cell **G11** using the following formula. Note that the formula works only if the sample of *Pulse rate* is pasted in the cells **A2:A332**. If you have placed the sample elsewhere, then you may modify the formula to accommodate the range that contains the sample data.

=AVERAGE(A2:A332)

7. The sample standard deviation can be calculated in cell **G12** using the following formula. Just like the sample mean calculated in the previous step, the range included in the formula changes if the sample of *Pulse rate* is pasted anywhere other than in the range **A2:A332**.

=STDEV.S(A2:A332)

8. Once calculated the table will look like the following.



9. The next parameter to calculate is the standard error of the mean. To calculate it, one needs to divide the sample standard deviation with the square root of sample size. Using the following formula in cell **G15**, one can calculate the standard error of the mean for this sample data.

=$G$12/SQRT($G$10)

10. The degrees of freedom can be calculated by subtracting 1 from the sample size. In this example, this can be calculated in cell **G16** using the following formula.

=$G$10-1

11. At the end of step 11, the table will look like the following.

Null Hypothesis	H_0: μ = 72
Alternative Hypothesis	H_1: μ ≠ 72
Data	
Hypothesized Mean	72.00
Level of Significance	5%
Sample Size	331
Sample Mean	69.87
Sample Standard Deviation	12.07
Intermediate Calculations	
Standard Error of the Mean	0.66
Degrees of Freedom	330

12. In the next step, one needs to calculate the *t*-test statistic. This can be calculated in cell **G17** using the following formula.

=($G$11-$G$8)/$G$15

In terms of statistics, it is calculated as

(Sample mean – Hypothesized Mean)/(Standard error of the mean)

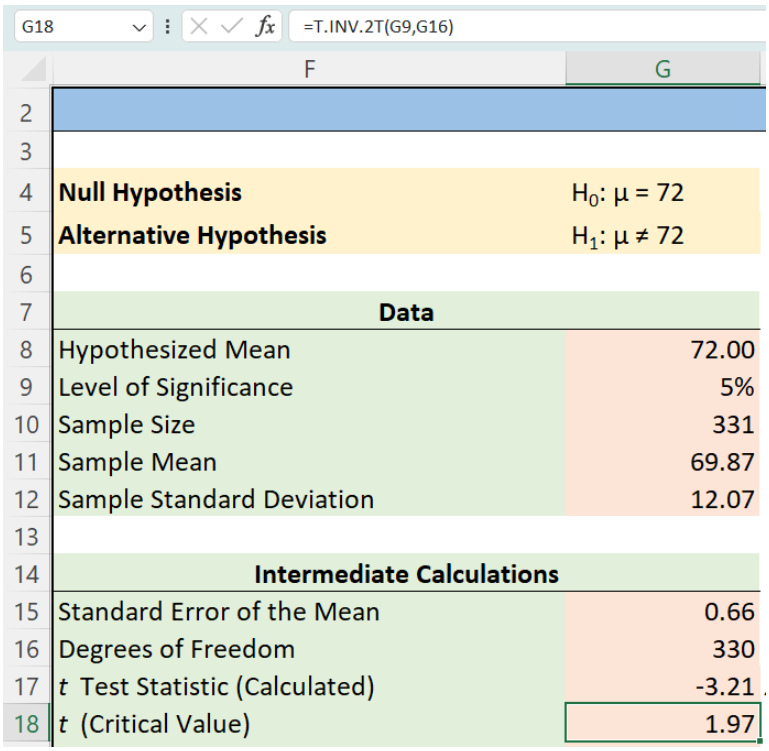
If the sample mean, hypothesized mean and standard error of the mean are not placed in cells **G11**, **G8** and **G15** respectively, then these cell references need to be changed.

13. As this is a two-tailed test, the critical value can be calculated in cell **G18** using the following formula.

=T.INV.2T(G9, G16)

where **G9** and **G16** corresponds to level of significance and degrees of freedom respectively.

14. After calculating the t-test statistic and the critical value, the table will look like the following. Note that some of the values will be different as the sample created will be different each time.



	F	G
2		
3		
4	**Null Hypothesis**	H_0: μ = 72
5	**Alternative Hypothesis**	H_1: μ ≠ 72
6		
7	**Data**	
8	Hypothesized Mean	72.00
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	69.87
12	Sample Standard Deviation	12.07
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.66
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	-3.21
18	*t* (Critical Value)	1.97

15. Before checking if the null hypothesis needs to be rejected, one may also find the *p*-value. One can make use of the **T.DIST** function in Excel that corresponds to the t-distribution to find the *p*-value. In cell **G19**, one can paste the following formula and obtain the *p*-value for the two-tailed test.

=T.DIST.2T(ABS(G17), G16)

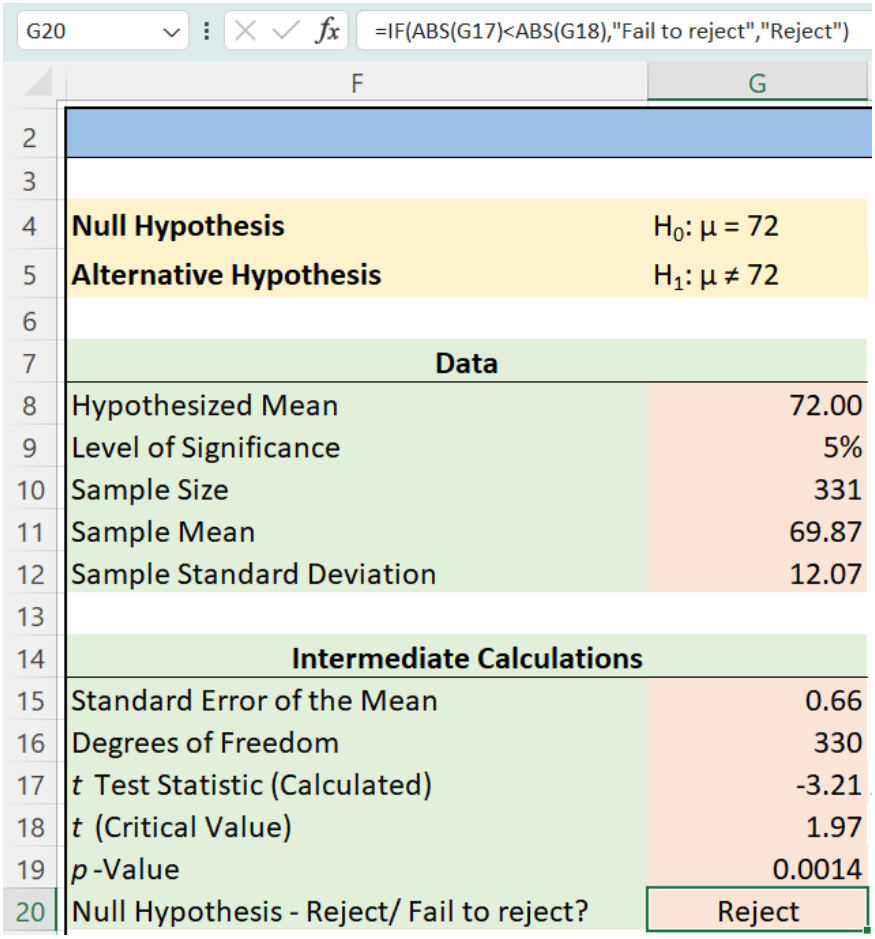
where the value in **G17** and **G16** corresponds to the t-statistic (calculated value) and the degrees of freedom respectively. **ABS()** is a function that returns the absolute value in Excel.

16. If the absolute value of the calculated test statistic is less than the absolute value of the critical value, then we fail to reject the null hypothesis. Otherwise, we reject the

hypothesis. This can be converted into an Excel formula and pasted in cell **G20** as below.

=IF(ABS(G17) < ABS(G18),"Fail to reject","Reject")

G17 and **G18** correspond to the t-test statistic (calculated) and the critical value respectively. The table updated with the results of steps 16 and 17 will look like the following.



	F	G
2		
3		
4	**Null Hypothesis**	H_0: μ = 72
5	**Alternative Hypothesis**	H_1: μ ≠ 72
6		
7	**Data**	
8	Hypothesized Mean	72.00
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	69.87
12	Sample Standard Deviation	12.07
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.66
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	-3.21
18	*t* (Critical Value)	1.97
19	*p*-Value	0.0014
20	Null Hypothesis - Reject/ Fail to reject?	Reject

17. To report the 95% confidence interval for the true mean of *Pulse rate*, one needs to calculate the *t*-statistic critical value. **T.INV.2T** returns the two-tailed inverse of the student's t-distribution. In cell **G23**, one can include **T.INV.2T** function as,

    =T.INV.2T(G9, G16)

    where **G9** and **G16** correspond to level of significance and the degrees of freedom respectively.
18. Once the critical *t*-value is calculated the lower critical value and the upper critical value can be calculated in cells **G24** and **G25** as follows.
    a. CI Lower Bound: = **G11 - G15*G23**
    b. CI Upper Bound: = **G11 + G15*G23**

    where **G11**, **G15** and **G23** represents sample mean, standard error of the mean and the critical *t*-value respectively.
19. After the confidence interval calculations, the table will look like the following.

	F	G
2		
3		
4	**Null Hypothesis**	H_0: μ = 72
5	**Alternative Hypothesis**	H_1: μ ≠ 72
6		
7	**Data**	
8	Hypothesized Mean	72.00
9	Level of Significance	5%
10	Sample Size	331
11	Sample Mean	69.87
12	Sample Standard Deviation	12.07
13		
14	**Intermediate Calculations**	
15	Standard Error of the Mean	0.66
16	Degrees of Freedom	330
17	*t* Test Statistic (Calculated)	-3.21
18	*t* (Critical Value)	1.97
19	*p*-Value	0.0014
20	Null Hypothesis - Reject/ Fail to reject?	Reject
21		
22	**Confidence Interval**	
23	Critical *t*-Value ($t_{\alpha/2}$)	1.97
24	CI Lower Bound	68.56
25	CI Upper Bound	71.17

Hypothesis tests about a Mean: σ Not Known (*t*-test)

Finding Critical Values:

To find the critical values, the **T.INV** function of Excel is used. **T.INV** stands for the inverse of *t*-distribution. In Excel's definition, the **T.INV** function has the following syntax.

=**T.INV**(probability, deg_freedom)

In simple words, the first argument inside the bracket tells Excel corresponds to the area to the left of the critical value and the second argument refers to the degrees of freedom.

This function returns the critical value from the *t*-distribution provided you put in the appropriate area and degrees of freedom.

Left-tailed test: = **T.INV(*α*, df)**

Right-tailed test: = **T.INV(1 – *α*, df)**

Two-tailed test: = **± T.INV.2T(ABS(*α*), df)**

Finding *p*-values:

To calculate the *p*-values, the **T.DIST** function of Excel is used. **T.DIST** stands for the *t*-distribution. The syntax of the function in Excel takes the following form.

=T.DIST(t, df, (cumulative))

where *'t'* corresponds to the calculated *t*-statistic value, *'df'* corresponds to the degrees of freedom and *'cumulative'* takes true or false value depending on whether the function should return the cumulative distribution function or the probability density function.

This function returns the area under the curve to the left of *t* when Cumulative = TRUE.

Left-tailed test: = **T.DIST(t, df, TRUE)**

Right-tailed test: = **1 - T.DIST(t, df, TRUE)**

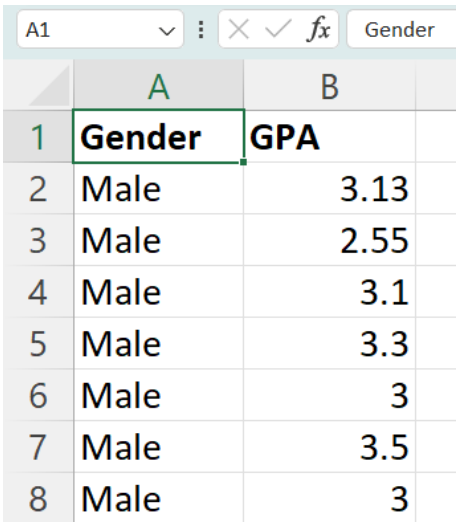
Two-tailed test: = **T.DIST.2T(ABS(t), df)**

COMPARING TWO POPULATION MEANS AND CONFIDENCE INTERVALS

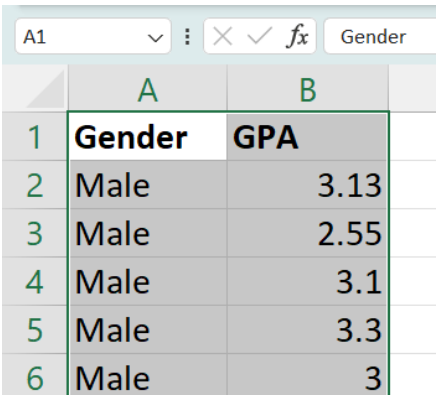
Question: Using the entire dataset, is there any difference between the average *GPA* by *Gender*?

Boxplot

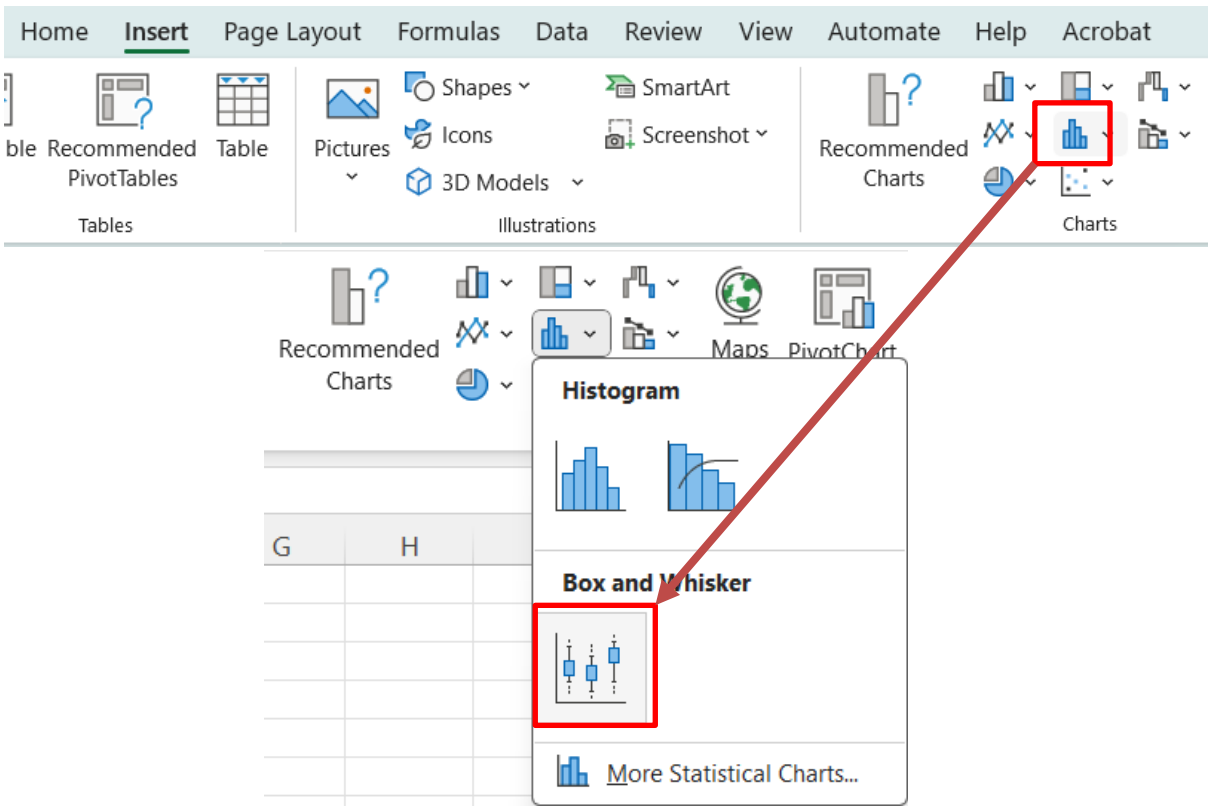
Boxplots can be created in Excel using the Charts function directly. To begin with, copy and paste the *Gender* and *GPA* data to a new worksheet. Preferably paste the values in the range **A1:B332**.



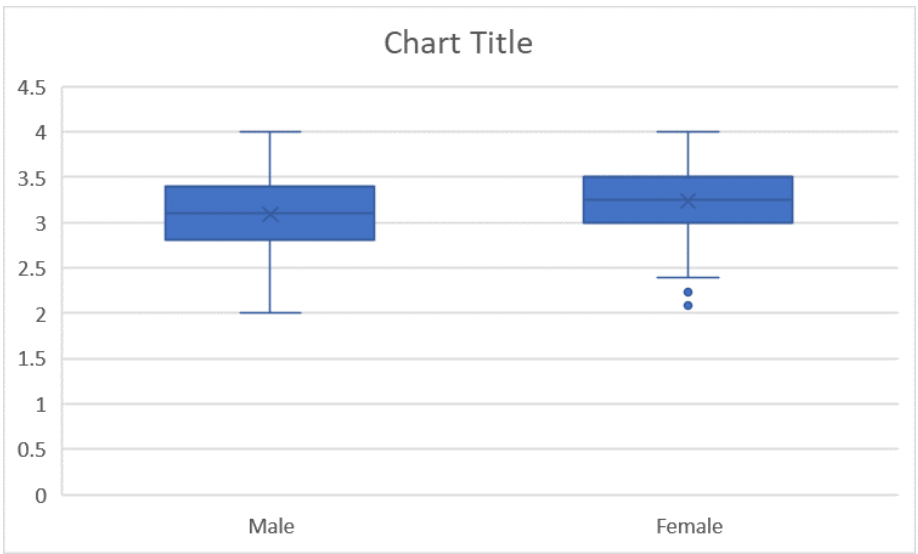
1. Select the *Gender* and *GPA* data in the range **A1:B332**.



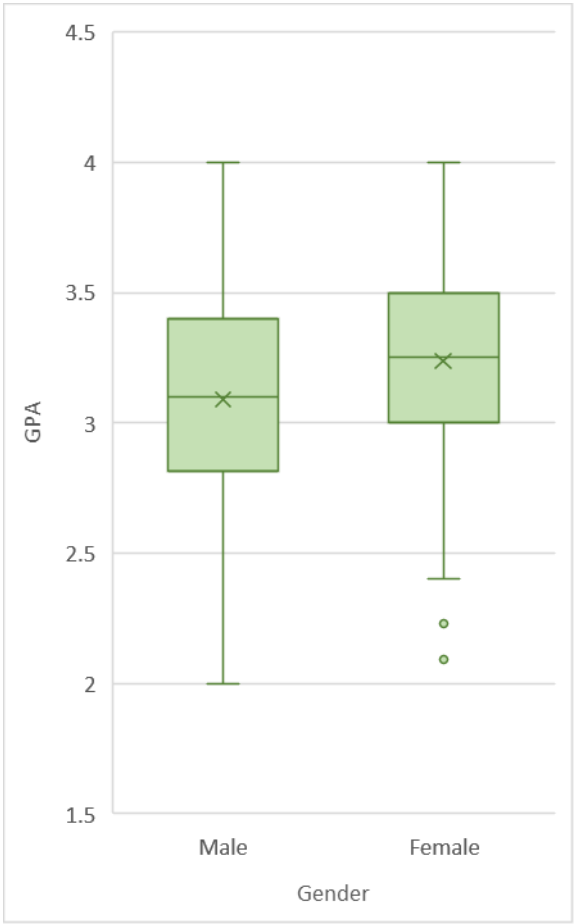
2. Navigate to the **Insert** tab in the Excel ribbon and click on **Insert Statistic Chart**. In the pop-up window that appears, click on the **Box and Whisker Chart**.



3. A boxplot with the *GPA* among *Male* and *Female* students will be inserted on the worksheet.



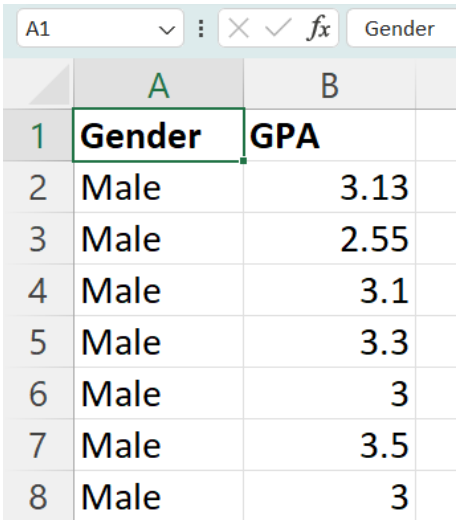
4. The boxplot can be formatted using the **Chart Elements** button on the top right of the chart. The detailed step-by-step guide was already discussed in the Excel manual for Quantitative Data Analysis. After inserting chart title, axes titles, and applying some formatting, the boxplot for the sample data can be obtained as shown below.


4.5
4
3.5
3
2.5
2
1.5
GPA
Male
Female
Gender

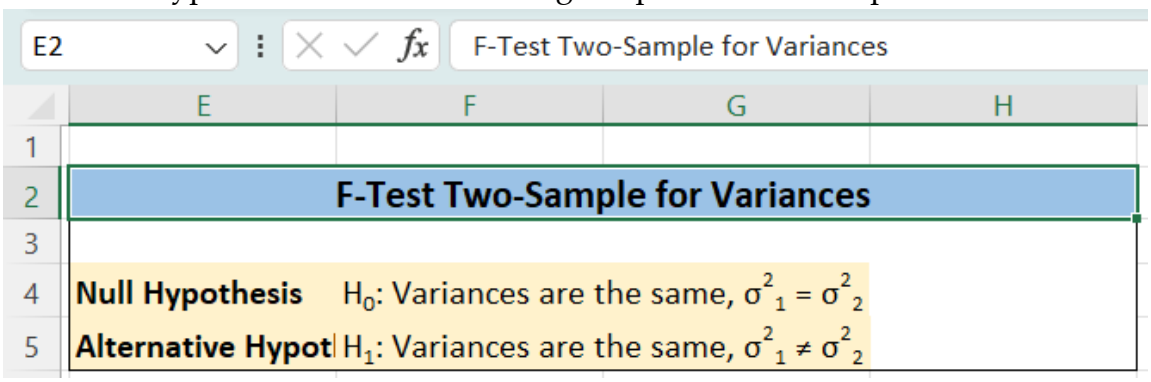
F-Test

Question: Using the entire dataset, is there any difference between the average *GPA* by *Gender*?

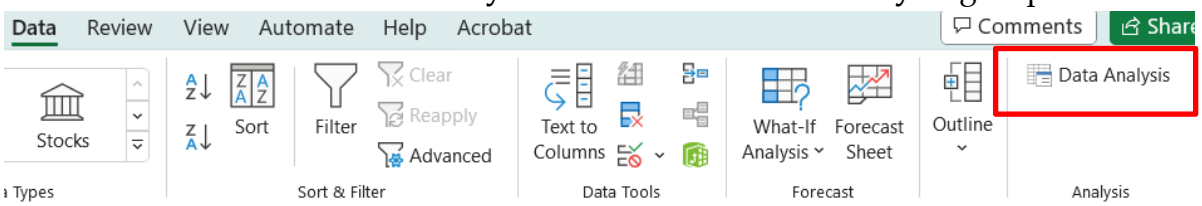
1. Copy and paste the *Gender* and *GPA* data from the StudentSurvey worksheet to a new worksheet. Preferably, paste them in the range **A1:B332**.



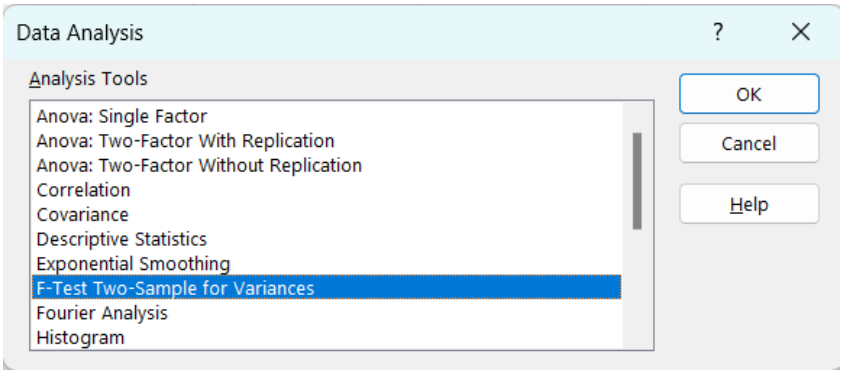
2. In the range **E2:H5**, insert an appropriate tile for the test and include the null and alternative hypothesis. Use the following template as an example.



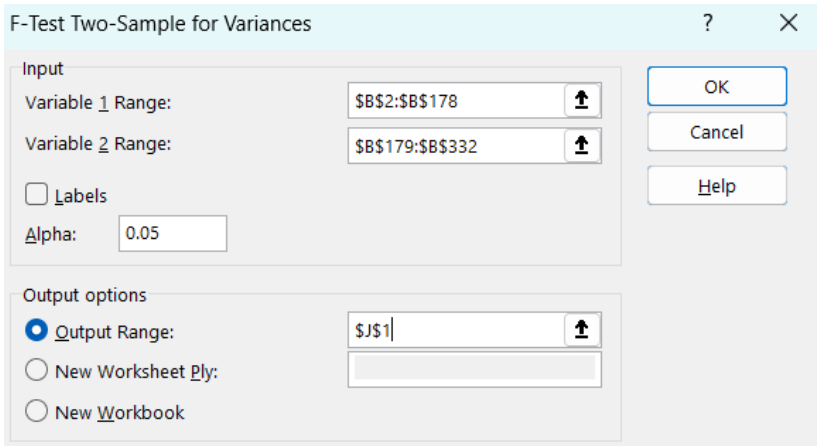
3. To perform the F-test on the sample data, navigate to the **Data** tab on the Excel ribbon and click on the **Data Analysis** tool situated in the **Analysis** group.



4. In the window that appears, select the **F-Test: Two-sample for Variances** and click **OK**.



5. In the next window that appears, the input range of variables, the alpha value and the output location needs to be specified. For **Variable 1 Range**, the *GPA* corresponding to *Male* students need to be selected. The input range will be **B2:B178**. Similarly, for **Variable 2 Range**, the *GPA* corresponding to *Female* students needs to be selected. The data range for this case will be **B179:B332**. Next, the **alpha** value can be inserted. For this demonstration, the **alpha** is set as *0.05*. Finally, one may choose the output range for the results to be pasted. Since a title for the analysis was already created in cell **E2**, the output range can be selected as cell **J1**. Once all these values are selected appropriately, click **OK**.



6. The F-test results will be then pasted in the range **J1:L10**.

F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	3.089492	3.238052
Variance	0.158687	0.148178
Observations	177	154
df	176	153
F	1.07092	
P(F<=f) one-tail	0.332282	
F Critical one-tail	1.295918	

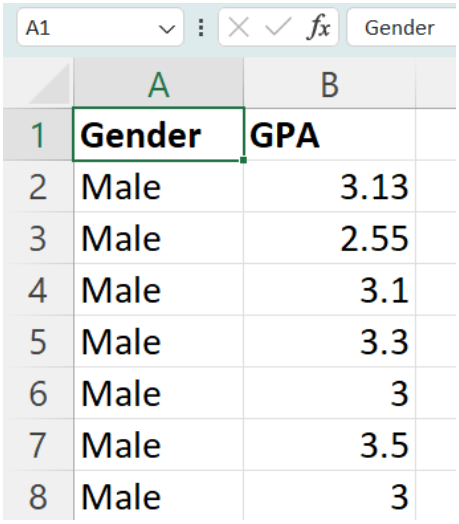
7. One can replace the headings, *Variable 1* and *Variable 2* with *Male* and *Female* respectively. The number formatting can be changed appropriately.

8. Also, one can use a similar template as the one used in the previous section - Hypothesis Testing to include remarks on the results. The results can be formatted to the following style.

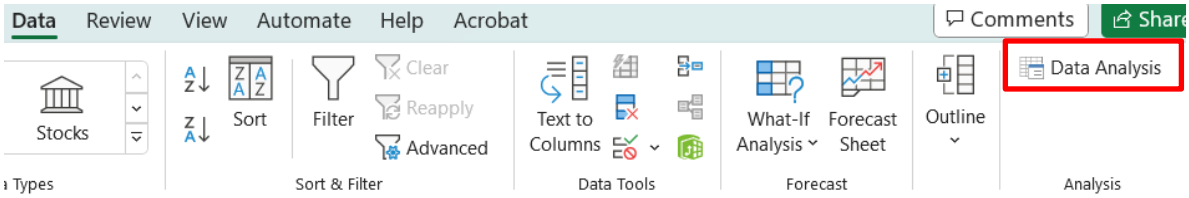
	E	F	G	H
2	**F-Test Two-Sample for Variances**			
3				
4	**Null Hypothesis**	H_0: Variances are the same, $\sigma^2_1 = \sigma^2_2$		
5	**Alternative Hypothesis**	H_1: Variances are the same, $\sigma^2_1 \neq \sigma^2_2$		
6				
7				
8				
9		**Male**	**Female**	**Remarks**
10	Mean	3.09	3.24	
11	Variance	0.16	0.15	
12	Observations	177	154	
13	df	176	153	
14	F	1.07		
15	P(F<=f) one-tail	0.3322824155		*As P>0.05, we fail to reject H_0.*
16	F Critical one-tail	1.30		
17	SD	0.40	0.38	

t-Test

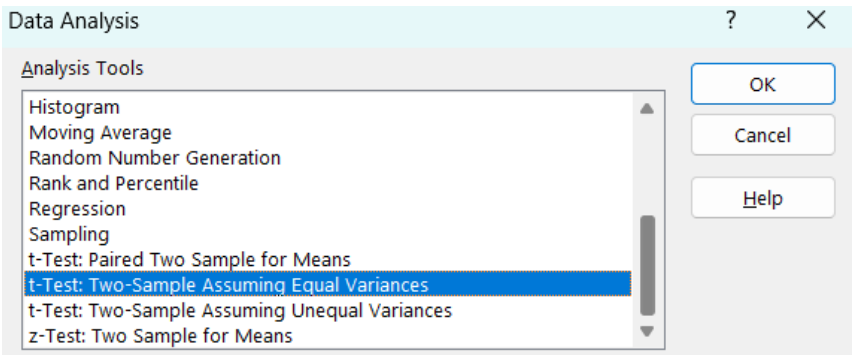
1. Use the *Gender* and *GPA* data in the range **A1:B332** previously copied for the F-test from the StudentSurvey worksheet.



2. To perform the **t-test** on the sample data, navigate to the **Data** tab on the Excel ribbon and click on the **Data Analysis** tool situated in the **Analysis** group.

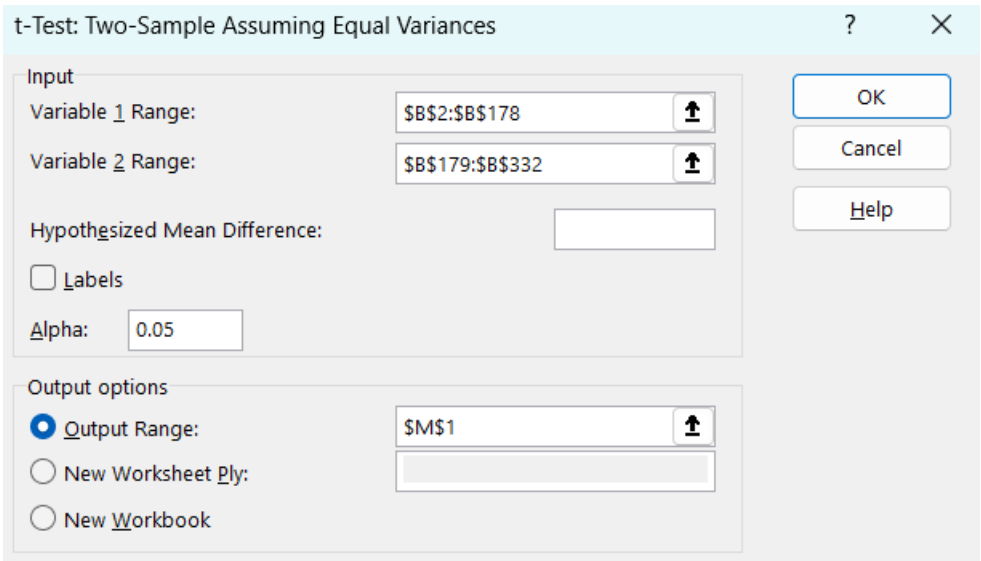


3. In the window that appears, select **t-Test: Two-Sample Assuming Equal Variances** and click **OK**.



4. A new window will appear. The input range of variables, the hypothesized mean difference, the alpha value, and the output location needs to be specified in the new window. For **Variable 1 Range**, the *GPA* corresponding to *Male* students in the range **B2:B178** needs to be selected. Similarly, for **Variable 2 Range**, the *GPA* corresponding to *Female* students in the range **B179:B332** needs to be selected. The **hypothesized mean difference** can be kept as *0*, as the null hypothesis for this demonstration identifies the mean turbidities in two decades has no difference.

Next, the **alpha** value can be inserted. For this demonstration, the **alpha** is set as *0.05*. Finally, one may choose the output range for the results to be pasted. The output range can be selected as cell **M1**. Once all these values are selected appropriately, click **OK**.



5. The t-Test results will be pasted in the range **M1:O13** as shown below.

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	3.089491525	3.2380519
Variance	0.158686672	0.1481779
Observations	177	154
Pooled Variance	0.153799604	
Hypothesized Mean	0	
df	329	
t Stat	-3.43762512	
P(T<=t) one-tail	0.000331108	
t Critical one-tail	1.649498293	
P(T<=t) two-tail	0.000662217	
t Critical two-tail	1.967200683	

6. Like in F-test, one can replace the headings, *Variable 1* and *Variable 2* with *Male* and *Female* respectively. The number formatting can be changed appropriately. The results can be formatted to a presentable format as shown below.

	E	F	G	H
19	t-Test: Two-Sample Assuming Equal Variances			
20		Male	Female	Remarks
21	Mean	3.09	3.24	
22	Variance	0.16	0.15	
23	Observations	177	154	
24	Pooled Variance	0.15		
25	Hypothesized Mean Difference	0		
26	df	329		
27	t Stat	-3.44		
28	P(T<=t) one-tail	3.31E-04		
29	t Critical one-tail	1.65		
30	P(T<=t) two-tail	6.62E-04		
31	t Critical two-tail	1.97		

7. To begin the confidence interval calculations, place captions in the range **E34:H36** as shown below.

Confidence Interval	
Critical t-Value ($t_{\alpha/2}$)	
CI Lower Bound	
CI Upper Bound	

8. The critical t-value is already obtained during the t-test calculations done using the Data Analysis tool. In the current demonstration, the value was placed in cell **F31**. One may copy that to cell **F34**.
9. To calculate the CI Lower Bound in this example, one may use the following formula.

=(F21-G21) - F34*SQRT(F24*((1/F23) + (1/G23)))

10. To calculate the CI Upper Bound, one may make use of the following formula.

=(F21-G21) + F34*SQRT(F24*((1/F23) + (1/G23)))

F36 =(F21-G21)+F34*SQRT(F24*((1/F23)+(1/G23)))

	E	F	G	H
33	Confidence Interval			
34	Critical t-Value ($t_{\alpha/2}$)	1.9672006834		
35	CI Lower Bound	-0.23		
36	CI Upper Bound	-0.06		

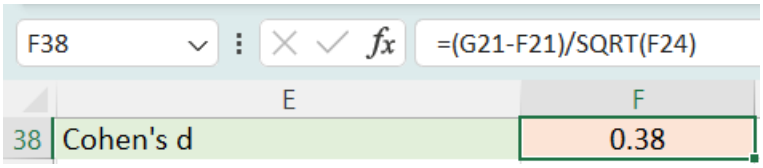
11. After the complete calculation of t-test statistic and the confidence intervals, the table will look like the following. Note that the equation to calculate the CI Lower and Upper Bounds are pasted beside the respective cells for better understanding.

	E	F	G	H
19	t-Test: Two-Sample Assuming Equal Variances			
20		Male	Female	Remarks
21	Mean	3.09	3.24	
22	Variance	0.16	0.15	
23	Observations	177	154	
24	Pooled Variance	0.15		
25	Hypothesized Mean Difference	0		
26	df	329		
27	t Stat	-3.44		
28	P(T<=t) one-tail	3.31E-04		
29	t Critical one-tail	1.65		
30	P(T<=t) two-tail	6.62E-04		
31	t Critical two-tail	1.97		
32				
33	Confidence Interval			
34	Critical t-Value ($t_{\alpha/2}$)	1.97		
35	CI Lower Bound	-0.23		
36	CI Upper Bound	-0.06		

12. Cohen's d can be calculated in cell **F38** by dividing the difference in the Mean *GPA* of Female and Male students, currently placed in cells **G21** and **F21** respectively, by the square root of the Pooled Variance obtained from the t-test in cell **F24**. The following equation can be placed in cell **F38**.

= (G21 - F21)/SQRT(F24)

13. The Cohen's d computed in cell **F38** will look like the following. Note that the number is rounded to two decimal places to be consistent with other values.

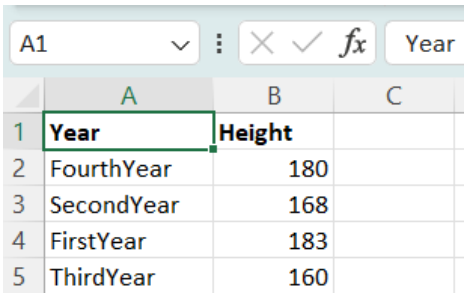


COMPARING THREE OR MORE POPULATION MEANS (ANOVA)

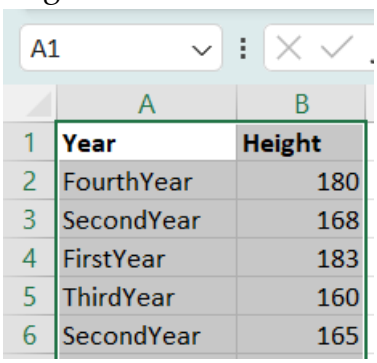
Question: Is the average *Height* of students different by their *Year* of study?

Boxplots

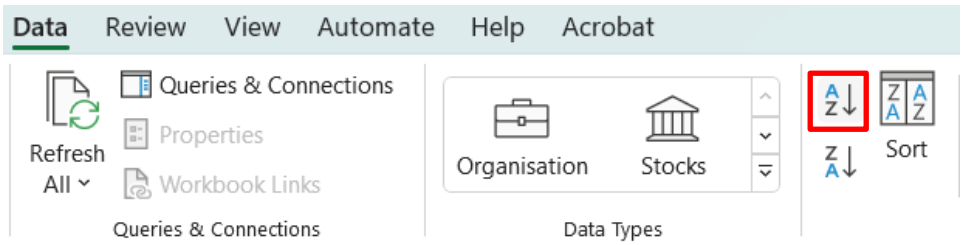
Boxplots can be created in Excel using the Charts function directly. To begin with, copy and paste the *Height* and *Year* data provided in the 'StudentSurvey' worksheet to a new worksheet. Preferably paste the values in the range **A1:B332**.



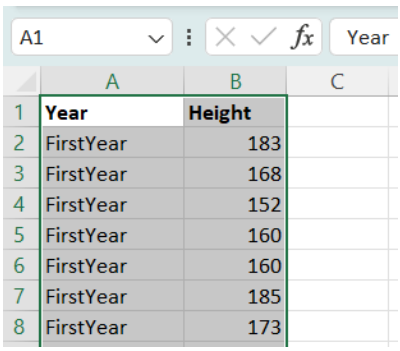
1. Select the entire data in the range **A1:B332**.



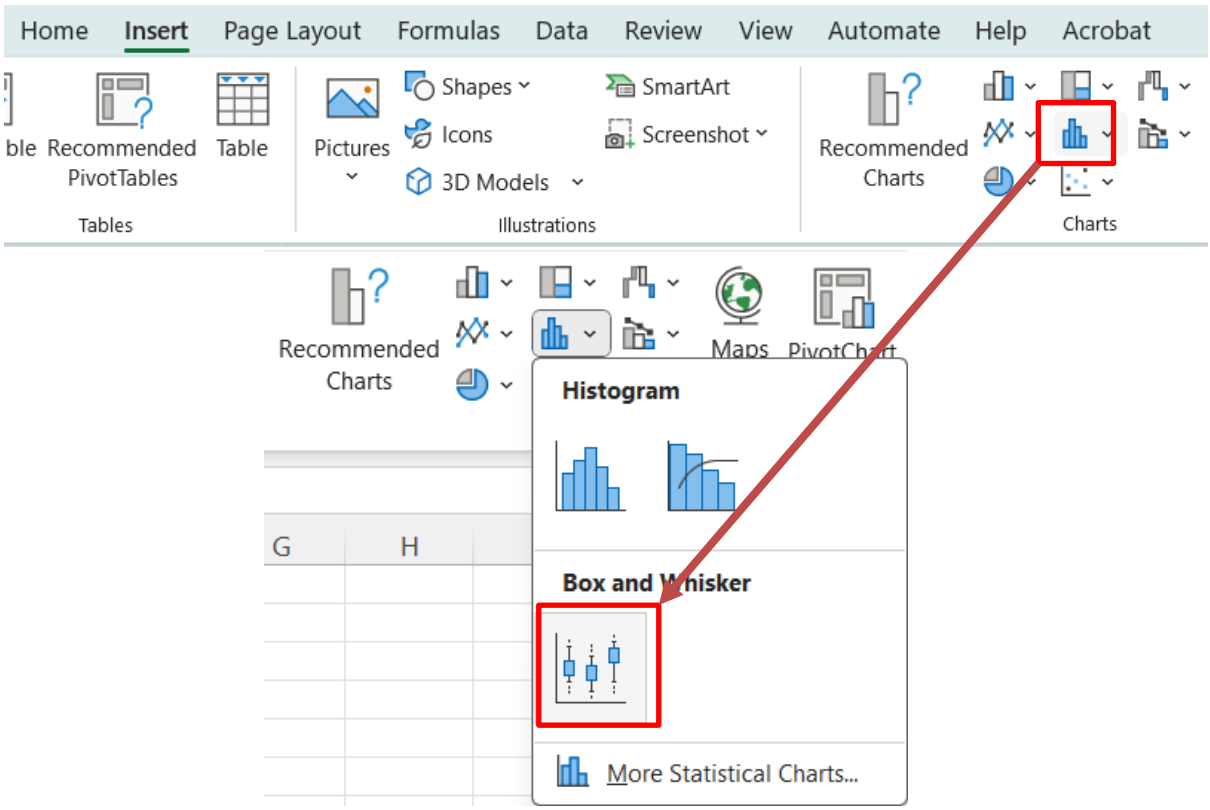
2. Navigate to the **Data** tab in the Excel ribbon and click on **Sort A to Z**.



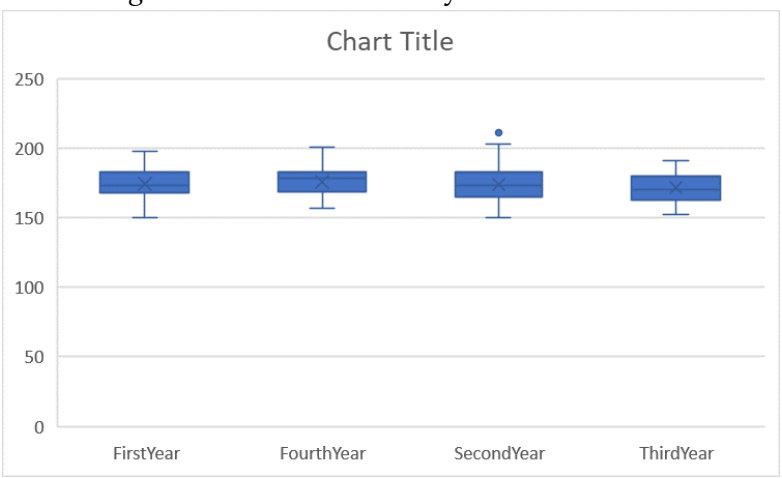
3. The above step will sort the data based on the *Year* in the order of *FirstYear*, *FourthYear*, *SecondYear* and *ThirdYear*. If needed one can click on the **Custom Sort** button under the **Data** tab and specify the order of sort manually.



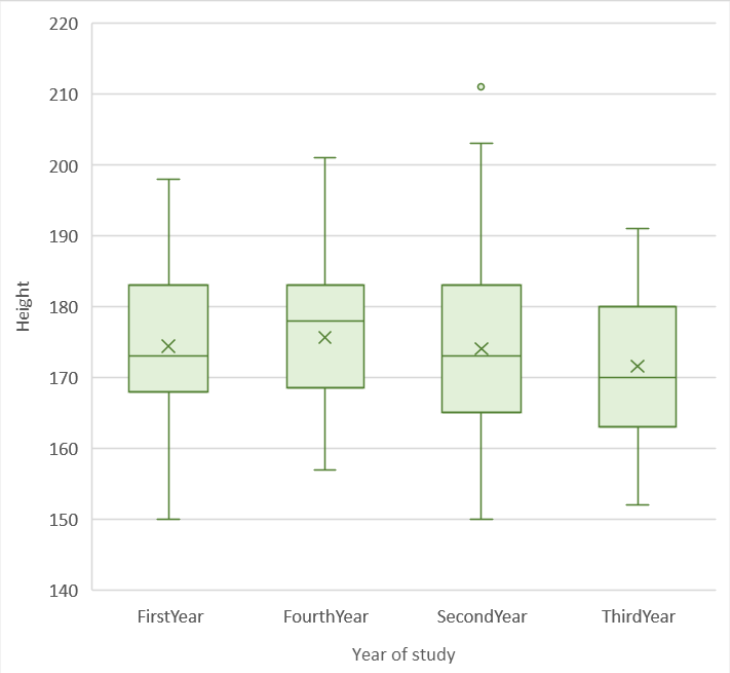
4. Select the entire data. Navigate to the **Insert** tab in the Excel ribbon and click on **Insert Statistic Chart**. In the pop-up window that appears, click on the **Box and Whisker Chart**.



5. A boxplot with the *Height* across *Year* of study will be inserted on the worksheet.



6. The above chart can be formatted using the **Chart Elements** button on the top right of the chart. The detailed step-by-step guide was already discussed in the Excel manual for Quantitative Data Analysis. After inserting chart title, axes titles, and applying some formatting, the boxplot for the sample data can be obtained as shown below.



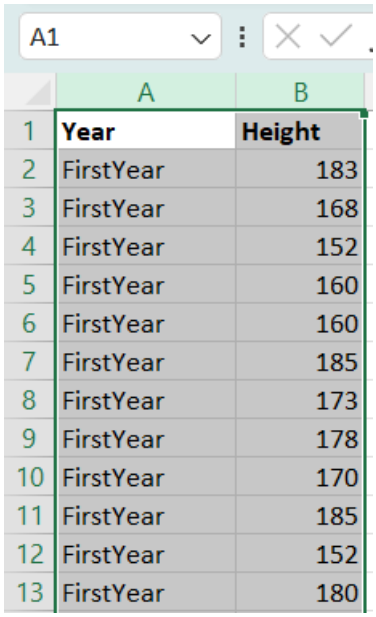
ANOVA

Question 1: Is the average *Height* of students different by their *Year* of study?

The null hypothesis identified for the above question is that there is no difference in the mean *Height* across the *Year* of study.

To perform the ANOVA test in Excel, the user needs to setup the data in the correct format.

1. Copy and paste the sorted *Height* and *Year* of study data used to create the Boxplot into a new worksheet. Preferably paste the values in the range **A1:B332**.

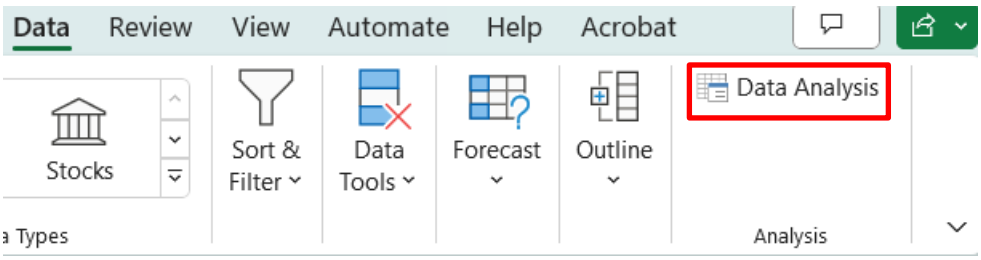


	A	B
1	**Year**	**Height**
2	FirstYear	183
3	FirstYear	168
4	FirstYear	152
5	FirstYear	160
6	FirstYear	160
7	FirstYear	185
8	FirstYear	173
9	FirstYear	178
10	FirstYear	170
11	FirstYear	185
12	FirstYear	152
13	FirstYear	180

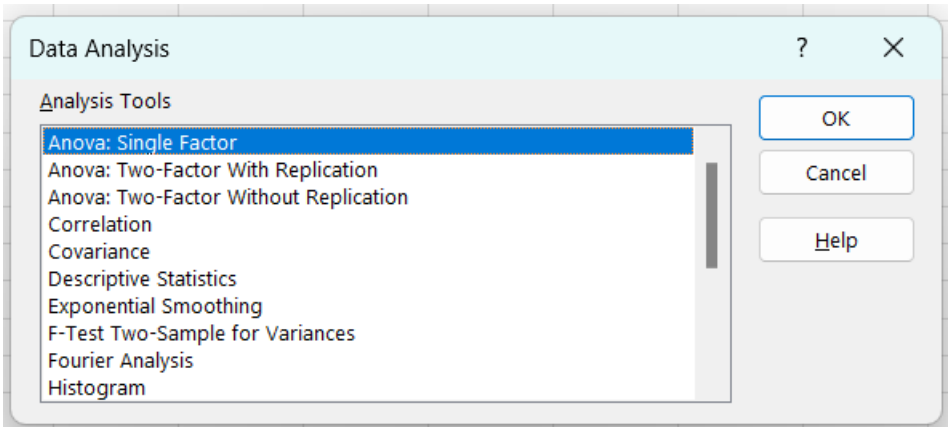
2. In the next step, arrange the *Height* across *Year* of study as shown below. It is preferred that the user pastes the values in the same ranges as shown in the demonstration. Here, the data are pasted in the range **E1:H184**.

	E	F	G	H
1	FirstYear	SecondYear	ThirdYear	FourthYear
2	183	168	160	180
3	168	165	173	183
4	152	165	188	183
5	160	188	157	183
6	160	165	185	183
7	185	183	183	178
8	173	157	165	163
9	178	170	160	178
10	170	165	165	180
11	185	173	170	201
12	152	188	191	170
13	180	178	178	183
14	183	180	155	157

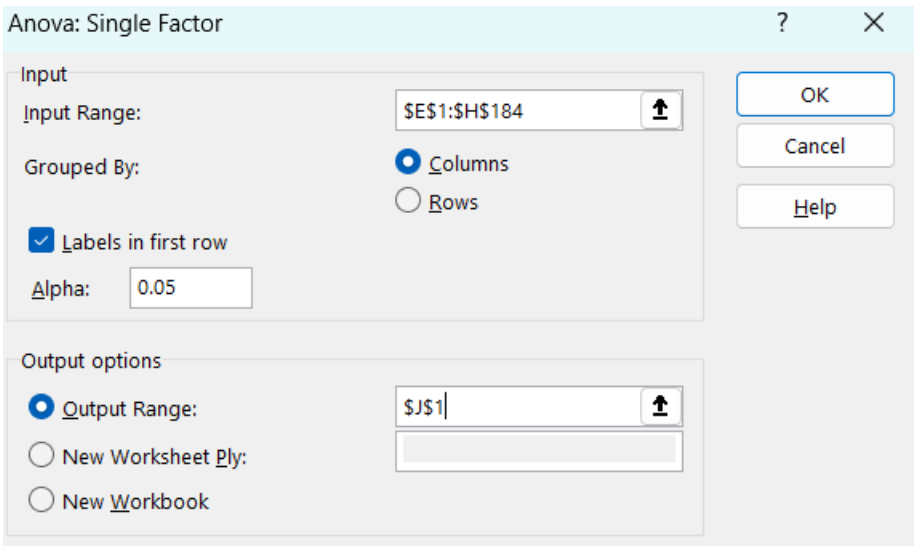
3. Navigate to the **Data** tab on the Excel ribbon and click on the **Data Analysis** tool.



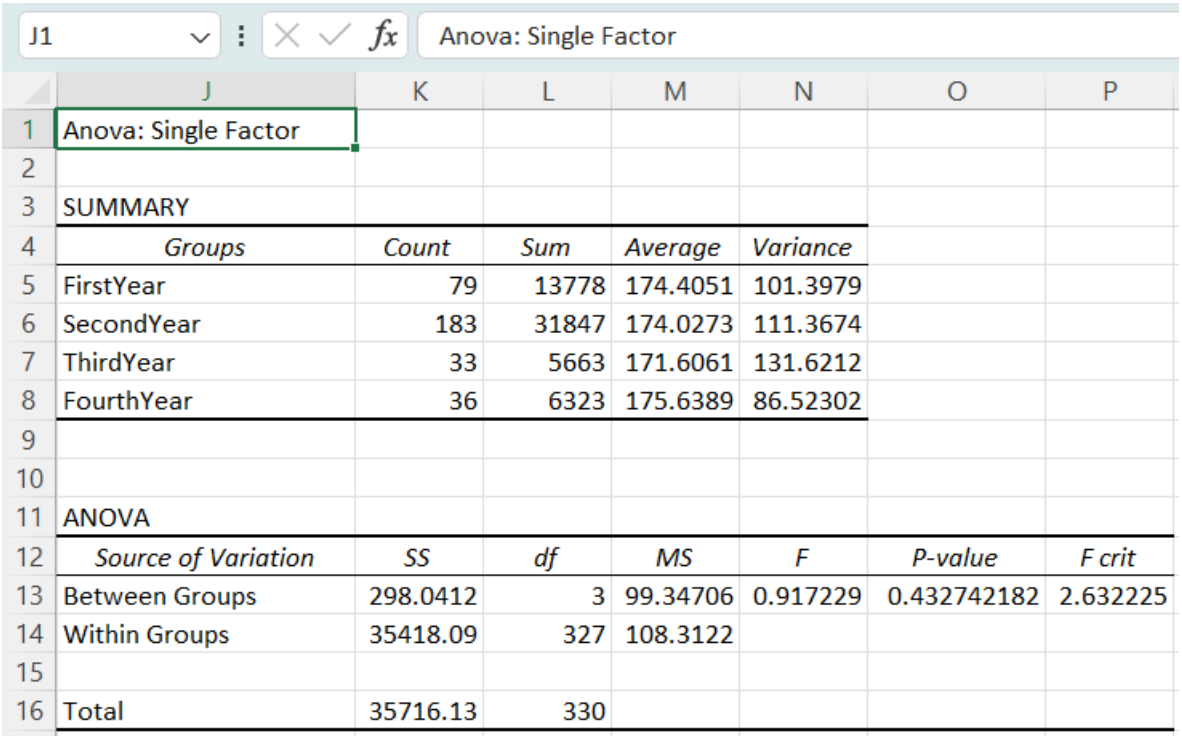
4. Select the **Anova: Single Factor** and click **OK**.



5. In the window that appears, select the input range **E1:H184** where the *Height* based on the *Year* of study are kept. Keep the "Grouped By:" option to "Columns". Check the box beside the "Labels in first row" option. Leave the "Alpha" value as 0.05. Finally, choose the range where you would like to have the output. **J1** is used in this demonstration. Click on **OK**.



6. The ANOVA results will be pasted in the current worksheet starting from the cell **J1**.



J1 | Anova: Single Factor

	J	K	L	M	N	O	P
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	FirstYear	79	13778	174.4051	101.3979		
6	SecondYear	183	31847	174.0273	111.3674		
7	ThirdYear	33	5663	171.6061	131.6212		
8	FourthYear	36	6323	175.6389	86.52302		
9							
10							
11	ANOVA						
12	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
13	Between Groups	298.0412	3	99.34706	0.917229	0.432742182	2.632225
14	Within Groups	35418.09	327	108.3122			
15							
16	Total	35716.13	330				

7. The values that are of interest from the above results are the *Summary*, *F*, *P-value* and the *F-crit*. Apply formatting of your choice and highlight the values. One may follow the following formatting.

	R	S	T	U	V
3	*Groups*	*Count*	*Sum*	*Average*	*Variance*
4	FirstYear	79	13778	174.41	101.40
5	SecondYear	183	31847	174.03	111.37
6	ThirdYear	33	5663	171.61	131.62
7	FourthYear	36	6323	175.64	86.52
8			*F*	*P-value*	*F crit*
9			0.92	0.43	2.63

8. As the p-value is greater than 0.05, the ANOVA is not significant. Hence, we fail to reject the null hypothesis and the post hoc multiple comparison tests are not required.

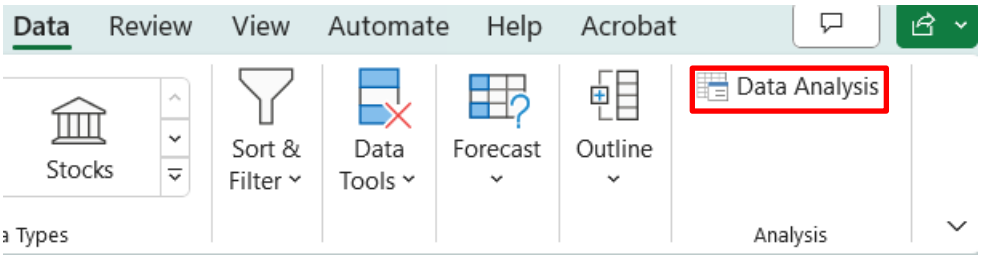
Question 2: Is the average students' *GPA* different for the preferred *Award* students would like to win?

To perform the ANOVA test in Excel, the user needs to setup the data in the correct format.

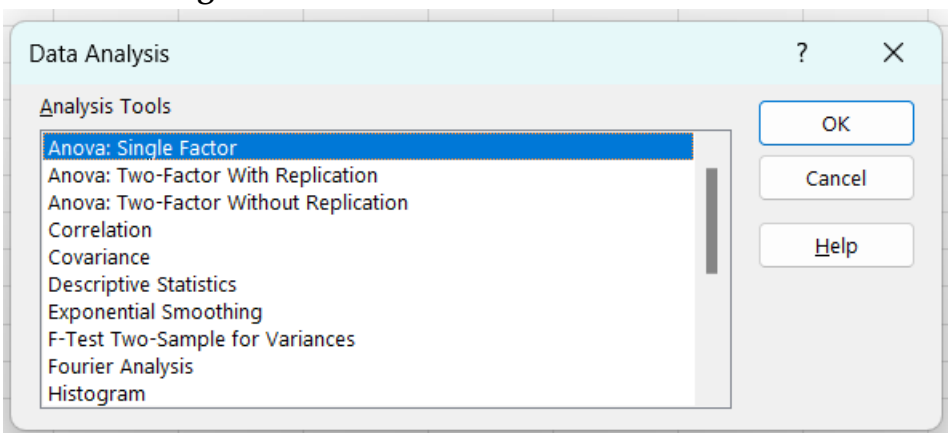
1. Copy and paste the *GPA* data provided in the 'StudentSurvey' worksheet to a new worksheet based on the *Preferred Award* as shown below. Paste the values in the range **A1:C169**.

	A	B	C
1	Academy	Nobel	Olympic
2	2.5	2.55	3.13
3	3.48	3.1	2.77
4	3.5	2.7	3.3
5	2.6	3.2	2.09
6	3.85	3.7	3.86
7	3	3.08	3
8	3.3	3	3.3
9	2.9	3.35	3.5

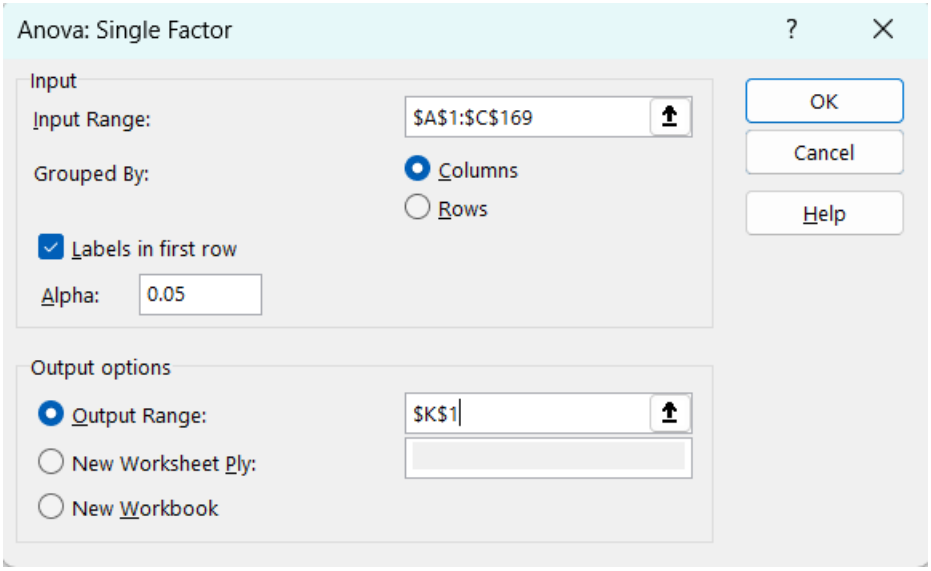
2. Navigate to the Data tab on the Excel ribbon and click on the Data Analysis tool.



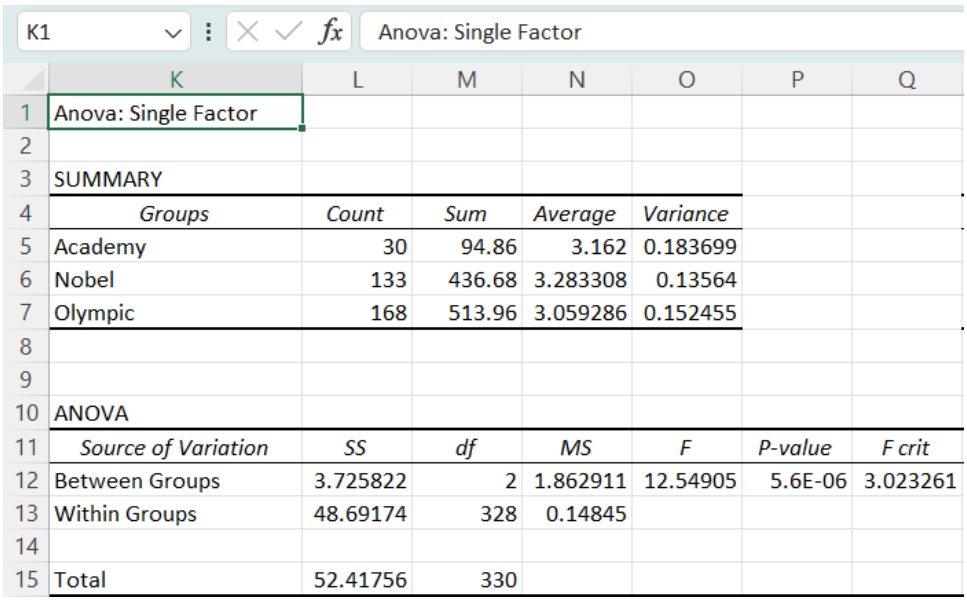
3. Select the **Anova: Single Factor** and click **OK**.



4. In the window that appears, select the input range **A1:C169** where the *GPA* for the three *Preferred Awards* is placed. Keep the "Grouped By:" option to "Columns". Check the box beside the "Labels in first row" option. Leave the "Alpha" value as 0.05. Finally, choose the range where you would like to have the output. **K1** is used in this demonstration. Click on **OK**.



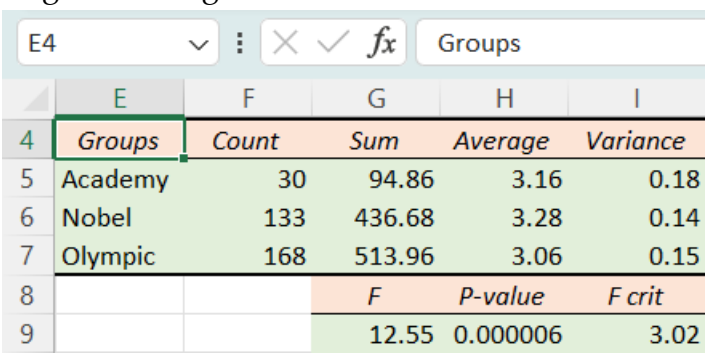
5. The ANOVA results will be pasted in the current worksheet starting from the cell **K1**.



K1 | Anova: Single Factor

	K	L	M	N	O	P	Q
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Academy	30	94.86	3.162	0.183699		
6	Nobel	133	436.68	3.283308	0.13564		
7	Olympic	168	513.96	3.059286	0.152455		
8							
9							
10	ANOVA						
11	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
12	Between Groups	3.725822	2	1.862911	12.54905	5.6E-06	3.023261
13	Within Groups	48.69174	328	0.14845			
14							
15	Total	52.41756	330				

6. The values that are of interest from the above results are the *Summary*, *F*, *P-value* and the *F-crit*. Apply formatting of your choice and highlight the values. One may follow the following formatting.



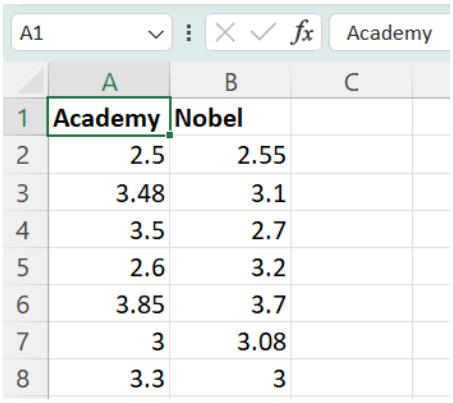
E4 | Groups

	E	F	G	H	I
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*
5	Academy	30	94.86	3.16	0.18
6	Nobel	133	436.68	3.28	0.14
7	Olympic	168	513.96	3.06	0.15
8			*F*	*P-value*	*F crit*
9			12.55	0.000006	3.02

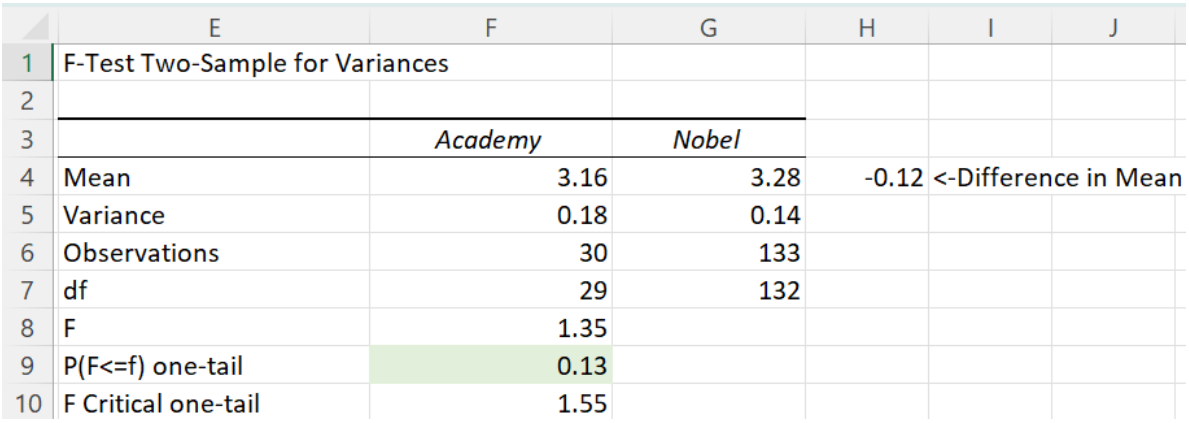
7. As the P-value is smaller than 0.05, the ANOVA is significant. Hence, post hoc multiple comparison tests are required to identify pairwise differences. To do this, one needs to setup worksheets for the following pairs.

Academy vs Nobel	Academy vs Olympic	Nobel vs Olympic

8. To begin with, copy the *GPA* for *Academy* and *Nobel Awards* to a new worksheet named as "Academy vs Nobel".



9. Carry out the F-test: Two-Sample for Variances first. The step-by-step instructions on how to do the F-test is discussed in the subsection - F-Test. In this demonstration, the results of the F-test are placed in the range starting in cell **E1**.



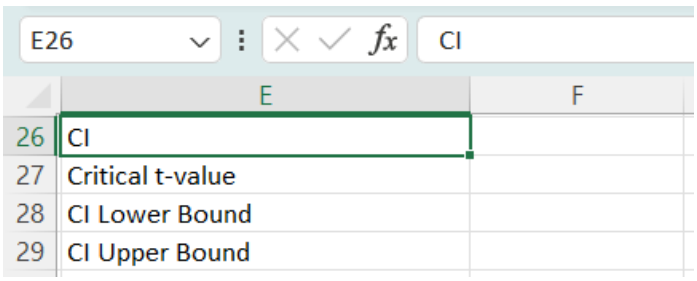
10. One can manually calculate the difference between the Mean of *GPA* of *Academy* and *Nobel Awards* in cell **H4** as shown above.
11. As the P-value is greater than 0.05, one needs to carry out the "t-test: Two-Sample Assuming Equal Variances". The step-by-step instructions on how to perform the "t-test: Two-Sample Assuming Equal Variances" are demonstrated in the subsection - t-Test. The results of the t-Tests are placed in the range starting in cell **E12** in this demonstration.

	E	F	G	H	I
12	t-Test: Two-Sample Assuming Equal Variances				
13					
14		*Academy*	*Nobel*		
15	Mean	3.16	3.28		
16	Variance	0.18	0.14		
17	Observations	30	133		
18	Pooled Variance	0.14			
19	Hypothesized Mean Differ	0			
20	df	161			
21	t Stat	-1.58			
22	P(T<=t) one-tail	0.06			
23	t Critical one-tail	1.65			
24	P(T<=t) two-tail	0.12		Cohen's d	-0.32
25	t Critical two-tail	1.97			

12. Cohen's d can be calculated in cell **I24** by dividing the difference in the Mean, calculated in cell **H4** by the square root of the Pooled Variance obtained from the t-test in cell **F18**. The formula to be placed in cell **I24** will look like the following.

=H4/SQRT(F18)

13. Now, that the critical t-value is obtained and considering that the P-value obtained in the F-test is less than 0.05, one can proceed to confidence interval calculations. Before proceeding to the calculations, suitable table can be setup in the range **E26:F29** as shown below. Note that the range in which the table is setup is important as the cell references affect the calculations.



14. The "critical t-value" is nothing, but the "t Critical two-tail value" obtained after the t-test. Place **=F25** in cell **F27** to copy the value.
15. The CI Lower Bound can be calculated using the following equation in cell **F28**.

= (F15 - G15) - F27*SQRT(F18*((1/F17) + (1/G17)))

16. Similarly, the CI Upper Bound can be calculated using the following equation in cell **F29**.

= (F15 - G15) + F27*SQRT(F18*((1/F17) + (1/G17)))

17. After both the calculations, the following results on CI can be obtained for the *GPA* of the *Academy* vs *Nobel Awards*.

	E	F
26	CI	
27	Critical t-value	1.97
28	CI Lower Bound	-0.27
29	CI Upper Bound	0.03

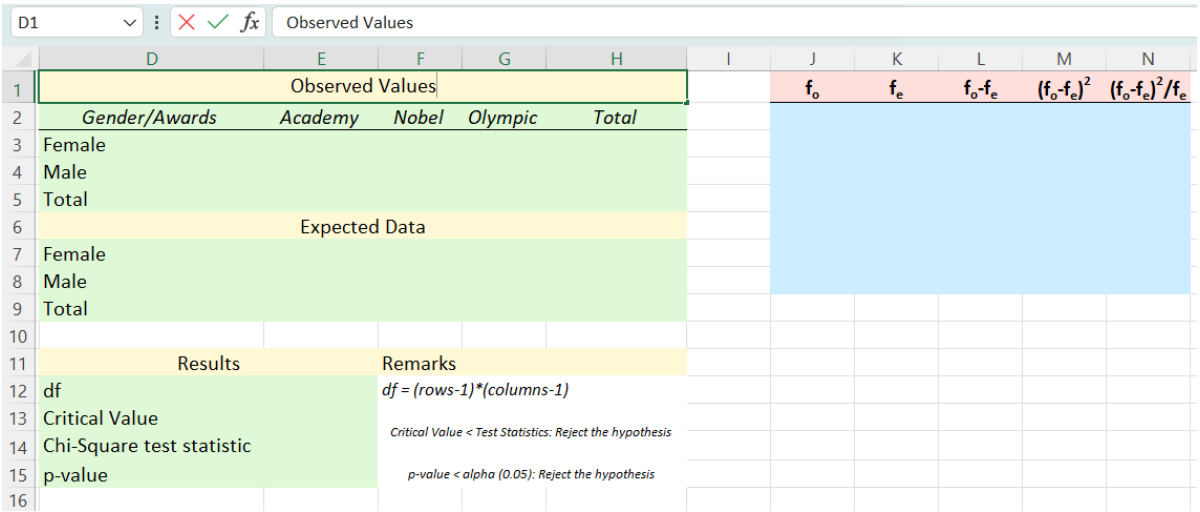
18. One can repeat the calculations for the remaining 2 pairs. However, note that if the P-value obtained in the F-test is less than 0.05, then one needs to carry out "t-test: Two-Sample Assuming Unequal Variances". Also, the equations to calculate the CI Lower Bound and Upper Bound will change as the pooled variance needs to be included in their calculations.

CHI SQUARE TEST

Question: Does preferred *Award* depend on *Gender*?

To carry out the Chi Square test, one needs to setup two tables – one for observed data and another for expected data.

1. To minimise the steps involved, one can make use of 'Template' as shown below. The cell references are vital in Chi-Square test calculations. Hence, pay utmost attention to the cell addresses in the equations if the table is moved elsewhere.



2. Copy the observed data from the two-way table on preferred *Award* and *Gender* that was created during the categorical data analysis and paste it to the observed data table in the worksheet. The two-way table already created for previous analysis is given below.

Gender/Awards	Academy	Nobel	Olympic
Female	19	68	67
Male	11	65	101
Grand Total	30	133	168

3. Once pasted, the observed data table will look like the following.

D	E	F	G	H
	Observed Values			
Gender/Awards	*Academy*	*Nobel*	*Olympic*	*Total*
Female	19.00	68.00	67.00	154.00
Male	11.00	65.00	101.00	177.00
Total	30.00	133.00	168.00	331.00

4. In the next step, one needs to calculate the expected values within the expected data table. To begin with, navigate to cell **E8**. The expected value of the *Academy Award* for *Female* students can be calculated by dividing the product of the total number of *Academy Awards* and total number of *Female* students by the total number of

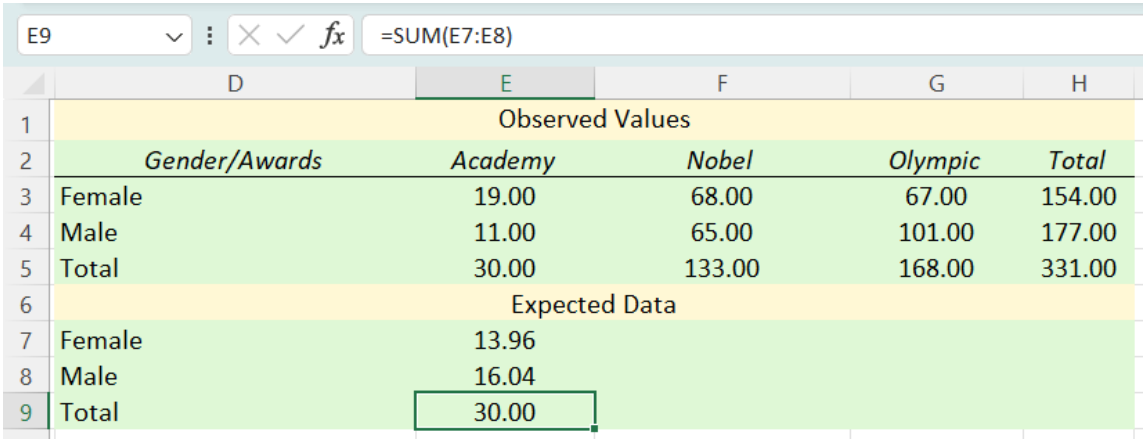
students. In cell **E7**, the following formula can be pasted to get the above expected value.

=E5*$H$3/$H$5

5. Similarly, in cell **E8**, the expected value of the *Academy Award* for the *Male* students can be calculated using the following formula.

=E5*$H$4/$H$5

6. Once these two values are calculated, the table will look like the following.

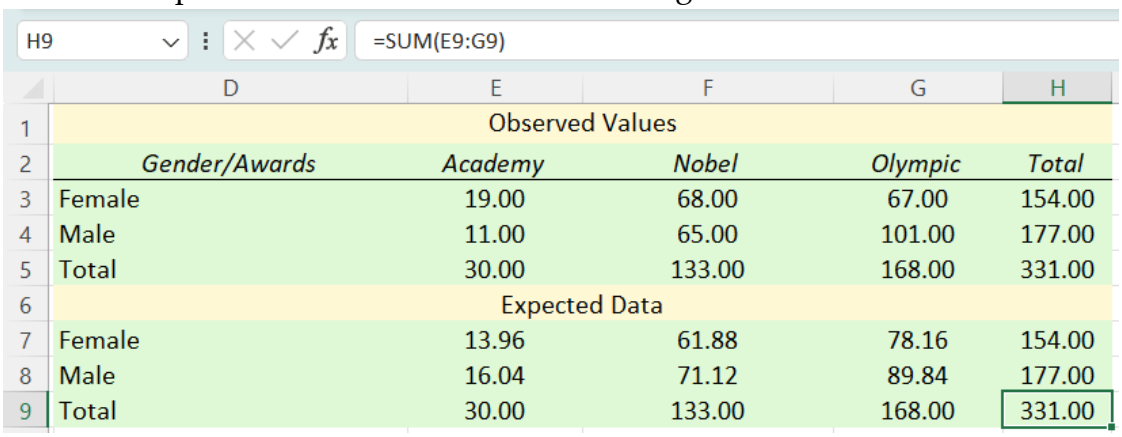


	D	E	F	G	H
1		Observed Values			
2	*Gender/Awards*	*Academy*	*Nobel*	*Olympic*	*Total*
3	Female	19.00	68.00	67.00	154.00
4	Male	11.00	65.00	101.00	177.00
5	Total	30.00	133.00	168.00	331.00
6		Expected Data			
7	Female	13.96			
8	Male	16.04			
9	Total	30.00			

7. The total row can be filled using the SUM function in Excel. In cell **E9**, one can use the following formula to find the sum of all the compliant data for the three treatment methods.

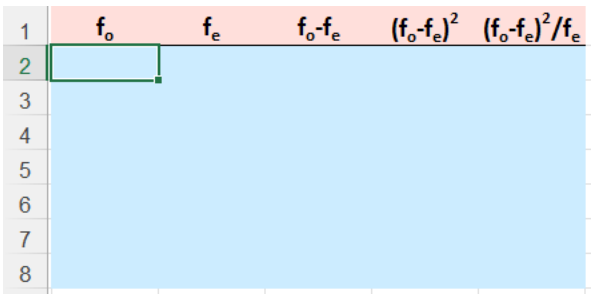
=SUM(E7:E8)

8. In the next step, one can repeat the same calculation for the *Nobel* and *Olympic Awards*. If the formulas mentioned in steps 4 and 5 are used correctly, one can us the AutoFill option in Excel to fill in the remaining four cells.

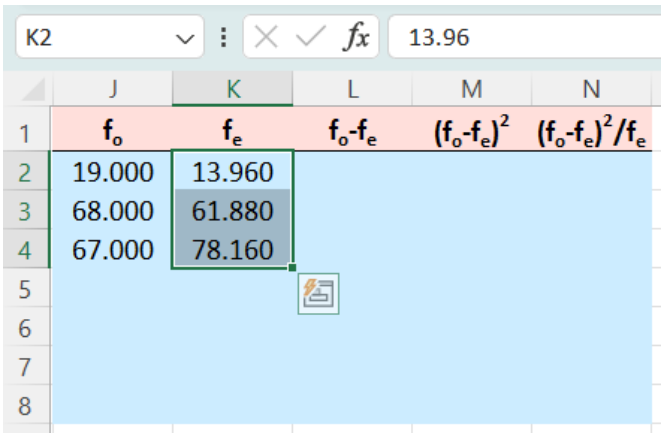


	D	E	F	G	H
1		Observed Values			
2	*Gender/Awards*	*Academy*	*Nobel*	*Olympic*	*Total*
3	Female	19.00	68.00	67.00	154.00
4	Male	11.00	65.00	101.00	177.00
5	Total	30.00	133.00	168.00	331.00
6		Expected Data			
7	Female	13.96	61.88	78.16	154.00
8	Male	16.04	71.12	89.84	177.00
9	Total	30.00	133.00	168.00	331.00

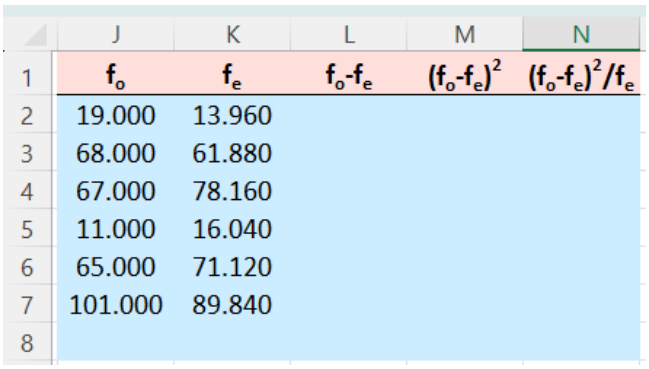
9. Up next, one needs to create another table for the calculation of Chi-Square test statistic. To do that, one can use the following empty table provided in the 'Template' as in step 1.



10. In the above table, f_0 and f_e are the observed and expected values respectively. One can copy the values from the observed data table and the expected data table and paste it here. However, utmost attention should be paid while this step is carried out. First, one can copy all the compliant data values for the three *Awards* corresponding to *Female* students to cells **J2:J4** from the observed data table. Note that only the values need to be copied and not the totals. Similarly, copy all the expected compliant data values for the three treatment methods to cells **K2:K4** from the expected data table.

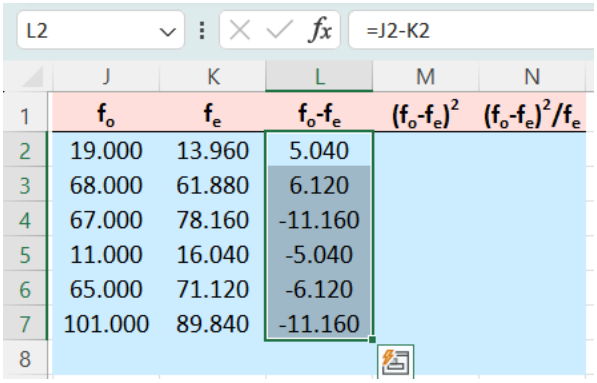


11. In the next step, copy the *Awards* values corresponding to *Male* students from the observed and expected data tables to the current table.



12. In the next column 'f_o-f_e', subtract the expected values from the corresponding observed data values. One can use the following formula in cell **L2** and make use of the AutoFill feature in Excel to fill in the rest of the rows.

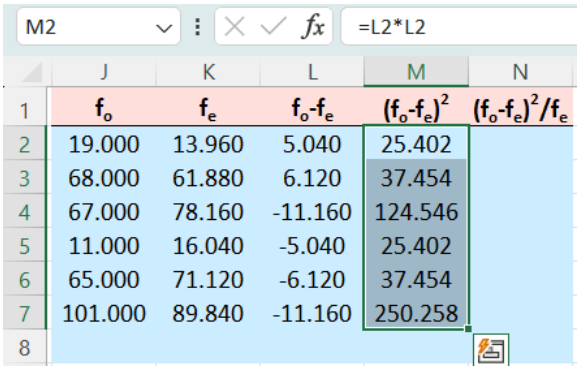
= J2 - K2



	J	K	L	M	N
1	f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
2	19.000	13.960	5.040		
3	68.000	61.880	6.120		
4	67.000	78.160	-11.160		
5	11.000	16.040	-5.040		
6	65.000	71.120	-6.120		
7	101.000	89.840	-11.160		
8					

13. To calculate the values in the next column '$(f_o-f_e)^2$', one can use the following formula in cell **M2** and use the AutoFill function in Excel to calculate the rest of the values in the column.

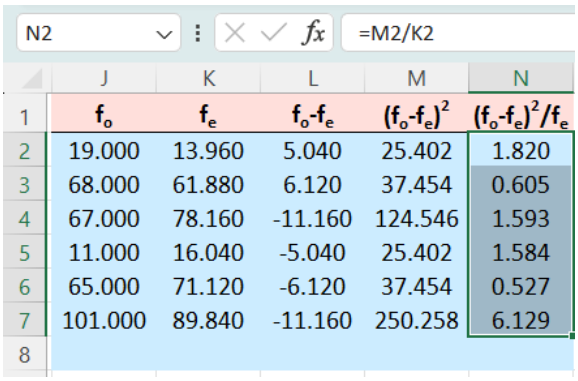
=L2*L2



	J	K	L	M	N
1	f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
2	19.000	13.960	5.040	25.402	
3	68.000	61.880	6.120	37.454	
4	67.000	78.160	-11.160	124.546	
5	11.000	16.040	-5.040	25.402	
6	65.000	71.120	-6.120	37.454	
7	101.000	89.840	-11.160	250.258	
8					

14. The last column '$(f_o-f_e)^2/f_e$', can be filled using the similar method as discussed in the previous two steps. Insert the following formula in cell **N2** and use the AutoFill function to fill in the rest of the values in the column.

=M2/K2



	J	K	L	M	N
1	f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
2	19.000	13.960	5.040	25.402	1.820
3	68.000	61.880	6.120	37.454	0.605
4	67.000	78.160	-11.160	124.546	1.593
5	11.000	16.040	-5.040	25.402	1.584
6	65.000	71.120	-6.120	37.454	0.527
7	101.000	89.840	-11.160	250.258	6.129
8					

15. The last row of the table can be used to fill in the total values of each of the columns. To do this, use the following formula in cell **J8** and use the AutoFill function in Excel to repeat the calculation in cells **K8:N8**. The value in cell **M8** is the Chi-Square test statistic value.

=SUM(J2:J7)

J8 =SUM(J2:J7)

	J	K	L	M	N
1	f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
2	19.000	13.960	5.040	25.402	1.820
3	68.000	61.880	6.120	37.454	0.605
4	67.000	78.160	-11.160	124.546	1.593
5	11.000	16.040	-5.040	25.402	1.584
6	65.000	71.120	-6.120	37.454	0.527
7	101.000	89.840	-11.160	250.258	6.129
8	331.000	331.000	-22.320	500.515	12.257

16. At this point, one has all the values to calculate the critical value and the p-value. To complete the calculations, one can use the last table provided in the 'Template' worksheet.

	D	E	F	G	H
11	Results		Remarks		
12	df		*df = (rows-1)*(columns-1)*		
13	Critical Value		*Critical Value < Test Statistics: Reject the hypothesis*		
14	Chi-Square test statistic				
15	p-value		*p-value < alpha (0.05): Reject the hypothesis*		

17. In the table, the first value to calculate is the degree of freedom (df). It is nothing but the (rows in observed data table – 1)*(columns in observed data table – 1). In this example, it will be **(3-1)*(2-1) = 2**.
18. To calculate the critical value, one needs the alpha and the degree of freedom that was calculated in the previous step. Alpha for this example is chosen to be 0.05. In cell **E13**, one can use the following formula to calculate the critical value.

=CHIINV(0.05,2)

E13 =CHIINV(0.05,2)

	D	E
11	Results	
12	df	2
13	Critical Value	5.991
14	Chi-Square test statistic	
15	p-value	

19. The Chi-Square test statistic was already calculated in step 15. One can copy the value from the second table to **E14** or insert **=N8** in **E14** to copy the value in cell **N8** to **E14.**
20. To calculate the p-value, one needs to use the values in the observed and expected data tables. In cell **E15**, use the following formula to calculate the p-value.

=CHISQ.TEST(E3:G4, E7:G8)

21. The completed tables will look like the following.

	D	E	F	G	H	I	J	K	L	M	N
1	Observed Values						f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
2	*Gender/Awards*	*Academy*	*Nobel*	*Olympic*	*Total*		19.000	13.960	5.040	25.402	1.820
3	Female	19.00	68.00	67.00	154.00		68.000	61.880	6.120	37.454	0.605
4	Male	11.00	65.00	101.00	177.00		67.000	78.160	-11.160	124.546	1.593
5	Total	30.00	133.00	168.00	331.00		11.000	16.040	-5.040	25.402	1.584
6	Expected Data						65.000	71.120	-6.120	37.454	0.527
7	Female	13.96	61.88	78.16	154.00		101.000	89.840	-11.160	250.258	6.129
8	Male	16.04	71.12	89.84	177.00		331.000	331.000	-22.320	500.515	12.257
9	Total	30.00	133.00	168.00	331.00						
10											
11	Results		Remarks								
12	df	2	*df = (rows-1)*(columns-1)*								
13	Critical Value	5.991	*Critical Value < Test Statistics: Reject the hypothesis*								
14	Chi-Square test statistic	12.257									
15	p-value	0.023	*p-value < alpha (0.05): Reject the hypothesis*								

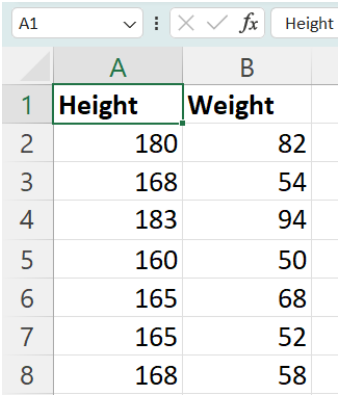
LINEAR REGRESSION TEST

Question: Do *Weights* of the students really depend on their *Heights*?

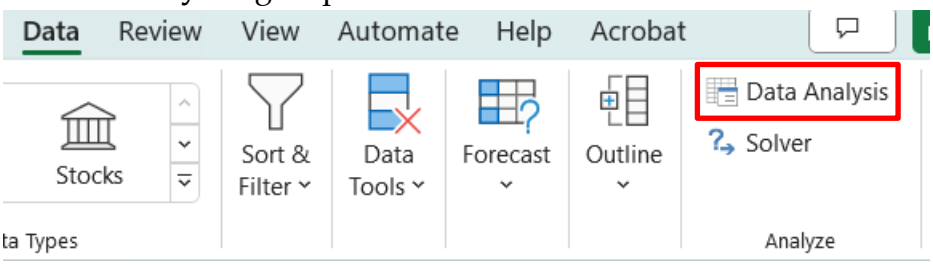
The steps to insert scatter plot containing the regression line and equation were already discussed in quantitative data analysis manual.

To carry out the regression statistics, one needs to use the *Height* and *Weight* data that is available in the worksheet named 'StudentSurvey'.

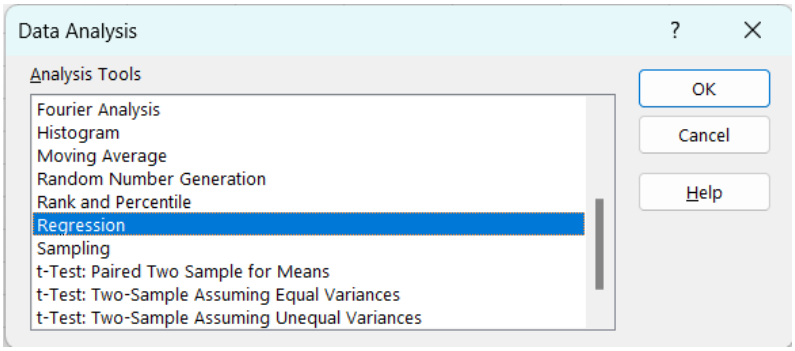
1. Copy the Height and Weight data from the worksheet named 'StudentSurvey' and paste it to a new worksheet preferably in the range **A1:B332**.



2. Navigate to the "**Data**" tab in the Excel ribbon and click on the "**Data Analysis**" tool under the "**Analyze**" group.

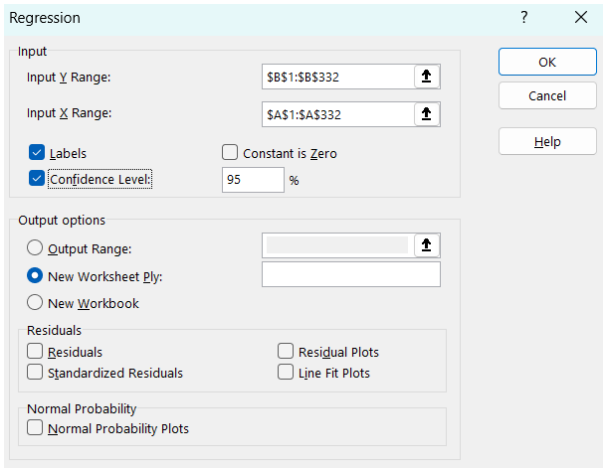


3. In the window that appears, select the "**Regression**" option and click on "**OK**".



4. A new window will appear. In the window that appears, input the Y and X ranges appropriately. If the data is pasted in the range **A1:B332**, the Input Y Range will be **B1:B332** and the Input X Range will be **A1:A332** including labels. The Confidence

Level can be set to 95%. Under the output options, set the output location to New Worksheet Ply and leave the rest of the options to the default values. Click **OK**.

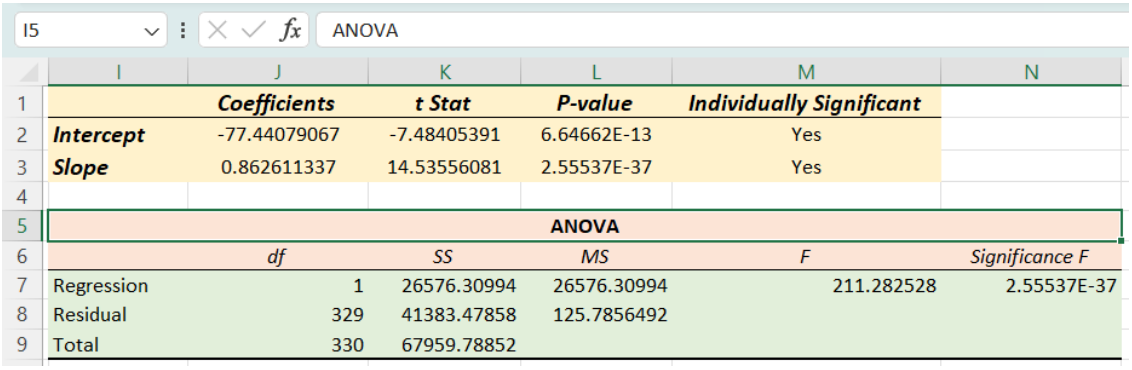


5. In a new worksheet, the following output will be displayed.

A1 fx SUMMARY OUTPUT

	A	B	C	D	E	F	G	H	I
1	SUMMARY	OUTPUT							
2									
3	*Regression Statistics*								
4	Multiple R	0.625347							
5	R Square	0.391059							
6	Adjusted R	0.389208							
7	Standard E	11.21542							
8	Observatic	331							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*ignificance F*			
12	Regressior	1	26576.31	26576.31	211.2825	2.56E-37			
13	Residual	329	41383.48	125.7856					
14	Total	330	67959.79						
15									
16		*Coefficients*	*tandard Erı*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
17	Intercept	-77.4408	10.34744	-7.48405	6.65E-13	-97.7963	-57.0853	-97.7963	-57.0853
18	Height	0.862611	0.059345	14.53556	2.56E-37	0.745868	0.979355	0.745868	0.979355

6. Among the obtained results, for the current exercise, the following data are only required.
    a. ANOVA
    b. Coefficients, t Stat and P-value.
7. One can copy and paste the Coefficients, t Stat and P-value alone to a different location in the worksheet and present them as shown below.



	I	J	K	L	M	N
1		***Coefficients***	***t Stat***	***P-value***	***Individually Significant***	
2	***Intercept***	-77.44079067	-7.48405391	6.64662E-13	Yes	
3	***Slope***	0.862611337	14.53556081	2.55537E-37	Yes	
4						
5			**ANOVA**			
6		*df*	*SS*	*MS*	*F*	*Significance F*
7	Regression	1	26576.30994	26576.30994	211.282528	2.55537E-37
8	Residual	329	41383.47858	125.7856492		
9	Total	330	67959.78852			

8. The number of decimal places can be adjusted to obtain the final tables as the following.

	Coefficients	***t Stat***	***P-value***	***Individually Significant***	
Intercept	-77.44	-7.48	6.64662E-13	Yes	
Slope	0.86	14.54	2.55537E-37	Yes	
		ANOVA			
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	26576.31	26576.31	211.28	2.55537E-37
Residual	329	41383.48	125.79		
Total	330	67959.79			

This material is part of our collection of resources that have been developed from over a decade of teaching introductory statistics and data analysis to undergraduate students at ATU Sligo and various Irish tertiary institutions, as well as providing Microsoft Excel classes and trainings as part of our Information Technology Modules. It assumes the use of computer software, particularly recommending MS Excel for beginners before progressing to more advanced statistical tools like SPSS or R. As such, we have used Excel to implement the various analysis and statistical tests.

Primarily aimed at undergraduate modules in science, health, business, and related disciplines that leverage MS Excel for data analysis, our materials are intended to be comprehensive resources. Accompanying these resources is a dedicated repository at https://thewee.link/NTUTORR-PBL-Resources, offering a lot of useful resources for both students and instructors.

Dr. Akinlolu Akande is a Lecturer in Mathematics and Information Technology in the Faculty of Science at Atlantic Technological University (ATU) Sligo, Ireland. He is a Senior Fellow of the Higher Education Academy (SFHEA), a recognition of his expertise in teaching and learning in higher education.

Dr. Syam Kumar R. is a Lecturer in Mathematics and Information Technology in the Faculty of Science at ATU Sligo, Ireland. He is committed to training and teaching students in the faculty, providing valuable support to students and colleagues.

Dr. David Obada is a Postdoctoral Fellow at the ATU Sligo, Ireland. He was also a research and teaching fellow at the Massachusetts Institute of Technology (MIT) in the USA and holds a Kaufmann Teaching Certificate from MIT.

Ollscoil Teicneolaíochta an Atlantaigh
Atlantic Technological University
NTUTORR Transforming Learning