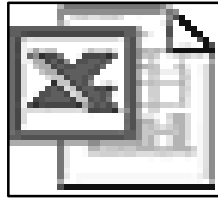


Mathematics LBS 5

Spreadsheet Mathematics: Statistics and Graphing

Lab 9: Scatter Diagrams and Lines of Best Fit



Microsoft Excel Logo and all screens captured by permission of Microsoft

Goal To produce scatter diagrams and lines of best fit by using an Excel worksheet and Chart Wizard .

Learning Outcomes After completing this section, the student should be able to

- enter data in an Excel worksheet
- use Chart Wizard to construct scatter diagrams and lines of best fit
- save the worksheets and the charts
- print the worksheets and charts
- complete several practice assignments

Prior Learning Before attempting this lab, it is expected that the student has

- completed **Spreadsheet Mathematics Lab 1: Entering, Formatting and Saving data in an Excel Spreadsheet**

Contents

- enter data in an **Excel Spreadsheet** ,create scatter plots and lines of best fit chart using **Chart Wizard**
- complete two practice assignments
- complete a hand-in assignment
- discuss questions or concerns with assignments with your teacher

Spreadsheet Math

Lab 8: Constructing Scatter Diagrams and Lines of Best Fit

New Words: scatter diagram, scatter plot, XY plot, variables, trend line.

Scatter plots are similar to line graphs. Data points are plotted on the horizontal (x) axis and the vertical (y) axis of a graph. **Scatter plots** show how much one variable (y) is affected by another variable (x). The relationship between two variables is called their correlation .

It is ideal to have large samples of data in order to make a scatter plot. The closer the plotted data points are on the graph when plotted to making a straight line, then the higher the correlation between the two variables. This shows a strong relationship between the two variables.

(Source: <http://www.mste.uiuc.edu/courses/ci330ms/youtsey/scatterinfo.html>)

In the data chart below, we can see that the weekly take home pay seems to depend on the number of hours worked in the week.

Hours Worked can be plotted on a graph as the independent variable on the **x-axis** of the graph.

Take Home Pay can be plotted on the same graph as the dependent variable on the **y-axis**.

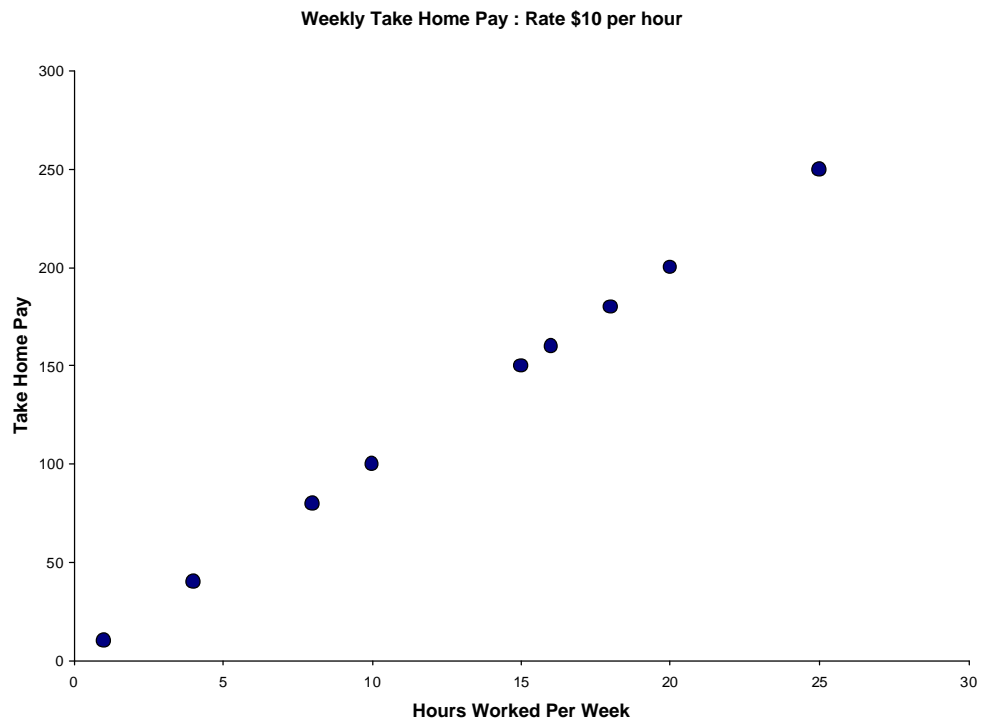
The points on the graph are called an **XY plot** or a **scatter plot**.

The **line of best fit** or the **trend line** is a line on a scatter plot which can be drawn near the points to clearly show the trend between two sets of data.

Data on Weekly Take Home Pay : Rate \$10 per hour

Hours Worked per Week	Weekly Take Home Pay
1	10
4	40
8	80
10	100
15	150
16	160
18	180
20	200
25	250

We can make a graph of this data by plotting the data on the x and y axis . You can see that the plotted points on the graph seem to be in a straight line ascending up to the right. This indicates that there is a strong positive linear correlation (relationship) between the number of hours worked and the take home pay.

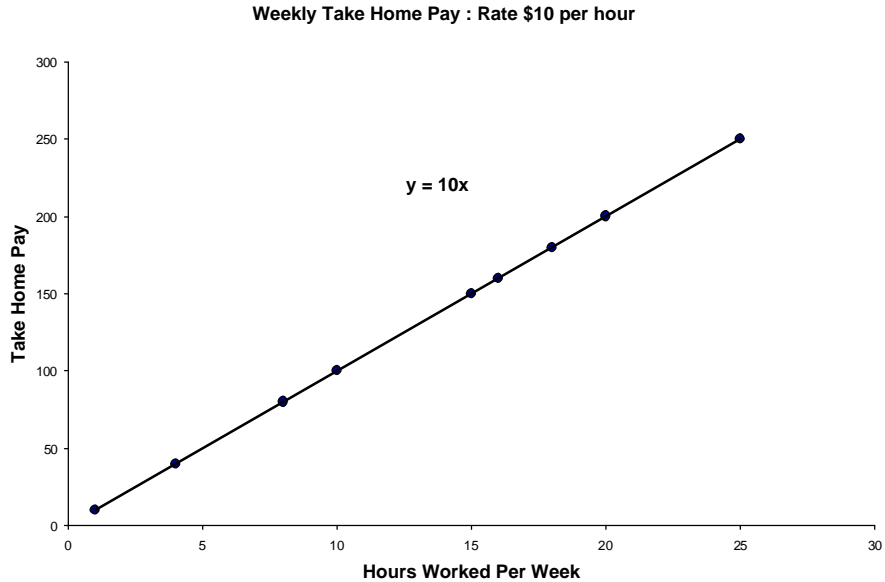


We can join the points on the graph by adding a ***trend line*** (shown on the following page).

This shows that the points are indeed in a straight line.

We can say that ***Take Home Pay*** is positively related to ***Hours Worked per Week***. (We are assuming that the rate of pay stays the same at \$10 per hour).

The scatter points and trend line linking the points can be constructed using a spreadsheet and Chart Wizard in Excel



As well, the equation of the line showing the dependence of the **y-variable** (**Take Home Pay**) on the **x- variable** (**Hours Worked**) appears on the graph.

The equation of the trend line is: $y = 10x$
 The **General Equation of a Straight line** is: $y = mx + b$

m represents the slope of the line and **b** represents the y-intercept or the coordinate of the point where the trend line crosses the y-axis of the graph.

By comparing our equation to the General Equation of a Straight line, we see that the **slope** of our line is **10**, and the **y - intercept** is **(0,0)**.

How would you describe the relationship between **Hours Worked** and **Take Home Pay** ?

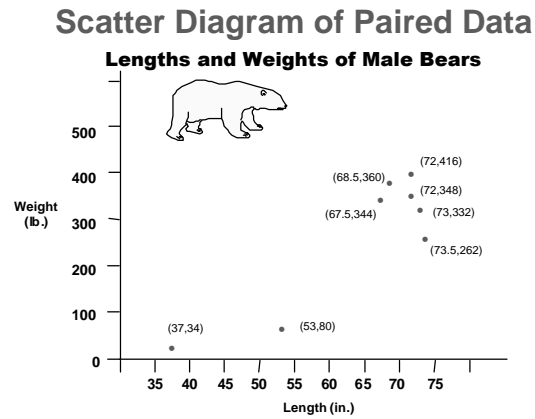
See the next page.

Research Data and Scatter Diagrams

Research data collected from experiments or real life situations often does not show a scatter diagram that is a straight line of dots. Realistically, the scatter plots will show clusters of dots ascending or descending to the right of the graph.

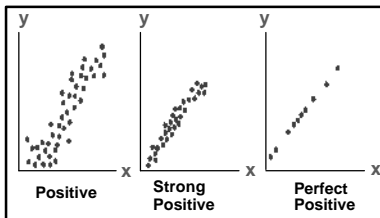
The examples below show different types of scatter plots and what the general interpretation is of each.

(Source: Elementary Statistics, 7th Edition. Mario F. Triola. Addison Wesley)



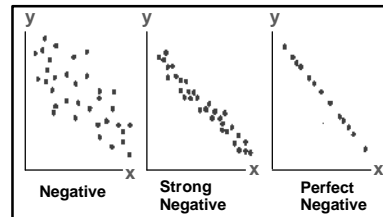
If the data points make a straight line going from the origin out to high x- and y- values, then the variables are said to have a **positive correlation** . If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the variables have a **negative correlation** .

Positive Linear Correlation



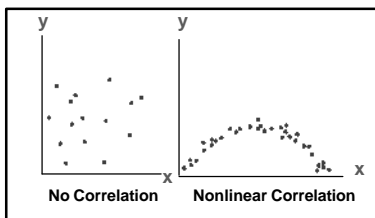
Scatter Diagrams

Negative Linear Correlation



Scatter Diagrams

No Linear Correlation



Scatter Diagrams

Examples:

You might find a **perfect positive correlation** when you compare the total amount of money spent on tickets at the movie theater with the number of people who go.

This means that every time that "**x**" number of people go, "**y**" amount of money is spent on tickets.

You might find a **perfect negative correlation**, when you compare the speed at which a car is going to the amount of time it takes to reach a destination. As the speed increases, the amount of time decreases.

You might find a **strong but not perfect positive correlation** if you compare the number of hours students spent study for an exam with the grade each student received. Students could spend the same number of hours studying but get different marks, so the relationship will not be perfectly positive.

You may get a clustering of scatter plots on the graph, not a perfectly straight line. You will have to add a trend line that is an average between the scatter plot points.

However, generally the rule will hold true that as the amount of time studying increases so does the marks received on the test.

Let's now go to Excel, plot some scatter plots and add some trend lines and equations of lines using an Excel spreadsheet and Chart Wizard.

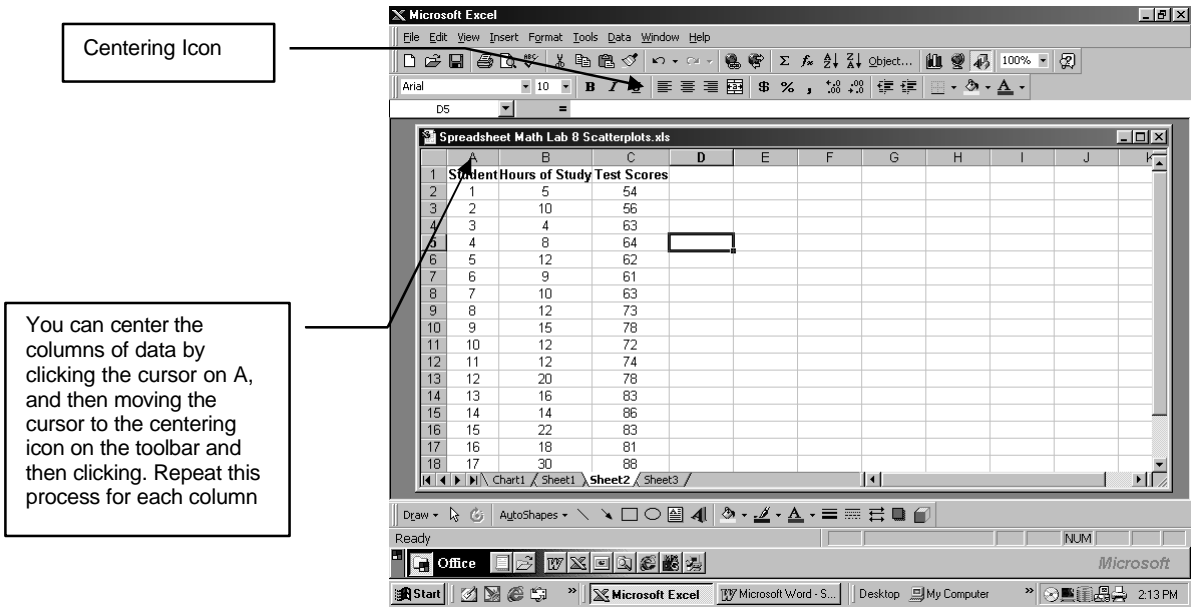
Example: Constructing a Scatter Plot and Line of Best Fit Using an Excel Spreadsheet and Chart Wizard

1. Collect some data from a real life situation. The data we will use for this example shows test scores received by 20 students. Information on the hours each student studied for the test is provided. We will attempt to determine if hours studied has a strong linear correlation to test scores.

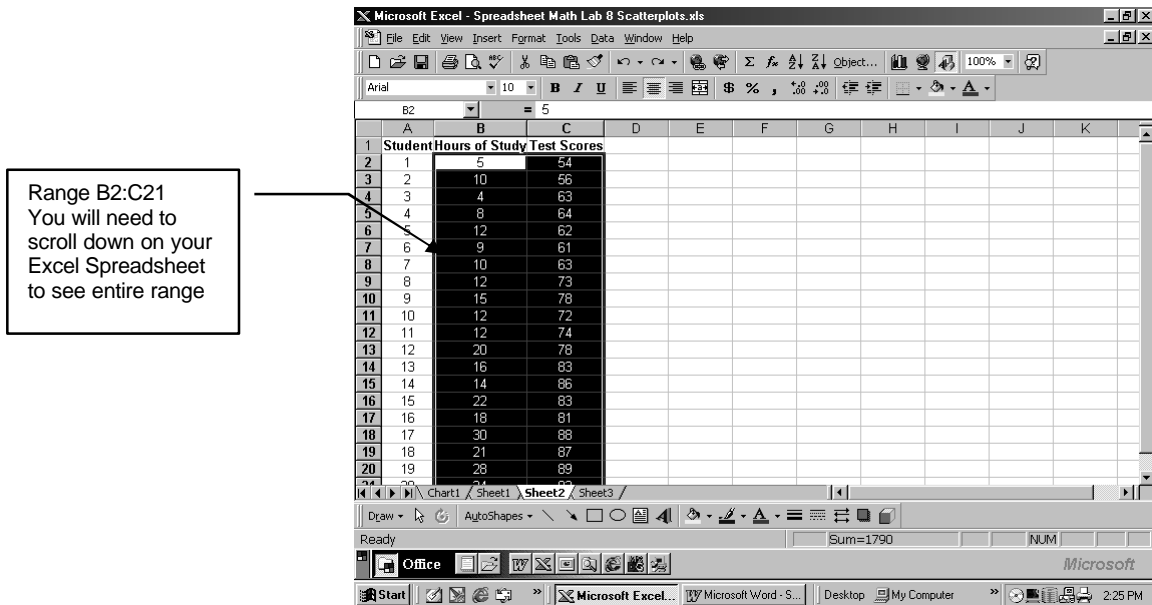
Student Test Scores and Hours of Study

Student	Hours of Study	Test Scores
1	5	54
2	10	56
3	4	63
4	8	64
5	12	62
6	9	61
7	10	63
8	12	73
9	15	78
10	12	72
11	12	74
12	20	78
13	16	83
14	14	86
15	22	83
16	18	81
17	30	88
18	21	87
19	28	89
20	24	93

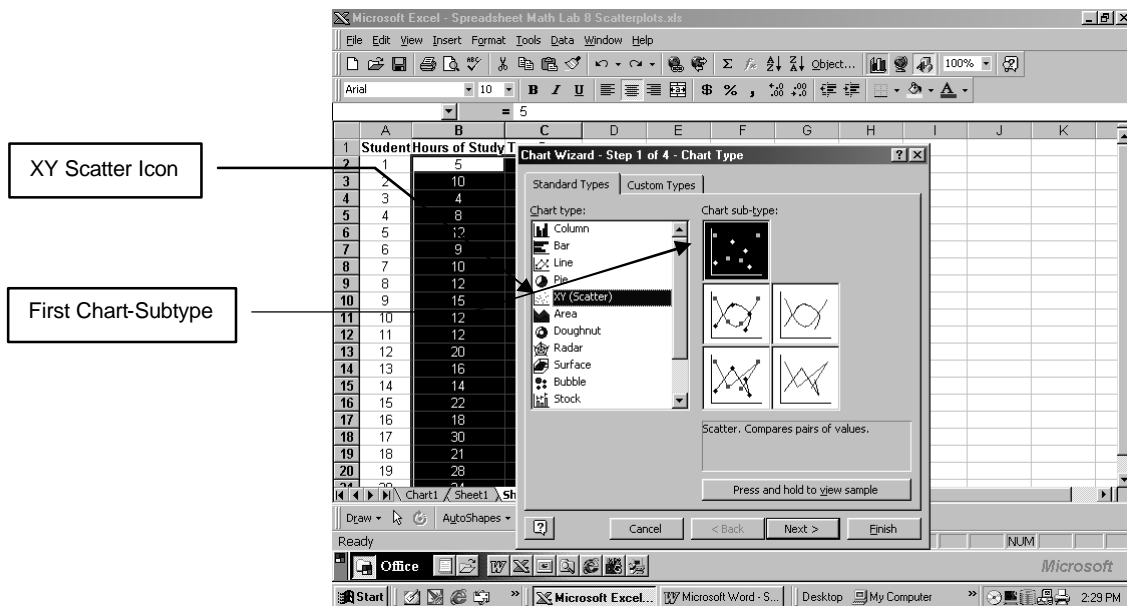
2. Open an Excel Spreadsheet and enter the data from the chart. You can see what the spreadsheet should look like by studying the screen on the following page.



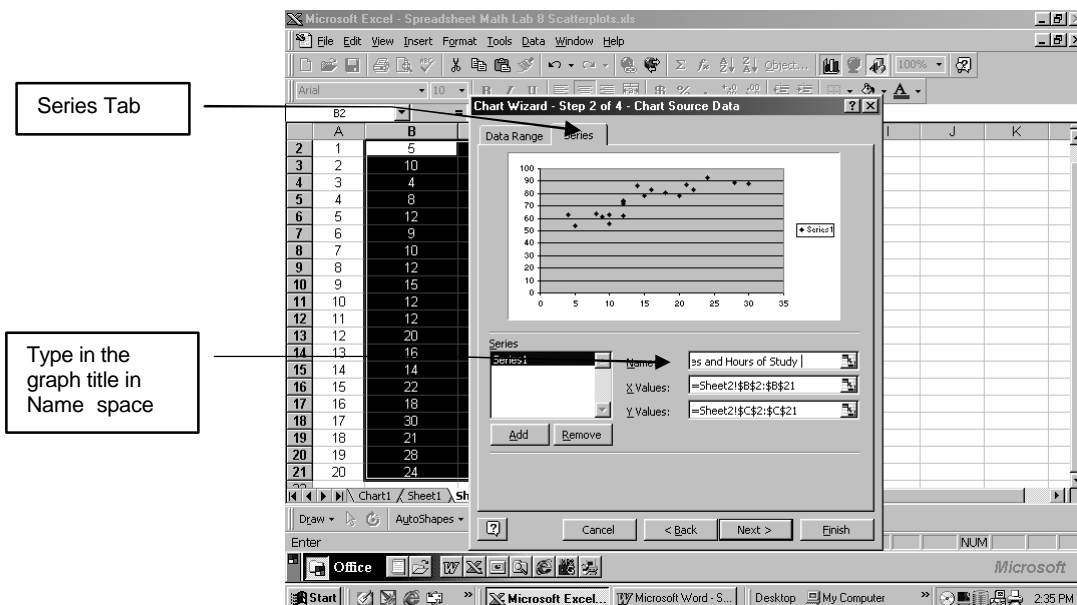
3. Select the range of data you want to plot on the graph. **Do not select** the **Student** column or any of the headings. In this example, the range is **B2:C21**.



4. Select Chart Wizard icon from the Excel Toolbar and **Click on XY Scatter**.



5. Choose the first Chart-subtype and **Click on Next >**. Select the **Series Tab**. Type in the title of your graph in the **Name:** space. We will use the title: **Student Test Scores and Hours of Study**. As you can see, the data points are clustered but are not in a straight line. **Click on Next >**.



6. Select the **Titles Tab** and type in the titles for the **Value x-axis (Hours of Study)** and the **Value y-axis (Test Scores)**.

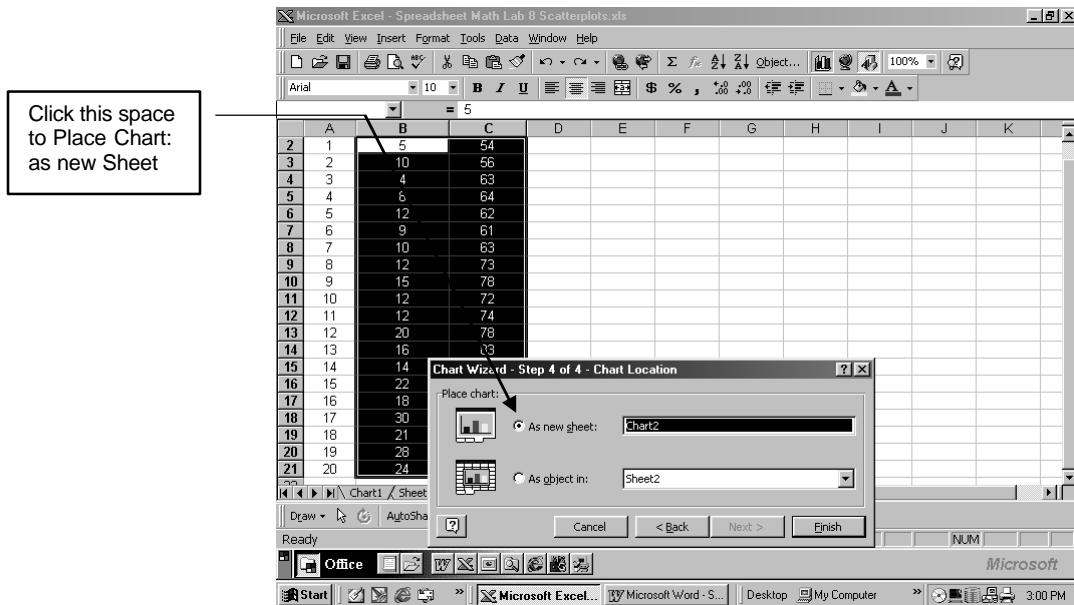
Titles Tab

Hours of Study	Test Scores
5	54
10	56
4	63
8	64
12	62
6	61

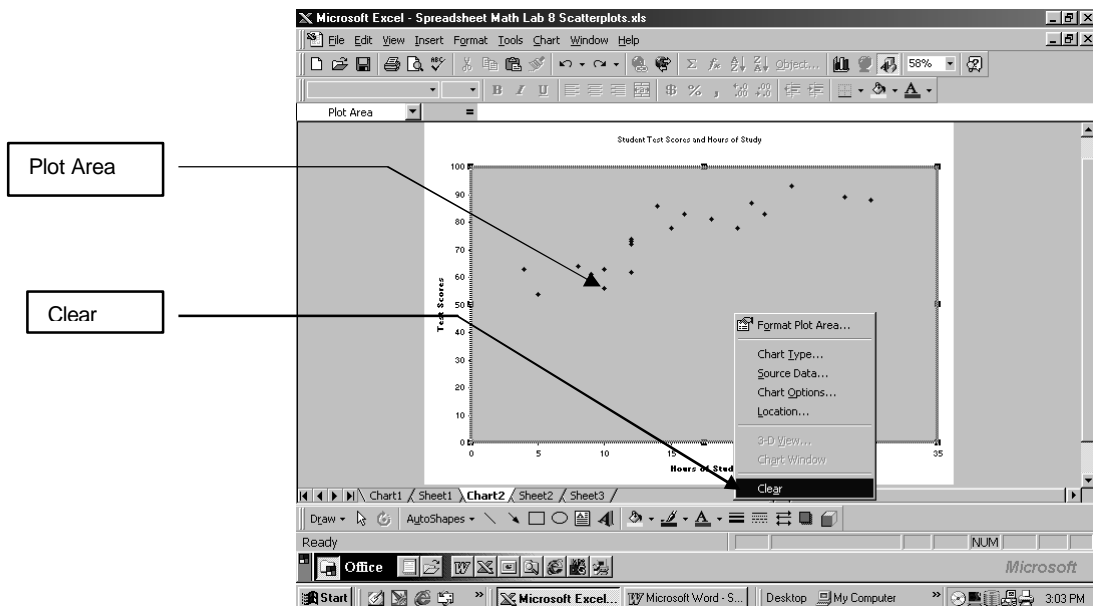
7. Select the **Gridlines Tab** and **Click off** the \dagger in the **Major Gridlines** space.
8. Select the **Legend Tab** and **Click off** the \dagger in the **Show Legend** space.

Hours of Study	Test Scores
5	54
10	56
4	63
8	64
12	62
6	61

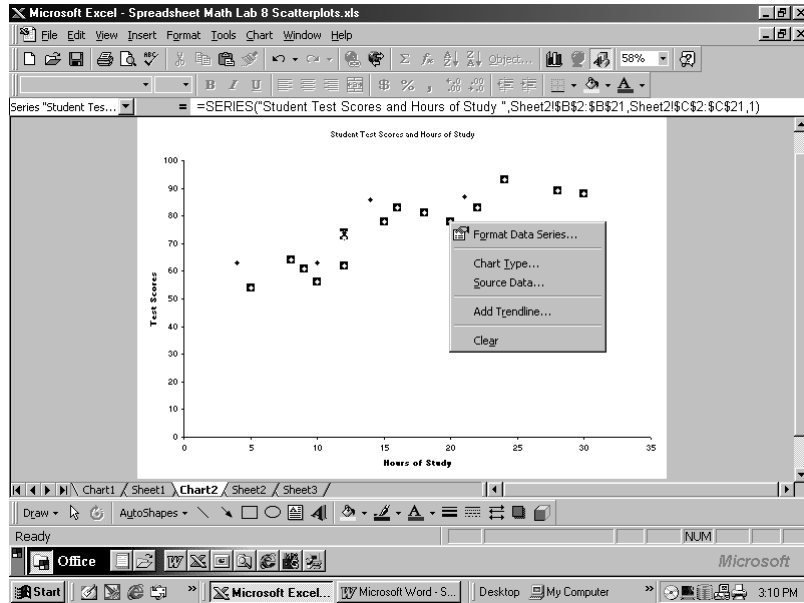
9. Click on **Next >**. Then **Place Chart: as a new Sheet**



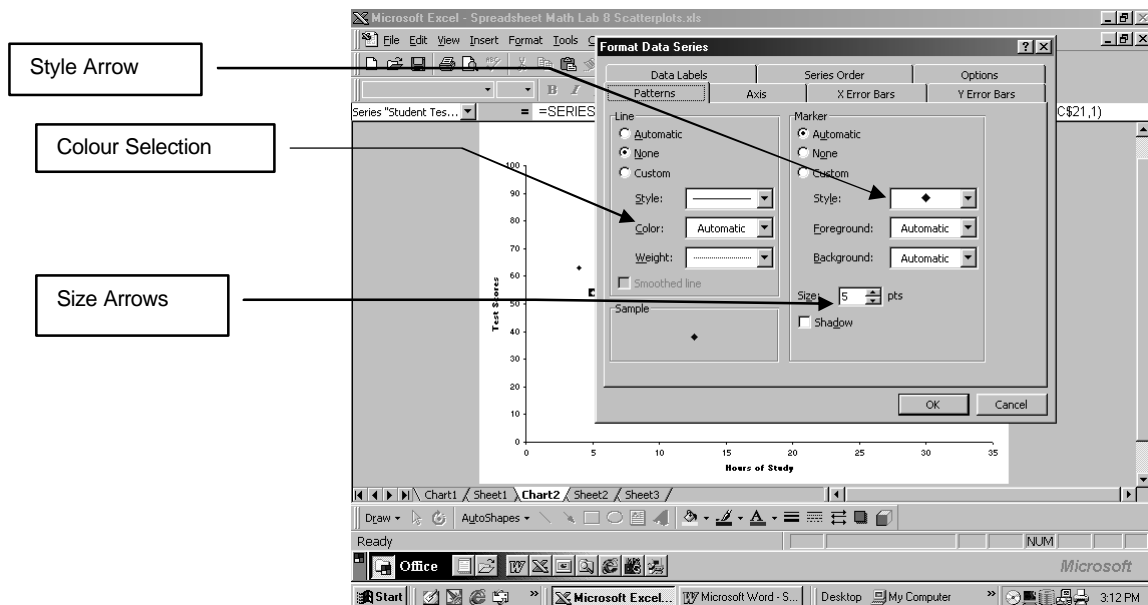
10. Click on **Finish >**. Place the cursor on the grey, **plot area** and **Right Click**. Then **Left Click** on **Clear** and the background will clear to white.



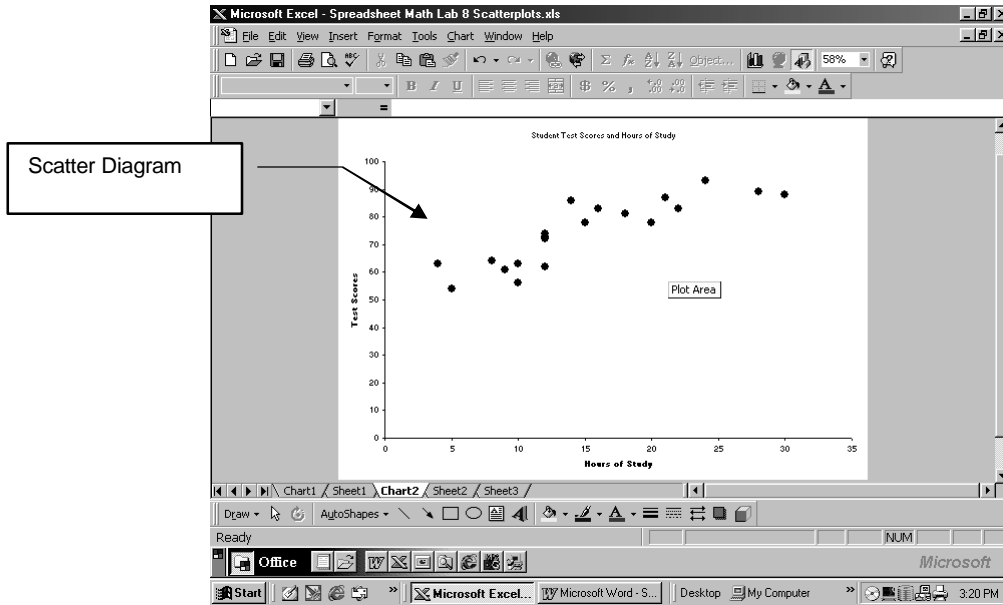
- Place cursor on one of the data plot points and **Right Click**. Next **Left Click** on **Format Data Series**.



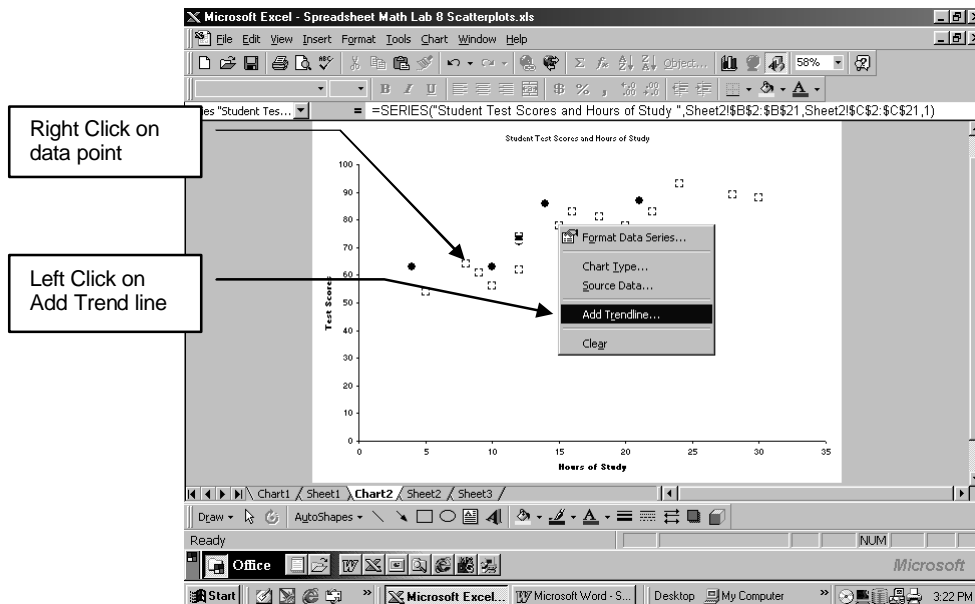
You will see the following screen. At the **Style down arrow**, select the style you like for the data points. You can increase or decrease the size of the data points by using **arrows at Size**: You can also experiment with colour.

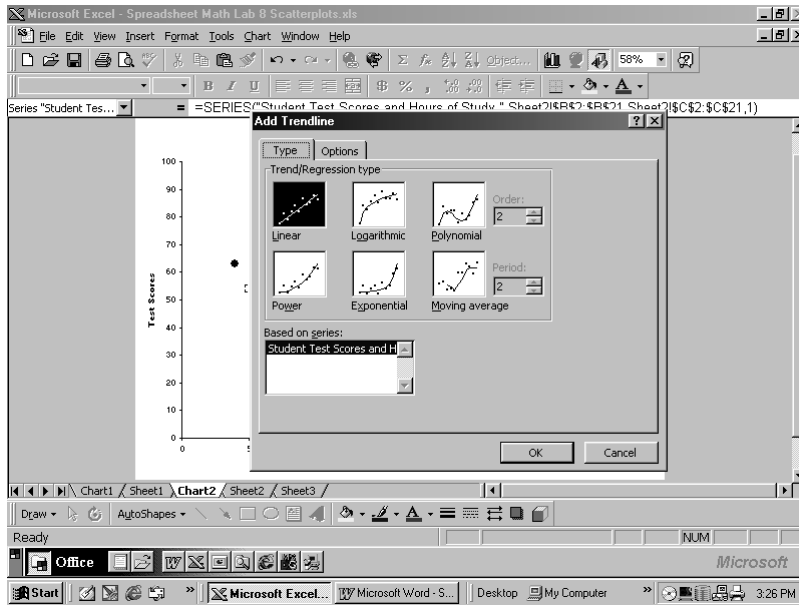


12. Click on **OK**

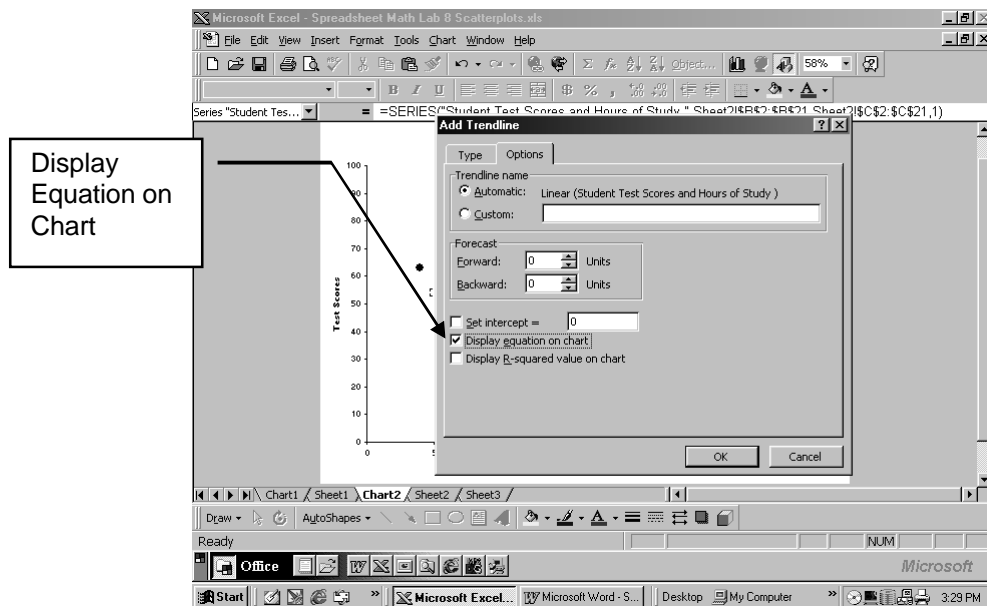


13. To add the trend line, **Right Click** on a data point and then **Click** on **Add Trend line**

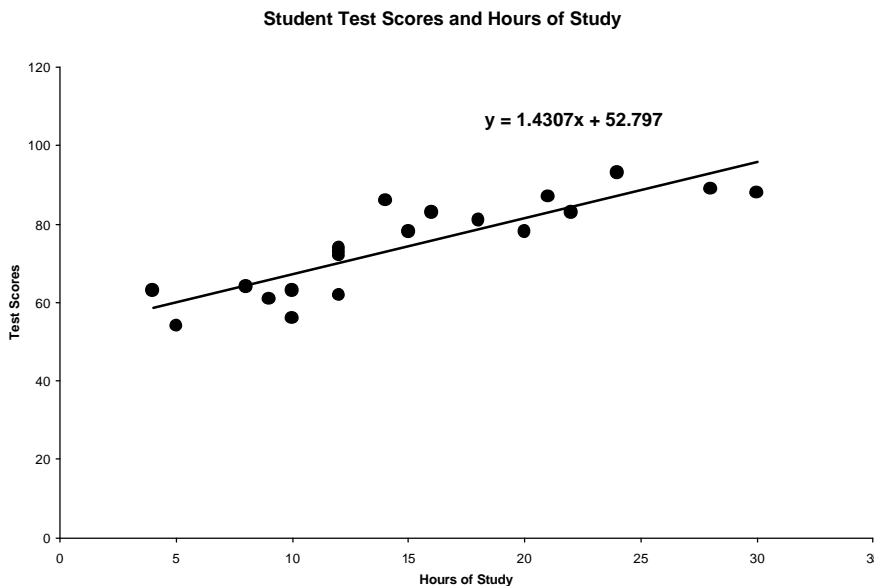
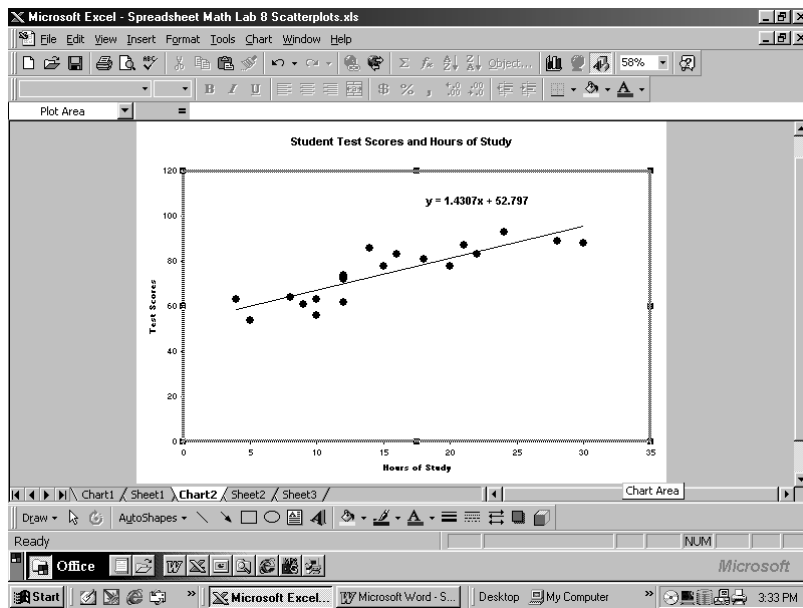




14. Select the first type of Trend line which is labelled **Linear**. Click on the **Options Tab**. Click on the space beside **Display Equation on Chart**. An **f** appears. Click on **OK**



15. If the title does not appear as before, you may need to **retype it in**. To do this simply **Click** on the title that does appear and change it to **Student Test Scores and Study Hours**



This is how your final **Scatter Diagram** and **Trend line** should appear.

Now, answer the questions on the following page.

Questions:

1. How would you describe the correlation between Test Scores and Hours of Study ?
2. What is the equation of the line of best fit ?
3. What is the slope of the line of best fit ?
4. What is the y-intercept of the line of best fit ?
5. What does this y-intercept tell you ?
6. Using the trend line on the scatter plot, estimate the mark of a student who studied for 25 hours.
7. Using the equation of the line of best fit for the trend line, calculate the mark a student who studied 25 hours might receive.

Practice Assignment 1:

Open an Excel spreadsheet and enter the data provided in the following chart.

Now, using the Excel spreadsheet and Chart Wizard and following the example, produce a scatter plot diagram of the data.

The chart should have:

- labelled axis
- a title
- a trend line
- the equation of the trend line.
- A pleasing colour scheme (if you chose colour)

Show your results to your instructor and discuss any problems you may have had in producing your scatter plot chart.

Data: Flying Start Machine 200 m - Race World Record Times

The following data chart shows the world record times for the 200 m race for flying start machines (a kind of bicycle)

Although world records were set in the years in the chart, you must enter all the years in your Excel spreadsheet in order to get a scatter plot.

Year	Record Time (sec)
1974	10.4
1975	10
1976	9.4
1977	9.1
1979	8.8
1985	7.2
1986	6.8
1992	6.5

Questions:

1. How would you describe the correlation between Year and World Record Times?
2. What is the equation of the line of best fit ?
3. What is the slope of the line of best fit ? Why is the slope a negative slope ?
4. Why do you think that a new world record time was not set from 1979 up to 1985 ?
5. Why do you think a new world record time has not been set since 1992 ?

6. Actually a world record time was set in 1988. Estimate the world record time set in 1988 from the scatter plot.
7. Now, using the equation of the line, calculate the actual world record time set in 1988.

Practice Assignment 2:

Open an Excel spreadsheet and enter the data provided in the following chart.

Now, using the Excel spreadsheet and Chart Wizard and following the example, produce a scatter plot diagram of the data.

The chart should have:

- labelled x and y axis
- a title
- trend lines
- the equation of the trend lines.
- A pleasing colour scheme (if you chose colour)

Show your results to your instructor and discuss any problems you may have had in producing your scatter plot chart.

Data: **Flyer Printing and Distribution Costs**

The following data chart shows the prices charged for printing and distributing various quantities of advertising flyers .

Quantity of Flyers	Printing Costs (\$)	Distribution Costs (\$)
1000	56.95	45
2000	78.90	90
3000	100.85	135
4000	122.80	180
5000	144.75	225
10000	254.50	450
15000	364.25	675
20000	474.00	900
25000	583.75	1125

Answer the following questions:

Questions:

1. How would you describe the correlation between Quantity of Flyers and Costs of Printing and Distribution ?

2. What are the equations of the line of best fit ?
3. What are the slopes of the lines of best fit ?
4. Which cost rises fastest ? How do you know ?
5. What are the y-intercepts of the lines of best fit ?
6. What is the basic cost of printing, before adding in any copies of flyers ?
7. Why does the Distribution Cost line have no y-intercept ?
8. Estimate what the Printing Cost and the Distribution Cost will be for 12000 flyers.

Hand In Assignment:

Open an Excel spreadsheet and enter the data provided in the following chart.

Now, using the Excel spreadsheet and Chart Wizard and following the example, produce a scatter plot diagram of the data.

The chart should have:

- labelled x and y axis
- a title
- trend lines
- the equation of the trend lines.
- A pleasing colour scheme (if you chose colour)

Show your results to your instructor and discuss any problems you may have had in producing your scatter plot chart.

Data: Olympic 100 metre Dash Times (1896 - 2000)

The following data chart shows the Olympic World Records set in the 100 metre dash from 1896 to 2000

Year	Time (sec)
1896	12.0
1900	11.0
1904	11.0
1906	11.2
1908	10.8
1912	10.8
1920	10.8
1924	10.6
1928	10.8
1932	10.3
1936	10.3
1948	10.3
1952	10.4
1956	10.5
1960	10.2
1964	10.0
1968	9.95
1972	10.14
1976	10.06
1980	10.25
1984	9.99
1988	9.92
1992	9.96
1996	9.84
2000	9.84

Answer the following questions:

Questions:

1. How would you describe the correlation between Year and Time for the times of the 100 metre dash ?
2. What is the equation of the line of best fit ?
3. What is the slope of the line of best fit ?
4. Make a prediction from the trend line what the 100 metre dash time will be in 2010 ?
5. Using the equation of the line, calculate the predicted time for the 100 metre dash in 2010.