

Line of Best Fit (Regression Analysis)

We have already learned how to create **scatter plots** by identifying, and graphing the independent and dependent variables.

In this section we will learn about a very useful technique for **analyzing trends** in data. These trends can be analyzed using **Regression Analysis** to look at and **model** trends in data in order to **forecast** or **backcast** within the model to make **predictions**. The **reliability** of such predictions will also be investigated.

You will be responsible to know and understand the importance of the following words by the end of this topic.

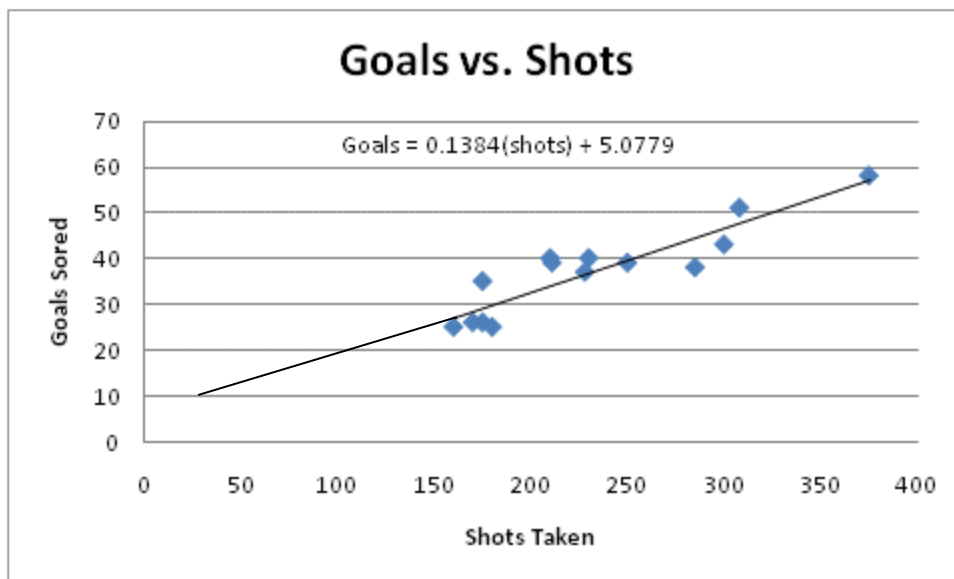
Key Words: **trendline, regression analysis (linear regression), model, predictions, forecasting, backcasting, interpolation, exptpolation, reliability (correlation confidence)**

Example:

Consider the data collected by the NHL for 14 of their players within one year. The number of shots they took and the number of goals they scored.

Shots	170	175	308	285	160	175	230	228	300	375	210	180	211	250
Goals	26	26	51	38	25	35	40	37	43	58	40	25	39	39

1. Create the scatter plot of the data in Excel. Describe the scatter plot in terms of the 4 attributes (type, direction, strength, outliers)



2. Add a **trendline (line of best fit)** – (this is called **Regression Analysis**). Write the equation below.
3. Using the equation that you found **predict/calculate** how many goals would be scored if a player took **280 shots**. Is this an example of interpolation or extrapolation?
4. Using the equation that you found predict how many goals would be scored if a player took **100 shots**. Is this an example of interpolation or extrapolation?
5. If a player scored **42 goals**, estimate the number of shots they are likely to have taken.

As a general rule (reliability rule):

Interpolation calculations tend to be more precise and **more accurate** as they are within the data range.

Extrapolation calculations are **less accurate** in general as they fall outside the range in which actual data was taken.

Try These:

The number of hours spent studying per week and the course grade were tabulated for a set of students. The data is below.

Hours Studied Per Week	12	11	8	5	2	1	10	9	6.5	7.5
Course Grade	95	97	78	72	65	50	85	75	72	75

Create a scatter plot of this data. Perform linear regression on this data. Display the equation on the graph and the coefficient of determination, r^2 . Answer the following:

Equation:

Coefficient of determination, r^2 :

Four attributes:

Description of trend:

Reliability:

Questions:

1. If a student was found to have studied for 3 hours per week what could you calculate their course grade to be?

2. A student was found to have a course grade of 58%. Calculate the approximate number of hours per week they are likely to have studied.