FCIS
AIN SHAMS
UNIVERSITY

Statistical Analysis Research

Non-Linear Regression

By

Name	ID
أميرة ناصر سيد متولي	2021170093
ضحى سيد محمود عبدالغني	20201700422
عبد الرحمن عادل طلعت صالح	20201701236
عبد الله بدر عبد العزيز عبد اللطيف	20201701093
علي أسامة علي إبراهيم الخولي	20201700502
هايدي أشرف مندي محمد	20201700953
هندية رزق ابراهيم محمد	20201700972
ياسمين احمد عبدالفتاح علي	20201700978

Under the supervision of

Prof/ Safia Abbas

2022-2023

Contents

INTRODUCTION

Statistics are everywhere today. You'll find statistics in almost every aspect of life.

Statistical analyses are present in virtually every scientific study. Indeed, these analyses determine whether the result of the study is significant and worthy of being published.

It's powerful stuff, but what is the field of statistics exactly?

The field of statistics is the science of learning from data. Statistical knowledge helps you use the proper methods to collect the data, employ the correct analyses, and effectively present the results.

Statistics is a crucial process behind how we make discoveries in science, make decisions based on data, and make predictions. Statistics allows you to understand a subject much more deeply.

Surprisingly, the field isn't only about numeric results. It also involves a wide range of practices, decisions, and methodologies for both collecting data and analyzing them in a manner that produces valid findings and sound conclusions.

In my view, statistics is an exciting field about the thrill of discovery, learning, and challenging your assumptions. Statistics facilitates the creation of new knowledge. Bit by bit, we push back the frontier of what is known.

There are many branches of statistics, including **Nonlinear Regression** which we will talk about and its strong importance in all fields.

What is Non-linear Regression...?

Also Called ("Non-linear Model"), Nonlinear regression is a technique to analyze a nonlinear relationship between one or more independent variables and a dependent variable. The values of the independent variables are considered to be exact, while the values of the dependent variables are subject to error. The Non-linear Regression Model class implements nonlinear regression in one variable.

General Form of the Non-linear Regression:

$$Y = f(X, \beta) + \varepsilon$$

X = a vector of p predictors, **β** = a vector of k parameters, **ε** = an error term.

f() represents some regression functions .

Error in nonlinear regression

- The standard error is a mathematical tool used in statistics to estimate the error.
- It's better to use the standard error in nonlinear regression to get more accurate value.
- Standard error (SE) = s (standard deviation)/√(n).
- n represents the number of records

In general, nonlinear models are just mathematical equations that describe how YY is produced by XX. In nonlinear regression we'll fit a model formula to pairs of (X, Y) (X, Y) data from an experiment. The best fitting model parameters responsible for giving nonlinear shape to the relationship between XX and YY are then determined by the regression method.

These regression fits produce estimates for the parameters of a nonlinear model. These model parameters are useful because they provide a way to quantify some biological processes (ex, rate and equilibrium constants, minimal and maximal responses, Km and Kd values, Hill slopes, etc.).

History of Non-linear Regression

Nonlinear regression dates to the 1920s to Fisher, Ronald Aylmer and Mackenzie, W.A.

However, the use and more detailed investigation of these models had to wait for advances in automatic calculations in the 1970s.

Pioneers such as Jenn rich (1969) and Malinvaud (1970) have advanced the econometric theory for nonlinear statistical models, while development of computing technology in the last few decades has allowed application of nonlinear models to statistical analysis of complicated relationships between variables.

Nonlinear Regression Model is Discover by: ("Bates and Watts 1988").

Importance of Non-linear Regression

1- Nonlinear regression is a powerful tool for analyzing scientific data, especially if you need to transform data to fit a linear regression.

2- The objective of nonlinear regression is to fit a model to the data you are analyzing the equations that fit the data best are unlikely to correspond to scientifically meaningful models.

3- Before microcomputers were popular, nonlinear regression was not readily available to most scientists. Instead, they transformed their data to make a linear graph, and then analyzed the transformed data with linear regression.

4- present graphical representations for assessing the quality of approximate confidence intervals.

The Difference between Linear and Nonlinear Regression Models

First, we should discuss what the regression models are:

A regression model attempts to determine the strength and character of the relationship between one dependent variable and a series of other variables (known as independent variables).

The two kinds of variables are:

- **Dependent:** Factor that you are attempting to predict or understand.
- **Independent:** Factors that you suspect to have an impact on the dependent variable.

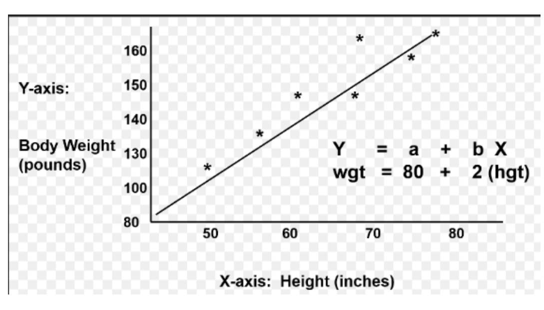
The difference between linear and nonlinear not just the linear produce a straight line and the nonlinear produce curves models as some thought. second, we will discuss the linear regression to differentiate both of them.

Linear Regression

Linear regression is a statistical modeling technique that represent the relationship between the dependent variable and one or more independent variables. It is one of the most common types of predictive analysis. The linear regression model follows one rule, the constant and the parameter it must be multiply by an independent variable "IV". if the equation has these rules, then it is a linear regression model.

Dependent variable = constant +parameter * IV + … + parameter * IV^2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$



Nonlinear Transformation

Some nonlinear regression problems can be moved to a linear domain by a suitable transformation of the model formulation.

Four common transformations to induce linearity are: logarithmic transformation, square root transformation, inverse transformation, and the square transformation

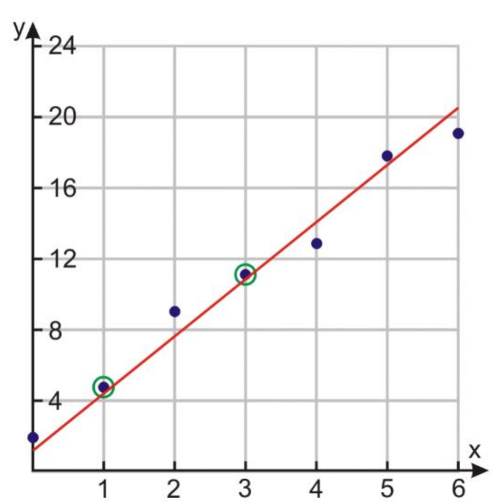
Examples:

$y = e^{\beta x}$ ⟶ $\ln y = \beta x$ if $y \geq 0$

$y = \frac{1}{1+\beta x}$ ⟶ $\frac{1}{y} - 1 = \beta x$ if $y \neq 0$

Ordinary least-squares regression method is a form of regression analysis that establishes the relationship between the dependent and independent variables along a linear line. This line refers to the "line of best fit".

Let us consider two variables, x and y. These are plotted on a graph with values of x on the x-axis and y on the y-axis. The dots represent these values in the below graph. A straight line is drawn through the dots – referred to as the line of best fit.



The objective of least squares regression is to ensure that the line drawn through the set of values provided establishes the closest relationship between the values

The regression line under the least squares method one can calculate using the following formula:

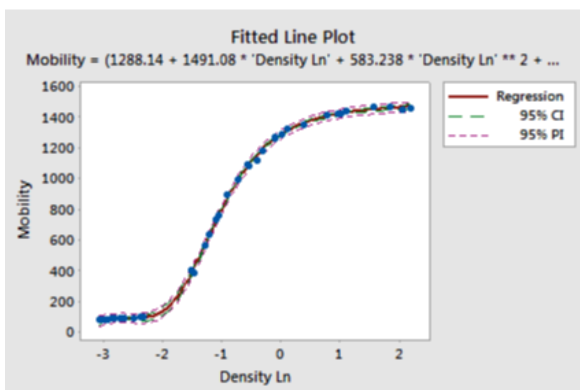
$$\hat{y}=b_0+b_1x \quad , \quad b_1=r\left(\frac{sy}{sx}\right) \quad , \quad b_0=\bar{y}-b_1\bar{x}$$

- **$\hat{y}$** = dependent variable
- **b_0** = y-intercept
- **r** = correlation
- $\bar{x} = mean\ of\ x$
- **S_y** = standard deviation of y

x = independent variable
b_1 = slope of the line
S_y = standard deviation of y
$\bar{y} = mean\ of\ y$

Now let's discuss the Nonlinear Regression model

(The Non-linear Regression definition in page 4)



When we should Choose Between Linear and Nonlinear Regression…?

The general and common method is to use linear regression first to determine whether it can fit the particular type of curve in your data. If you can't obtain the fit using linear regression, then you should need to choose nonlinear regression. Linear regression is easier to use, simpler to interpret, and you obtain more statistics that help you to interpret the model. While linear regression can model curves, it is limited in the shapes of curve.

Sometimes it can't fit the specific curve in your data. while nonlinear regression can fit many types of curves, but it can require more effort to find the best fit and to interpret the role of the independent variables.

The following comparison will clear it: -

The benefits of the **linear regression** model:

- easy to implement.
- easy to interpret.
- easily obtained statistics to assess model.

The benefits of the **Non-linear regression** model:

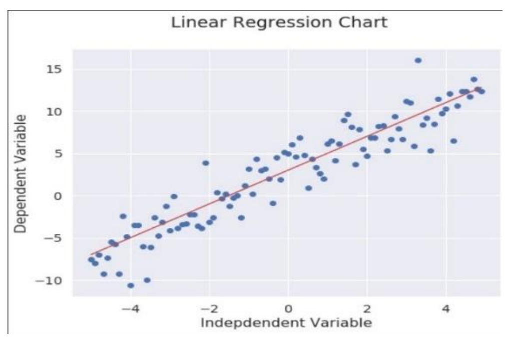
- hard to implemented.
- hard to interpret.
- Can fit more complicated trends in data (but it hard to accomplish).

Data set Shape

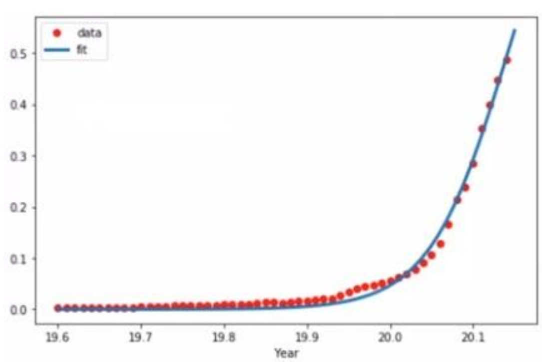
As we know that the data set is consists of features and output and we use different machine learning algorithms to predict the output of these features to make model with high accuracy.

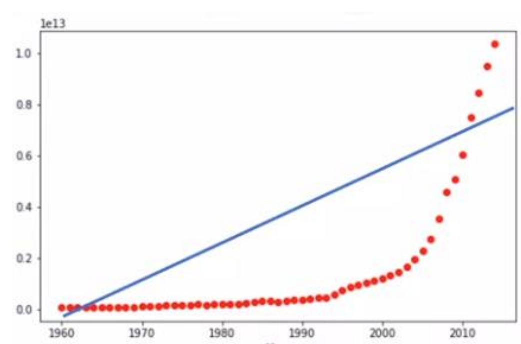
To determine which algorithm is better to we must see the graph of the data first as its shape and curve shows how we will work on it.

- In this figure the data in the graph takes a straight-line shape, so we notice that it’s better to apply linear regression algorithm to get high accuracy and decrease error.



- In this figure the data in the graph is curved, so if we applied linear regression the line won’t fit all the data points and will give low frequency and high error, so it’s better to apply nonlinear regression.





Non-linear Regression Functions.

Polynomial regression: is a form of Linear regression where only due to the Non-linear relationship between dependent and independent variables we add some polynomial terms to linear regression to convert it into Polynomial regression.

Suppose we have **X** as independent data and **Y** as dependent data. Before feeding data to a mode in preprocessing stage we convert the input variables into polynomial terms using some degree.

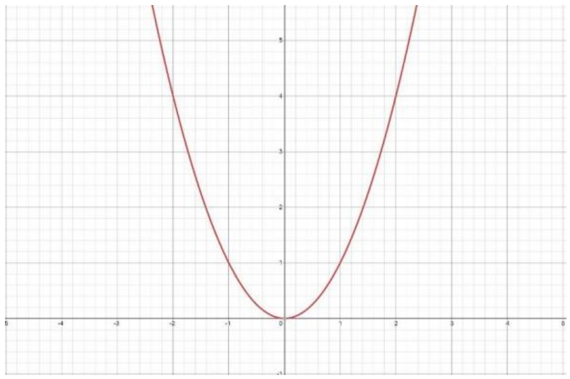
Consider an example my input value is 35 and the degree of a polynomial is 2 so I will find 35 power 0, 35 power 1, and 35 power 2 And this helps to interpret the non-linear relationship in data.

The equation of polynomial becomes something like this.

$$y = a0 + a_1x_1 + a_2x_1^2 + \ldots + a_nx_1^n$$

The degree of order which to use is a Hyperparameter, and we need to choose it wisely. But using a high degree of polynomial tries to overfit the data and for smaller values of degree, the model tries to underfit, so we need to find the optimum value of a degree.

1- Quadratic Regression: is the process of determining the equation of a parabola that best fits a set of data. This set of data is a given set of graph points that make up the shape of a parabola.



The equation of the parabola is

$$y = ax^2 + bx + c$$

where a can never equal zero.

For this process, you must follow the following steps:

Step (1)

Make a table with all your **x** and **y** values. When you plug these values into a graphing calculator, they should form a parabola:

x	*y*
1	32.5
3	37.3
5	36.4
7	32.4
9	28.5

Step (2)

Create **5** additional columns for quadratic regression: **x**, **xy** and **y** values and calculate.

You'll want to use Microsoft Excel or a calculator for this step:

x	y	x^2	x^3	x^4	xy	x^2y
1	32.5	1	1	1	32.5	32.5
3	37.3	9	27	81	111.9	335.7
5	36.4	25	125	625	182	910
7	32.4	49	343	2,401	226.8	1,587.6
9	28.5	81	729	6,561	256.5	2,308.5

Step (3)

At the bottom of each column, calculate the sums:

x	y	x^2	x^3	x^4	xy	x^2y
1	32.5	1	1	1	32.5	32.5
3	37.3	9	27	81	111.9	335.7
5	36.4	25	125	625	182	910
7	32.4	49	343	2,401	226.8	1,587.6
9	28.5	81	729	6,561	256.5	2,308.5
25	**167.1**	**165**	**1,225**	**9,669**	**809.7**	**5,174.3**

Step (4)

Below is the matrix equation for determining the parabolic curve. ∑ represents the summation, meaning that you will plug the relevant sum into the equation. For example, **$\sum xi^4$** would be the sum of column x^4, 9,669. Using the matrix equation, fill in all the sums:

$$a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i$$

$$9669a + 1225b + 165c = 5174.3$$

$$a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i$$

$$1225a + 165b + 25c = 809.7$$

$$a\sum x_i^2 + b\sum x_i + cn_i = \sum y_i$$

$$165a + 25b + 5c = 167.1$$

Step (5)

Solve for **a**, **b**, and **c** by isolating each of these variables using an online calculator. Your result should be the following:

a = -0.3660714
b = 3.015714
c = 30.42179

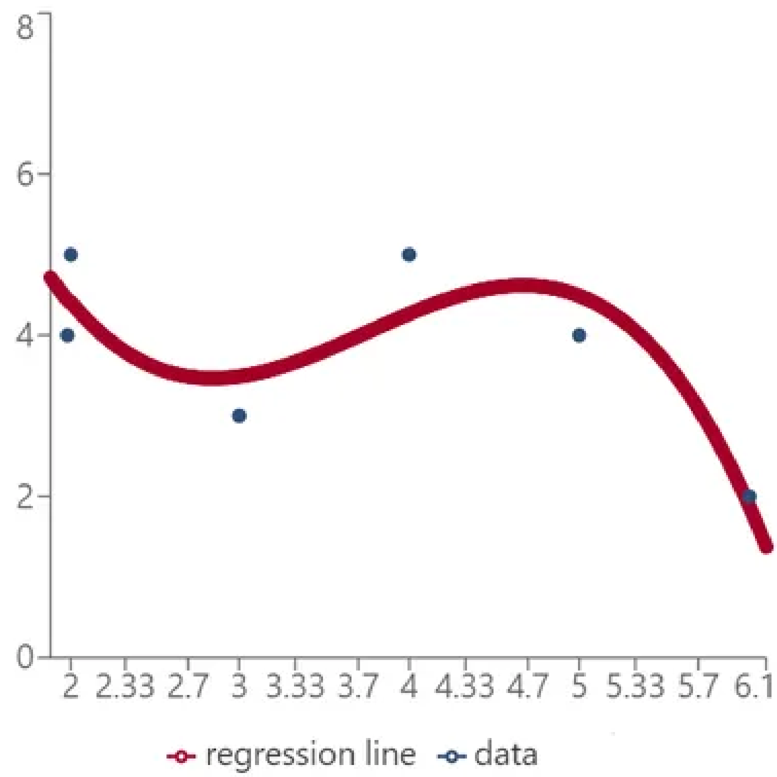
Step (6)

Insert these values (**rounding to the 3rd decimal point**) into our quadratic equation:

$$y = ax^2 + bx + c$$

$$y = -0.366x^2 + 3.016x + 30.422$$

2- **Cubic regression:** is a statistical technique finds the cubic polynomial (**a polynomial of degree 3**) that best fits our dataset. This is a special case of polynomial regression, other examples including



The cubic regression function takes the form:

$$y = a + bx + cx^2 + dx^3$$

where a, b, c, d are real numbers, called coefficients of the cubic regression model. As you can see, we model how the change in x affects the value of y. In other words, we assume here that x is the independent (explanatory) variable and y is the dependent (response) variable.

Let us introduce some necessary notation:

We let X be a matrix with four columns and n rows, where n is the number of data points. We fill the first column with ones, the second with the observed values **$x_1, \ldots, x_n$** of the explanatory (independent) variable, the third with **squares** of these observed values, and the fourth with **cubes** of these observed values:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 & x_N^3 \end{bmatrix}$$

This matrix is often called the **model matrix.**

We let y be a column vector containing the values **$y_1, \ldots, y_n$** of the response (**dependent**) variable:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

We let **β** be the column of the coefficients of the cubic regression model that we're looking for:

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Keep in mind that the order matters start with a at the top and finish with d at the bottom!

Now, to determine the actual values of the coefficients, we just use the so-called normal equation:

$$\beta = (X^TX)^{-1}X^Ty$$

where:

X^T - Transpose of X
$(X^TX)^{-1}$ - Inverse of X^TX
And the operation between every two matrices is matrix multiplication.

Example (1):

Let us find the cubic regression function for the following dataset:

(0, 1) , (2, 0) , (3, 3) , (4, 5) , (5, 4)

Here are our matrices:

The matrix X:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \end{bmatrix}$$

The vector Y:

$$\begin{bmatrix} 1 \\ 0 \\ 3 \\ 5 \\ 4 \end{bmatrix}$$

We apply the formula step-by-step:

First, we determine **X^T**:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 & 5 \\ 0 & 4 & 9 & 16 & 25 \\ 0 & 8 & 27 & 64 & 125 \end{bmatrix}$$

Next, we compute $\mathbf{X^T X}$:

$$\begin{bmatrix} 5 & 14 & 54 & 224 \\ 14 & 54 & 224 & 978 \\ 54 & 224 & 978 & 4424 \\ 224 & 978 & 4424 & 20514 \end{bmatrix}$$

Then, we find $\mathbf{(X^T X)^{-1}}$:

$$\begin{bmatrix} 0.9987 & -0.9544 & 0.2844 & -0.0267 \\ -0.9544 & 5.5128 & -2.7877 & 0.3488 \\ 0.2844 & -2.7877 & 1.4987 & -0.1934 \\ -0.0267 & 0.3488 & -0.1934 & 0.0254 \end{bmatrix}$$

Finally, we perform the matrix multiplication $\mathbf{(X^T X)^{-1} X^T y}$. The linear regression coefficients we wanted to find are:

$$\begin{bmatrix} 0.9973 \\ -5.0755 \\ 3.0687 \\ -0.3868 \end{bmatrix}$$

Therefore, the cubic regression function that best fits our data is:

$$\mathbf{y = 0.9973 - 5.0755x + 3.0687x^2 - 0.3868x^3}.$$

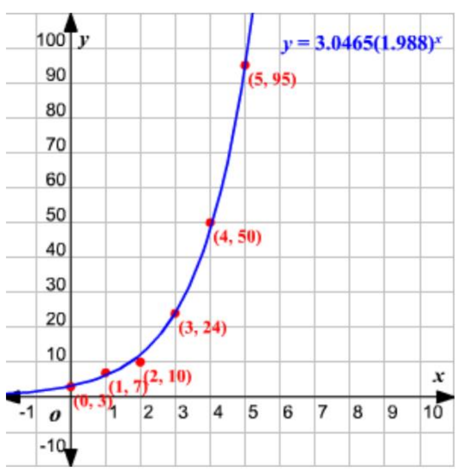
As you can see, to find the cubic linear regression formula by hand, we need to perform a lot of calculations.

Exponential Regression: is used to model situations in which growth begins slowly and then accelerates rapidly without bound, or where decay begins rapidly and then slows down to get closer and closer to zero. This returns an equation of the form: $\hat{y} = a\, b^x$

- b **must** be non-negative.
- When b > 1, we have an exponential growth model.
- When 0 < b < 1, we have an exponential decay model.
- a = y-intercept, a≠0.
- b = slope of the curve.

Let us consider two variables, x and y. x is independent and y is dependent. These are plotted on a graph with values of x on the x-axis and y on the y-axis. The dots represent these values in the below graph.

X	0	1	2	3	4	5
y	3	7	10	24	50	95



The objective of Exponential regression is to ensure that the curve drawn through the set of values provided establishes the closest relationship between the values. This’s better than linear regression because we can cover almost all data, so the error will be fewer.

Logarithmic Regression: is used to model situations in which growth or decay accelerates rapidly at first and then slows over time. This returns an equation of the form:

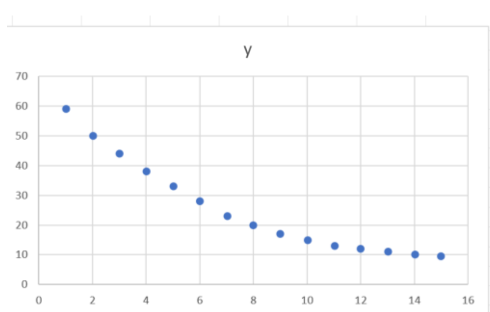
$$\hat{y} = a + b*\ln(x)$$

- a>0, x>0
- a = y-intercept, a≠1
- b = slope of the curve

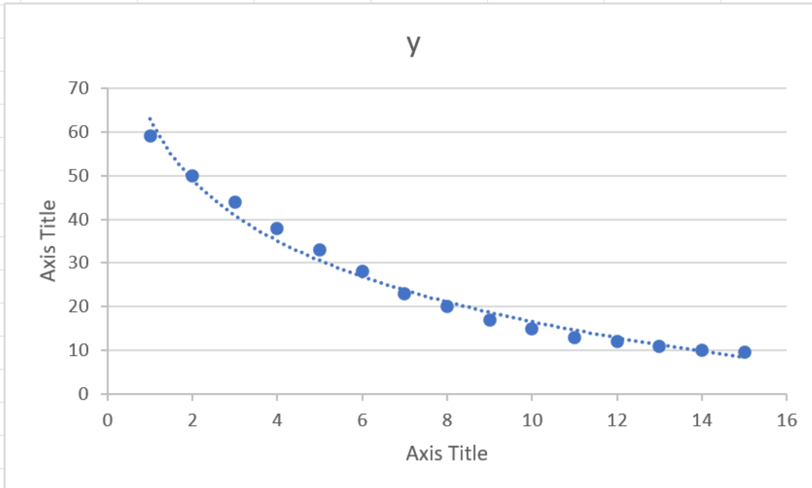
The relation between two variables as the following table which x is independent, and y is dependent.

x	y	ln(x)
1	59	0
2	50	0.693147
3	44	1.098612
4	38	1.386294
5	33	1.609438
6	28	1.791759
7	23	1.94591
8	20	2.079442
9	17	2.197225
10	15	2.302585
11	13	2.397895
12	12	2.484907
13	11	2.564949
14	10	2.639057
15	9.5	2.70805

The scatter plot:



From the scatter plot, we find the graph as ideal Logarithmic function.



And the statistical regression:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.992243
R Square	0.984546
Adjusted R	0.983357
Standard E	2.053604
Observatio	15

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	1	3492.675	3492.675	828.1803	3.7E-13
Residual	13	54.82475	4.217288		
Total	14	3547.5			

	Coefficients	*andard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	63.0686	1.409032	44.76024	1.25E-15	60.02457	66.11263	60.02457	66.11263
X Variable	-20.1987	0.701877	-28.7781	3.7E-13	-21.715	-18.6824	-21.715	-18.6824

$\hat{y}$=63.0686-20.1987*ln(x)

This equation fit the data **98.45%.**

Applications of Non-linear Regression in real life

A nonlinear regression model is used to accommodate different mean functions, even though it is **less flexible** than a linear regression model.

Some of its advantages include **predictability**, **parsimony**, and **interpretability**. Financial forecasting is one way that a nonlinear regression can be applied.

The Financial Prices

A scatterplot of changing financial prices over time shows an association between changes in prices and time. Because the relationship is nonlinear, a nonlinear regression model is the best model to use.

Examples:

Dollar price:

Form 2003 till now:

Stock Prices:

Tesla Stock Prices:

Apple Stock Prices:

Agricultural Research:

Non-linear regression is of great importance in agricultural research. Because many crops and soil processes are better captured by nonlinear than linear models.

Biology:

Since most biological processes are nonlinear in nature, we can find nonlinear model applications in forestry research.

Research and Development:

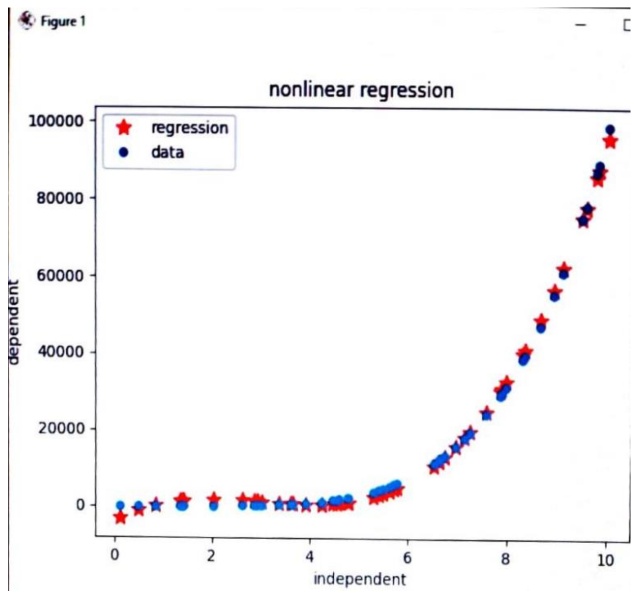
In research and development, it is used in the process of formulation of the problem and deriving statistical solutions to the calibration problem.

Python Example

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.optimize import curve_fit
```

 # data generation

```
n = 50
x= np.random.random(size=(n))*10
m= 5
c= 7
y= x**m-c*x**(m//2) -c*x
```



 # fit the regression model
 # calculate polynomial

```
z = np.polyfit(x, y, 3) #get x,y , degree
f = np.poly1d(z) #represent equation
```

 # calculate new x's and y's

```
y_fit = f(x)
plt.plot(x, y_fit, '*r', label='regression', markersize=10)
plt.plot(x,y, 'o', label='data', markersize=5)
plt.title("nonlinear regression ")
plt.ylabel("dependent")
plt.xlabel("independent")
plt.legend()
plt.show()
```

Summary

If the data shows a curvy trend, then linear regression will not produce very accurate results when compared to a non-linear regression because, as the name implies, linear regression presumes that the data is linear. Non-linear regression categorizes into **Exponential**, **Logarithmic**, **Cubic** and **Quadratic** non-linear regression. Studying non-linear regression is important for our applications in real life.

References

1- encyclopedia.com

2- tjmurphy.gethub

3- libretexts.org

4- Book of Statistics by Jim

5- wallstreetmojo.com

6- omnicalculato.com

7- corporatefinanceinstitute.com

8- otexts.com