BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF MATHEMATICS

ÚSTAV MATEMATIKY

MODIFICATION OF REGRESSION FUNCTION

MODIFIKACE REGRESNÍ FUNKCE

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR | **BSc Seyi Popoola, BSC.**

AUTOR PRÁCE

SUPERVISOR | **doc. RNDr. Libor Žák, Ph.D.**

VEDOUCÍ PRÁCE

BRNO 2021

Assignment Master's Thesis

Institut:	Institute of Mathematics
Student:	**BSc Seyi Popoola, BSC.**
Degree programm:	Applied Sciences in Engineering
Branch:	Mathematical Engineering
Supervisor:	**doc. RNDr. Libor Žák, Ph.D.**
Academic year:	2020/21

As provided for by the Act No. 111/98 Coll. on higher education institutions and the BUT Study and Examination Regulations, the director of the Institute hereby assigns the following topic of Master's Thesis:

Modification of Regression Function

Brief Description:

When investigating dependencies, we may encounter cases where the measured data include sections with different dependencies. However, it is difficult to find a value where the dependency changes. The work will deal with the search for a point of change and relevant dependencies, including further statistical evaluation. The procedures found are applied to real data.

Master's Thesis goals:

Description of regression analysis.
Finding an analytical solution.
Finding a numerical solution.

Recommended bibliography:

DRAPER, N.R., SMITH, H. Applied Regression Analysis (3rd ed.). John Wiley, 1998. ISBN 978-0-4-1-17082-2.

FOX, J. Applied Regression Analysis, Linear Models and Related Methods. Sage, 1997.

SEN, A., SRIVASTAVA, M. Regression Analysis — Theory, Methods, and Applications, Springer-Verlag, Berlin, 2011 (4th printing).

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2020/21

In Brno,

L. S.

prof. RNDr. Josef Šlapal, CSc.
Director of the Institute

doc. Ing. Jaroslav Katolický, Ph.D.
FME dean

Abstrakt

Summary

The regression analysis is a modelling technique that establishes, mathematically, the relationship between entities of a particular subject. Although the modelling is done in such a way that one variable is seen as a subject of the other(s), regression does not imply causation. The modeling has assumptions such as linearity, normality, little or no multicollinearity, homoscedasticity as conditions for optimal relationship establishment. The simplest of the regression technique is the linear regression which also is the most commonly used. It involves the use of a straight line model to define the best pattern of relationship. This best pattern is assessed by the measure of goodness of fit which describes the amount of variation in the response variable explained by the stimuli (or stimulus). Change-point regression is a type of linear regression that takes into account a change in course of the movement of the relationship under study. This type of change in course is taken into account by modelling the regression in segments to account for the entire relationship observable in the data at hand. This model was carried out using the least square method. The data upon which this methodology is applied is the Italy COVID-19 data. The data was subjected to a linear regression and evaluated after which it was subjected to this change point test and the test shows the presence of a change in course. The sections which the test divides the data into two were modelled individually and their regression lines were obtained. The two sections were plotted on a graph with their regression lines intercepting at the crest of the plot.

Klíčová slova

Keywords

Regression Analysis, Least Square and Lagrange Multiplier Estimator, Slyvester Criterion, The Linear Regression Analysis, Regression Line, Non Linear Regression Analzsis, Non Linear Regresssion Line, Change-point Analysis, Method for Detecting Change-Point, Description of Italy Covid - 19 Data, The Change-Point Test.

POPOOLA, S.Modifikace regresní funkce. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2022. 92 s. Vedoucí doc. RNDr. Libor Žák, Ph.D.

I declare that I have written this diploma thesis Modification of Regression Funtion on my own under the direction of my supervisor, doc. RNDr. Libor Zak, PhD., and using the sources listed in the bibliography.

BSc Seyi James Popoola, BSC.

Dr. Libor Zak, PhD has been an ideal teacher, mentor, and thesis supervisor, offering advice and encouragement with a perfect blend of insight and humor. I'm proud of, and grateful for, my time working with professor Zak.

I wish to thank my colleagues who were more than generous with their expertise and precious time: Bature Jonathan, for his countless hours of reflecting, reading, encouraging, inspirations and most of all patience throughout the entire process; and to my other colleagues and friends Korede Olaosun Samuel(My twinie as i loved to call him), Adewale Mudathir, Abu Bunmi, Abdulkabir (sir Yello); I can't thank you guys enough for you supportive efforts and encouragement on this tough Journey, we finally sail through all the difficult trials. I'm indeed grateful.

I would like to acknowledge the Mathmods and Intermaths Scholars.

I would like to acknowledge and thank my school division (the University of L'aquila, Italy, and Brno University of Technology, Czech Republic) for allowing me to conduct my research and providing any assistance requested.

Most importantly, I am grateful for my family's unconditional, unequivocal, and loving support. Most especially my Mum whose constant fasting and prayers, love and support keep me motivated and confident. My accomplishments and success are because she believed so much in me: I give you the respect Mum for not giving up on me. Deepest thanks to my siblings, who keep me grounded, remind me of what is important in life, and they are always in support of my adventures.

Finally, I owe my deepest gratitude to my partner (Boluwatife mi Ifeoluwa Agbeke) my lovely wife by the grace of God, I am forever thankful for constantly listening to me rant and talk things out, for proofreading over and over (even during difficult times), for cracking jokes when things became too serious, and for the sacrifices you have made in every area of our life, thank you for the unconditional love, patience and support throughout the entire thesis process and everyday; also to her siblings and to all Aregbesola's family in general. I love you all.

To God I give everything.

BSc Seyi James Popoola, BSC.

Contents

List of Figures

1. DESCRIPTION OF REGRESSION ANALYSIS

1.1. INTRODUCTION TO REGRESSION ANALYSIS

Regression is a tools mathematician and statistician use to modal data. These tools are used in our daily activities such as finance, predictions about the future, investing, and other disciplines that attempts to determine the strength and character of the relationship between one or more dependent variable (which is mostly denoted by Y) and a series of other variables (known as independent variables denoted as x) [17].
Regression analysis is a statistical system which helps us to dissect and understand the relationship between two or multiple variables of interest. The optimized process for performing regression analysis helps to understand which factors are important, which factors may be overlooked, and how they're affecting each other [23]. Regression analysis is one of the most extensively habituated system among logical models of association employed in business exploration as spoken earlier. Regression analysis attempts to dissect the relationship between a dependent variable and a group of independent variables (one or additional variables).

For better understanding we can describe this analysis has a set of statistical processes for assessing the relationship between the dependent variable (often known as the 'outcome/results variable') and one or multiples Independent variables (often referred to as 'predictors', 'covariates', or 'features/observations'). It's a method used for estimating relationships between a dependent variable and one or more independent variables. it can also be used for assessing the strength of relationships between variables and for future modeling relationship between them [16]. Regression analysis includes many variations, such as linear, Multiple linear and non-linear.

Example: We can use regression model to analysis the age and height of people in a Community, because people's height increases with age and this shows that they have a linear relationship.

In another scenario which was stated by Redman [14]: Assuming you're an incoming supervisor attempting to predict the following monthly purchases. You comprehend that dozens, possibly many variables from the climate to a contender's advancement to the talk of a better than ever model can affect the number. Maybe individuals in your association even have a hypothesis regarding what will have the greatest impact on sales. "Believe me. The more downpour we have, the more we sell." "a month and a half after the contender's advancement, deals bounce" .

Regression analysis is a way to find out mathematically which of those variables/factors actually have an effect [14].
This analysis gives answer to the following questions:

1. Which variables make the biggest difference?

2. Which could be able to disregard?

3. How do those variables collaborate with one another?

Concerning the Redman's scenario which was mentioned above, monthly purchases is our dependent variables and the suspected variable have an impact on it.

1.2. TYPES OF REGRESSION ANALYSIS

We have many types of regression model but to talk of few starting from

- Linear regression model:
  The linear regression method is also a simple regression type, although it includes dependant variable and predictor variable that connect to one another either directly or linearly. It can be determine which is the best fit line with linear regression then set up a predictor error among the predicted value and the main observed. The downside of linear regression is the responsiveness to outlier in the data, therefore it is regularly utilized for minor data or predictions [1].

$$y = \beta_0 + \beta_1 x + \mathcal{E}_i \tag{1.2.1}$$

  The model of linear regression is utilized to portray a connection between factors which are relative to one another. Meaning, the reliant variable builds/diminishes with the autonomous variable [24].

  The linear regression graph has a straight direct line plotted between the factors. Regardless of whether the focuses are not actually in an orderly fashion (which is generally the situation) we can nonetheless see a sample and make sense out of it.

  Example: As a person ages, the level of glucose in his body also increases.

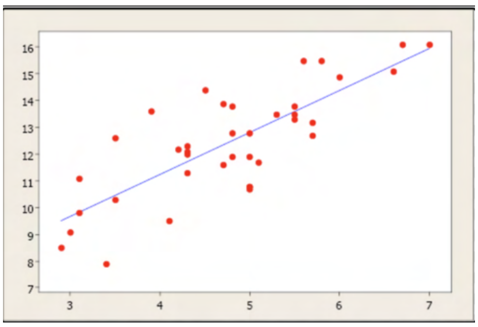


Figure 1.1: Sample graph of a Linear regression [4]

- Multiple regression model:
  The multiple regression technique helps to correspond the connection among a dependant variable and two or more independent variable. When more independent variable is included it makes it a more complex regression analysis study. For instance, the evaluation that if more rain coat sell in the meteorologist forecasts rainy

weather particularly in spring or across all seasons. Also, evaluation of salary incomes for education, experience and proximity to a city area [1].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mathcal{E}_i$$

This model is also utilized when more than one free factor influences a dependent variable. While anticipating result variables, it is essential to gauge how every independent variable moves in its current circumstance and what their progressions will mean for the result or target variable [24].

Example: The chances of a student failing in his test may depend on various input variables like hard work, family issues, health issues etc.

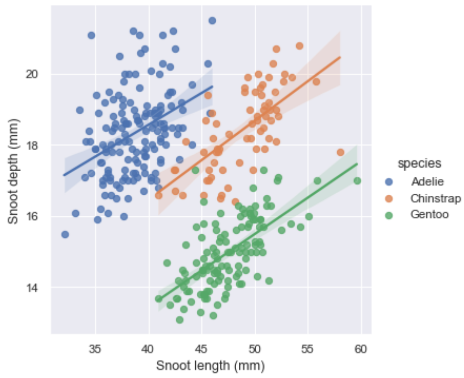


Figure 1.2: Sample graph of a Multiple linear regression [11]

- Non Linear regression model:
  The non linear regression is a mathematical expression that utilize a formed line usually a curve to suit an equation to some data [1].

  For example:

  $$y = e^{\beta_0} e^{\beta_1 x} \tag{1.2.2}$$

  The non linear regression model are utilized owing to the fact that their capacity to fit several mean functions [1]. For the non linear, the diagram doesn't show a linear movement in the model. Contingent upon how the reaction variable responds to the input variable, the line do rise or fall showing the tallness or profundity of the impact of the reaction variable.

  Example: A patient's reaction to treatment can be fortunate or unfortunate relying upon their body inclination and resolve.

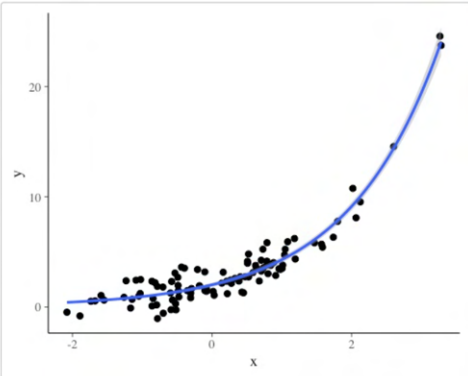


Figure 1.3: Sample graph of a Non linear regression [7]

Some examples of Non-linear regression model

1. Logistic regression model

   This model is most normally utilized when the objective variable or the reliant variable is unmitigated [12]. For instance, regardless of whether a tumor is threatening or harmless, or whether an email is valuable or spam.

   We have 3 types of logistic models

   - Binary logistic models

     This model only have two possible result. For example, a tumor is threatening or harmless [12].

   - Multinomial logistic models

     These kinds of models have at least three potential results without really any request for inclination or positioning. For instance, what kinds of drinks are more preferred(smoothie, milkshake, juice, tea, espresso, and so on) [12].

   - Ordinal logistic models

     These sorts of models have at least three potential results and these results have a request for inclination. For instance, Movie evaluations from 1 to 5 stars [12].

2. Michaelis-Menten Regression model

   Michaelis-Menten Kinetics model serve as the highest prominent Kinetics model. In biochemistry, it is utilized for modeling enzyme kinetics. This model is tagged following a physician from Canada called Maud Menten including a biochemist from Germany called Leonor Michaelis. This model report amount of enzymatic results ratio towards the attention regarding an underlayer. The equation appear as shown [12].

$$v = \frac{v_{max}[S]}{K_{M+}[S]}$$

- Vmax - maximum rate achieved by the system
- KM - Michaelis coefficient
- S - concentration of the substrate
- V - rate of the enzymatic reaction

3. Generalized Additive Models

   These models fit non-parametric bends to given information without requiring a particular numerical model to depict the nonlinear connection between the factors. They are extremely helpful as they permit us to recognize the connections among reliant and autonomous factors without requiring a specific parametric structure [12].

1.3. TERMINOLOGIES USED IN REGRESSION ANALYSIS

- Outliers:
  In a direct words, an outlier is an extreme value. Assuming there is an presumption in the data set that own a very high or very low value as contrast to the other observation in the data, i.e it does not belong to the population, observation like that is called an outlier. An outlier is a problem because most times it hampers the outcome we generate.

- Multicollinearity:
  Multicollinear can be described as when the independent variables are extremely correlated to each other. Numerous types of regression techniques presume that multicollinear should not be available in the data set. The reason is because it makes the job difficult in choosing the most paramount independent variable, or it causes problems in ranking variables base on its importance.

- Heteroscedasticity:
  This is seen as when the variation between the target variable including the independent variable is not constant. For example - The more one's income increase, the higher the variability of food consumption. A poor person will spend constant amount by eating less expensive food always, while a wealthy person may sometimes purchase inexpensive food and some other times, consume expensive meals. Those with more income show a substantial variability of food consumption.

- Heteroscedasticity:
  This is seen as when the variation between the target variable including the independent variable is not constant. For example - The more one's income increase, the higher the variability of food consumption. A poor person will spend constant amount by eating less expensive food always, while a wealthy person may sometimes purchase inexpensive food and some other times, consume expensive meals. Those with more income show a substantial variability of food consumption.

- Undercut and Overfit:
  Overfitting is when our algorithm works well on training set but cannot perform better on the test sets. It can also be described as a problem of high variance. Also, when we use irrelevant explanatory variables, it may lead to overfitting.

  Underfit is when our algorithm works so poorly that is unable to fit even a training set. This is also known as a problem of high bias.

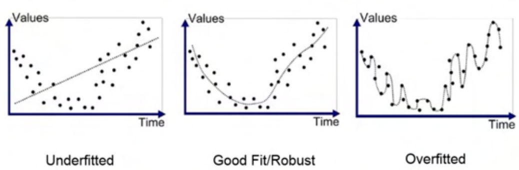


Figure 1.4: Graphs [25]

Knowing the variance between the variables is key factor that is examined as part of regression analysis. We need to understand the measures of variation in other to understand the variance [25].

SST = SSR + SSE

Total Sum of Squares   *Regression Sum of Squares*   *Error Sum of Squares*

Figure 1.5: SST, SSR AND SSE [25]

- SST = Total sum of squares (Overall/Total Variation)
  Calculate the variation of the Y_i values around a mean value of Y [25].
- SSR = Regression sum of squares (Explained Variation)
  Variation is traceable to the relationship between X and Y [25].
- SSE = Error sum of squares (Unexplained Variable)
  Variation in Y traceable to factors other than X [25].

After taking all these factor into consideration, before we can start obtaining if the model is performing well, we need to examine the assumption of Linear Regression.

1.4. DESCRIPTION AND DERIVATION OF REGRESSION FORMULAS

From the general linear model of the form

$$y = X\beta + \mathcal{E}_i \tag{1.4.1}$$

where y is a N x 1 vector of noticed reactions, X is a matrix of fixed constants of N x p dimension, β is a vector of fixed however obscure boundaries of p x 1 dimension, and e is a vector of (unnoticed) errors of N x 1 dimension with no mean. This model is known as a linear model since the mean of the reaction vector y is linear in the obscure boundaries β. Our advantage is to appraise the boundaries of this model and test speculations in regards to direct blends of the boundaries. A few models normally utilized in statistical/mathematical techniques are instances of the general linear model (1.3.1). As additional depicted in this section, these incorporate basic linear regression and multiple regression models, one-way examination of difference (ANOVA), two-way crossed examination with or without collaboration, the analysis of covariance (ANACOVA) model, blended impacts models, and some time series models. We will examine these models and give a few instances of models that are not unique cases.

1.5. THEORETICAL BASIS OF LINEAR REGRESSION

Let $(X_1, \cdots, X_k, Y)^T = (\boldsymbol{X}, Y)^T$ be a random vector whose components have finite second moments. We are looking for the best linear approximation of the quantity Y using $\boldsymbol{X}$. So we are looking for a random variable α and a random vector $\boldsymbol{\beta} : Y = \alpha + \boldsymbol{\beta}^T \boldsymbol{X}$, where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)^T$. The best quality of the approximation is assessed by the standard deviation:

$$E\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right)^2$$

Holds:

$$E\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right)^2 \geq H(Y) - \operatorname{cov}(Y, \boldsymbol{X}) \operatorname{var}(\boldsymbol{X})^{-1} \operatorname{cov}(\boldsymbol{X}, Y)$$

and equality is achieved just when it is

$$\alpha = E(Y) - \boldsymbol{\beta}^T E(\boldsymbol{X}), \boldsymbol{\beta} = \operatorname{var}(\boldsymbol{X})^{-1} \operatorname{cov}(\boldsymbol{X}, Y)$$

Proof:

Let's mark $\boldsymbol{V} = \operatorname{var}(\boldsymbol{X})$. If Z has a finite variance, then we can write:

$$D(Z) = E\left(Z^2\right) - (E(Z))^2$$

, then $E\left(Z^2\right) = D(Z) + (E(Z))^2$.

Accordingly for $Z = Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)$

we get:

$$E\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right)^2 = H\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right) + \left(E\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right)\right)^2 \geq H\left(Y - \left(\alpha + \boldsymbol{\beta}^T \boldsymbol{X}\right)\right).$$

Equality is achieved just when:

$$E\left(Y-\left(\alpha+\boldsymbol{\beta}^T\boldsymbol{X}\right)\right)=E(Y)-\left(\alpha+\boldsymbol{\beta}^T E(\boldsymbol{X})\right)=0, \text{ then } E(Y)=\alpha+\boldsymbol{\beta}^T E(\boldsymbol{X})$$

Holds:

$$D\left(Y-\left(\alpha+\boldsymbol{\beta}^T\boldsymbol{X}\right)\right)=D\left(Y-\boldsymbol{\beta}^T\boldsymbol{X}\right)=D(Y)-C\left(Y,\boldsymbol{\beta}^T\boldsymbol{X}\right)-C\left(\boldsymbol{\beta}^T\boldsymbol{X},Y\right)+H\left(\boldsymbol{\beta}^T\boldsymbol{X}\right)=$$
$$=D(Y)-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{\beta}-\boldsymbol{\beta}^T\operatorname{cov}(\boldsymbol{X},Y)+\boldsymbol{\beta}^T\operatorname{var}(\boldsymbol{X})\boldsymbol{\beta}=$$
$$=D(Y)-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{\beta}-\boldsymbol{\beta}^T\operatorname{cov}(\boldsymbol{X},Y)+\boldsymbol{\beta}^T\boldsymbol{V}\boldsymbol{\beta}+\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)-$$
$$\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)$$
$$=D(Y)-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)+$$
$$\left(-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{\beta}-\boldsymbol{\beta}^T\operatorname{cov}(\boldsymbol{X},Y)+\boldsymbol{\beta}^T\boldsymbol{V}\boldsymbol{\beta}+\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)\right)$$

where,

$$\left(-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{\beta}-\boldsymbol{\beta}^T\operatorname{cov}(\boldsymbol{X},Y)+\boldsymbol{\beta}^T\boldsymbol{V}\boldsymbol{\beta}+\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)=\right.$$
$$\left(\beta-\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y))\right)^T\boldsymbol{V}\left(\beta-\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)\right)$$

Then,

$$D\left(Y-\left(\alpha+\boldsymbol{\beta}^T\boldsymbol{X}\right)=D(Y)-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)+\right.$$
$$\left(\beta-\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y))\right)^T\boldsymbol{V}\left(\beta-\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)\right)$$

and,

$$D\left(Y-\left(\alpha+\boldsymbol{\beta}^T\boldsymbol{X}\right)=D(Y)-\operatorname{cov}(Y,\boldsymbol{X})\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{X},Y)\right.$$

If and only if,

$$\boldsymbol{\beta}-\operatorname{var}(\boldsymbol{X})^{-1}\operatorname{cov}(\boldsymbol{X},Y)=0 \tag{1.5.1}$$

Then,

$$\boldsymbol{\beta}=\operatorname{var}(\boldsymbol{X})^{-1}\operatorname{cov}(\boldsymbol{X},Y) \tag{1.5.2}$$

1.6. REGRESSION LINE

Entered data: $(\boldsymbol{x},\boldsymbol{y})$, where $\boldsymbol{x}\equiv\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$, $\boldsymbol{y}\equiv\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$ Let's mark: $\boldsymbol{X}\equiv(1,\boldsymbol{x})$, where $\mathbf{1}\equiv\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$, therefore $\boldsymbol{X}=\begin{pmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}$

Formulas:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{1.6.1}$$

We use this equation to represent the two-dimensional vector $\widehat{\beta}$ in connection with our normal or estimating equations $\widehat{\beta}_0$ and $\widehat{\beta}_1$. Thus, it, too, is called an estimating equation.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}, \tag{1.6.2}$$

$\hat{\boldsymbol{y}}$ is modeled or predicted regression equation.

$$\boldsymbol{\mathcal{E}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \tag{1.6.3}$$

$\boldsymbol{\mathcal{E}}$ is the error sum of squares. It measures the error/difference between the experiment data/observation and the estimated model.

Taking the expressions of the formulas given above

1. .) (a.) $\boldsymbol{D} = \boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$,

   b.) $\boldsymbol{X}^T\boldsymbol{y} = \begin{pmatrix} \sum_{i=i}^n y_i \\ \sum_{i=i}^n x_i y_i \end{pmatrix}$

Prove of (1a)

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1^\top \\ \mathrm{x}^\top \end{pmatrix} \times \begin{pmatrix} 1 & \mathrm{x} \end{pmatrix} = \begin{pmatrix} 1^\top 1 & 1^\top \mathrm{x} \\ \mathrm{x}^\top 1 & \mathrm{x}^\top\mathrm{x} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

Note:

$$\begin{aligned} 1^\top 1 &= \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= (1 \times 1 + \cdots 1 \times 1) \\ &= n \end{aligned}$$

$$\begin{aligned} 1^\top \mathrm{x} &= \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \times \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= (1 \times x_1 + \cdots 1 \times x_n) \\ &= \sum_{i=1}^n x_i \end{aligned}$$

$$\mathrm{x}^\top \mathrm{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \times \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$
$$= (x_1 \times x_1 + \cdots x_n \times x_n)$$
$$= \sum_{i=1}^{n} x_i^2$$

Prove of (1b)

$$\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 1^\top \\ \\ \mathrm{x}^\top \end{pmatrix} \times \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$
$$= \begin{pmatrix} 1 & \cdots & 1 \\ \\ x_1 & \cdots & x_n \end{pmatrix} \times \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$
$$= \begin{pmatrix} y_1 + \cdots + y_n \\ \\ x_1 y_1 + \cdots + x_n y_n \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{i=1}^n y_i \\ \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

2. .) (a.) $\det(\boldsymbol{D}) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2,$

$(b.)$ $$\boldsymbol{D}^{-1} = \frac{1}{\det(\boldsymbol{D})} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ \\ -\sum_{i=1}^n x_i & n \end{pmatrix}$$

Prove of (2a)

Since $\mathrm{X}^\top \mathrm{X}$ is 2×2, we obtain the determinant by subtracting the product of the elements of the secondary diagonal from the product of the elements of the main diagonal diagonal.

Hence

$$\det\left(\mathrm{X}^\top \mathrm{X}\right) = n \sum_{I=1}^{n} x_i^2 - \left(\sum_{I=1}^{n} x_i\right)^2$$

<u>Prove of (2b)</u>

$$\text{Inverse } = \frac{\text{Adjoint}}{\text{determinant}}$$

But since $\mathrm{X}^\top\mathrm{X}$ is 2×2, the inverse is computed by simply swapping the diagonal entries, putting negatives in front of the secondary diagonal entries (the swapped one), and dividing everything by the determinant of the original matrix. Hence, we have:

$$\left(\mathrm{X}^\top\mathrm{X}\right)^{-1} = \frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}$$

3. .) $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} : \beta_1 = \frac{1}{\det(\boldsymbol{H})}\left(n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i\right), \beta_0 = \bar{y} - \beta_1\bar{x}$

<u>Prove of (3)</u>

$$\begin{aligned}
\widehat{\beta} &= \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\left(\mathrm{X}^\top\mathrm{y}\right) \\
&= \frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \times \begin{pmatrix} \sum_{n=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\
&= \frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)} \begin{pmatrix} \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \\ -\sum_{i=1}^n x_i \sum_{i=1}^n y_i + n\sum_{i=1}^n x_i y_i \end{pmatrix}
\end{aligned}$$

Hence,

$$\beta_1 = \frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)}\left(n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i\right)$$

Note, from b_2, we have;

$$\frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)} n\sum_{i=1}^n x_i y_i = \beta_1 + \frac{1}{\det\left(\mathrm{X}^\top\mathrm{X}\right)} \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

Also note that:

$$\sum_{i=1}^n x_i = n\bar{x}$$

Now,

$$\begin{aligned}
\beta_0 &= \frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \\
&= n\bar{y}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i^2 - \bar{x}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} n \sum_{i=1}^{n} x_i y_i \\
&= n\bar{y}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i^2 - \bar{x}\left(\beta_1 + \frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \\
&= n\bar{y}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i^2 - \bar{x}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - \beta_1 \bar{x} \\
&= \frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \sum_{i=1}^{n} y_i \left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2\right) - \beta_1 \bar{x} \\
&= \frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \frac{1}{n} \sum_{i=1}^{n} y_i \left(n\sum_{i=1}^{n} x_i^2 - \left[\sum_{i=1}^{n} x_i\right]^2\right) - \beta_1 \bar{x} \\
&= \bar{y}\frac{1}{\det\left(\mathrm{X}^\top \mathrm{X}\right)} \det\left(\mathrm{X}^\top \mathrm{X}\right) - \beta_1 \bar{x}
\end{aligned}$$

Then

$$= \bar{y} - \beta_1 \bar{x} \tag{1.6.4}$$

1.6.1. TOTAL SUM OF SQUARES

In statistics, the total sum of squares (S_T) describes the variation between the values of a dependent variable and the sample mean.

$$S_T = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \tag{1.6.5}$$

y_i – the sample value, $\bar{y}$ - the sample mean

<u>Prove</u>

From

$$S_T = (\boldsymbol{y} - 1\overline{\boldsymbol{y}})^T(\boldsymbol{y} - 1\overline{\boldsymbol{y}}) \tag{1.6.6}$$

$$
\begin{aligned}
S_T &= (\mathrm{y} - 1\bar{y})^\top (\mathrm{y} - 1\bar{y}) \\
&= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top 1\bar{y} - \bar{y}^\top 1^\top \mathrm{y} + \bar{y}^\top 1^\top 1\bar{y} \\
&= \sum_{i=1}^{n} y_i^2 - \bar{y}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2 \\
&= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - n\bar{y}^2 + n\bar{y}^2
\end{aligned}
$$

Then,

$$
= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \tag{1.6.7}
$$

Note:

$$
\begin{aligned}
\mathrm{y}^\top \mathrm{y} &= \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \times \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \left(y_1^2 + \cdots + y_n^2\right) \\
&= \sum_{i=1}^{n} y_i^2
\end{aligned}
$$

Also,

$$
\bar{y}^\top = \bar{y} \tag{1.6.8}
$$

1.6.2. RESIDUAL SUM OF SQUARES

: Residual sum of squares (S_E) measures the variability of model errors. Another way to explain it is that it shows how a regression model cannot explain the variation in the dependent variable. Regression models with lower residual sums of squares generally explain the data better, while regression models with higher residual sums of squares generally do not explain the data well.

$$
S_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1.6.9}
$$

$$
S_E = \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i \tag{1.6.10}
$$

y_i is the observation, $\widehat{y}$ is the regression line estimated value

<u>Prove</u>

$$
S_E = (\boldsymbol{y} - \widehat{\boldsymbol{y}})^T (\boldsymbol{y} - \hat{\boldsymbol{y}}) \tag{1.6.11}
$$

$$
\begin{aligned}
S_E &= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \hat{\mathrm{y}} - \hat{\mathrm{y}}^\top \mathrm{y} + \hat{\mathrm{y}}^\top \hat{\mathrm{y}} \\
&= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \left(\mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y}\right) - \mathrm{y}^\top \mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} + \\
&\mathrm{y}^\top \mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} \\
&= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \mathrm{X}\beta - \mathrm{y}^\top \mathrm{X}\beta + \mathrm{y}^\top \mathrm{X}\mathrm{I}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} \\
&= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \mathrm{X}\beta - \mathrm{y}^\top \mathrm{X}\beta + \mathrm{y}^\top \mathrm{X}\beta \\
&= \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \mathrm{X}\beta
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{y}^\top \mathrm{X} &= \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \times \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \\
&= \begin{pmatrix} y_1 + \cdots + y_n & x_1 y_1 + \cdots + x_n y_n \end{pmatrix} \\
&= \left(\sum_{i=1}^{n} y_i \quad \sum_{i=1}^{n} x_i y_i\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{y}^\top \mathrm{X}\beta &= \left(\sum_{i=1}^{n} y_i \quad \sum_{i=1}^{n} x_i y_i\right) \times \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \\
&= \beta_0 \sum_{i=1}^{n} y_i + \beta_1 \sum_{i=1}^{n} x_i y_i
\end{aligned}
$$

Hence,

$$
S_E = \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \mathrm{X}\beta
$$

$$
= \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i \tag{1.6.12}
$$

1.6.3. REGRESSION SUM OF SQUARES

Regression sum of squares (S_R) assesses the degree to which the modeled data is accurately represented by the regression model. You can calculate regression sum of squares by using the following formula:

$$
S_R = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 \tag{1.6.13}
$$

$$
S_R = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2 \tag{1.6.14}
$$

$\widehat{y}_i$ is the regression line estimated value, $\bar{y}$ is the sample mean

Prove

$$S_R = (\hat{\boldsymbol{y}} - 1\overline{\boldsymbol{y}})^T(\hat{\boldsymbol{y}} - \mathbf{1}\overline{\boldsymbol{y}}) \tag{1.6.15}$$

$$\begin{aligned}
S_R &= (\hat{\mathrm{y}} - 1\bar{y})^\top(\hat{\mathrm{y}} - 1\bar{y}) \\
&= \hat{\mathrm{y}}^\top\hat{\mathrm{y}} - \hat{\mathrm{y}}^\top 1\bar{y} - \bar{y}^\top 1^\top \hat{y} + \bar{y}^\top 1^\top 1\bar{y} \\
&= \sum_{i=1}^{n} \hat{y}_i^2 - \bar{y}\sum_{i=1}^{n} \hat{y}_i - \bar{y}\sum_{i=1}^{n} \hat{y}_i + n\bar{y}^2 \\
&= \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y}\sum_{i=1}^{n} (y_i - \mathcal{E}_i) + n\bar{y}^2 \\
&= \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y}\left[\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \mathcal{E}_i\right] + n\bar{y} \\
&= \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y}[n\bar{y} - 0] + n\bar{y}^2 \\
&= \sum_{i=1}^{n} \hat{y}_i^2 - 2n\bar{y}^2 + n\bar{y}^2
\end{aligned}$$

Then,

$$= \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y} \tag{1.6.16}$$

1.6.4. RELATIONSHIP BETWEEN (TOTAL, ERROR, AND REGRESSION) SUM OF SQUARES

The following equation summarizes the relationship between the three types of sum of squares (i.e. the total sum of square (S_T), regression sum of square S_R and the residual sum of square S_E)

$$S_T = S_R + S_E \tag{1.6.17}$$

$$\boldsymbol{D} \cdot \boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{y}$$

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=i}^{n} y_i \\ \sum_{i=i}^{n} x_i y_i \end{pmatrix}$$

$$\beta_0 n + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=i}^{n} y_i \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=i}^{n} x_i y_i \tag{1.6.18}$$

Prove

We have already generated the following results

$\hat{y}_i = \beta_0 + \beta_1 x_i$

$S_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$

$S_T = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ - (total variation)

$S_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

$S_R = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2$ - (regression variation)

$S_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$S_E = \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i$ - (variation by linear model)

So,

$$
\begin{aligned}
&S_A + S_E = \sum_{i=1}^{n} \left(\beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2\right) - n\bar{y}^2 + \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = \\
&\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 + \sum_{i=1}^{n} \beta_0^2 + \sum_{i=1}^{n} 2\beta_0\beta_1 x_i + \sum_{i=1}^{n} \beta_1^2 x_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = \\
&S_T + \sum_{i=1}^{n} \beta_0^2 + \sum_{i=1}^{n} 2\beta_0\beta_1 x_i + \sum_{i=1}^{n} \beta_1^2 x_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i \\
&\sum_{i=1}^{n} \beta_0^2 + \sum_{i=1}^{n} 2\beta_0\beta_1 x_i + \sum_{i=1}^{n} \beta_1^2 x_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = \\
&= n\beta_0^2 + 2\beta_0\beta_1 \sum_{i=1}^{n} x_i + \beta_1^2 \sum_{i=1}^{n} x_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = \\
&= \beta_0 \left(n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i\right) + \beta_1 \left(\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2\right) - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = \\
&= \beta_0 \sum_{i=1}^{n} y_i + \beta_1 \sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i = 0 \qquad (1.6.19)
\end{aligned}
$$

1.7. SIMPLE LINEAR REGRESSION

From the sample problem, The least complex form of the straight model emerges with one of the fundamental issues in rudimentary insights, where y_i are haphazardly tested from a populace with obscure mean μ and fluctuation σ^2. For this situation, Xβ takes an extremely basic structure

$$X\beta = 1(\mu)$$

in order that the scalar μ is simply the unknown coefficient vector β

Appraise the model whereby the reaction variable y_i is corresponding to an independent variable x_i, stated by

$$y_i = \beta_0 + \beta_1 x_i + \mathcal{E}_i, \quad i = 1, \ldots, n \tag{1.7.1}$$

Whereby $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ are generally presumed to exist uncorrelation random variables accompanied by mean zero together with constant variance σ^2. Let's presume that $x_1, x_2, \ldots, x_n$ are set of constant variables, detected without inaccuracy, afterwards equation 1.2 is a unique case of equation 1.1 with

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad \mathrm{X}\beta = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_{n-1} \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \cdots \\ \mathcal{E}_n \end{bmatrix}$$

In such a way that X is $n \times 1$, X is $n \times 2, \beta$ is 2×1, and $\mathcal{E}$ is $n \times 1$. Observe that x_i were calculated with mistake, afterwards the model in equation (1.3.1) is not a unique case related to Model in equation (1.5.1), on account of this, the matrix X is random, not specified [39].

1.8. MULTIPLE LINEAR REGRESSION

Let's take a look at this model whereby the result variable y_i is linearly connected to many independent variables $x_{i1}, x_{i2}, \ldots, x_{ik}$, indicated

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \mathcal{E}_i, \quad i = 1, \ldots, n \tag{1.8.1}$$

Whereby once more again $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ are regularly presumed as uncorrelation random variables along mean zero including variance constant σ^2.

Let's presume that $x_{i1}, x_{i2}, \ldots, x_{ik}$ are stable constants noticed without mistake/error, afterwards the regression model in equation (1.6.1) is not a unique case of the common linear model in equation (1.3.1) [39]

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad \mathrm{X}\beta = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & \cdots & x_{3k} \\ \cdots & & \cdots & \cdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \cdots \\ \mathcal{E}_n \end{bmatrix}$$

- y is an $n \times 1$ vector of observations on the dependent variable.
- X is an $n \times k$ matrix where we have observations on k independent variables for n observation.
- β is a $k \times 1$ vector of unknown population parameters that we want to estimate.

- $\mathcal{E}$ is an n×1 vector of disturbances or errors

y predicts X while β_0 is the intercept terms and β_1 is the slope terms.

There are some components errors $\mathcal{E}$ we fail to observe or notice and this error result to the failure of data not falling on the straight line including representing the difference among the true and presumed realization of y. There are various reasons that cause this difference, for instance,, variables may be subjective, the outcome of all the deleted variables, and inherent randomness in the observation etc. We presume that error $\mathcal{E}$ is detected as precisely distributed and independent random variable along constant variance and mean zero variance constants σ^2. Afterwards, we will also presume that error $\mathcal{E}$ is distributed normally [41].

It is seen that the independent variable are being controlled by the experiment, therefore it is examined as non-theoretical while y is seen as a random variable with

$$E(y) = \beta_0 + \beta_1 X \tag{1.8.2}$$

E($\boldsymbol{\beta}$) = $\widehat{\boldsymbol{\beta}}$. Implies that $\boldsymbol{\beta}$ is an unbiased estimate of $\boldsymbol{\beta}$.

and

$$\mathrm{Var}(y) = \sigma^2 \tag{1.8.3}$$

Var($\boldsymbol{\beta}$) = $\sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Estimated coefficients are described by the variances and co-variances.

Variance of X can sometimes be a random variable. In situations like this, we consider the conditional mean and variance of y given $X = x$ as

$$E(y \mid x) = \beta_0 + \beta_1 x \tag{1.8.4}$$

and the conditional variance of y given $X = x$ as

$$\mathrm{Var}(y \mid x) = \sigma^2 \tag{1.8.5}$$

The model is fully set out, when the values of intercept β_0, slope β_1 and variance σ^2 are studied/known. The parameters (β_0, β_1 and σ^2) are broadly not known in operation and error $\mathcal{E}$ is not noticed. The calculation of the statistical model of $y = \beta_0 + \beta_1 x + \mathcal{E}$ is base on the computation. for instance, estimation of intercept β_0, slope β_1 and variance σ^2. To know the rate of these parameters, n pairs of observation (X_i, y_i) where $i = 1, \ldots, n$ on (X, y) are analyse and they are utilized to decide the unknown parameters [41].

We decide the estimate of the parameters by utilizing different methods, but the two popular methods are:

- the method for least squares and
- the maximum likelihood

1.9. REGRESSION SYMBOLS

Before we go into more details of our analysis, we are going to address these symbols ($\beta, \beta_0, \beta_1, \widehat{\beta_0}, \widehat{\beta_1}$ and $SE\beta$) that confuse students in regression analysis.

- β, β_0, β_1 are the unstandardized beta
- $SE\beta$ the standard error for the unstandardized beta
- $\widehat{\beta}, \widehat{\beta_0}, \widehat{\beta_1}$ are the standardized beta

1.9.1. REGRESSION TABLE

Let's take a look at this regression table as an example

Source	β	SEB	$\widehat{\beta}$
Variable 1	1.35	0.34	.34
Variable 2	1.10	3.41	.05
Variable 3	−1.83	0.11	−.16

The unstandardized beta (β) is the first symbol in our sample table and what it represent is the slope of the regression line between the dependent and the independent variables. Starting from the first variable which is variable 1 rise by 1.35 units together with variable 3, for every rise in variable 3, the dependent variable reduced by -1.83 units [18].

The standard error for the unstandardized beta ($SE\beta$) is the following symbol on the table. Standard deviation is similar to this value. A larger number indicates a more dispersed distribution of points from the regression line. Statistical significance is less likely to be found when the numbers are spread out [18].

The standardized beta ($\widehat{\beta}$) is the last symbol on the table. A correlation coefficient works in much the same way. If the relationship is positive, it will range from 0 to 1. If it is negative, it will range from 0 to -1 this depends on the direction. Values closer to 1 or -1 indicate stronger relationships. Since all the variables are on a scale of 0 to 1, it is easy to see which of the variables had the strongest relationship with the dependent variable. Among those variables in the table above, Variable 3 had the strongest correlation/relationship. The standardized beta ($\widehat{\beta}$) can also be described as when a predictor variable is changed by one unit, the standardized beta coefficient changes by the same amount in the outcome variable. In the case of negative beta coefficients, the outcome variable will decrease by the beta coefficient value for every 1-unit increase in the predictor variable [18].

- WHAT IS REGRESSION COEFFICIENT?
  Estimates of the unknown population parameters, also known as regression coefficients, show how predictor/independent variables and dependent/responses are related. A coefficient is the value that multiplies the value of an independent variable in linear regression [26].

- DIFFERENCE BETWEEN BETA AND BETA HAT:
  Beta is an non-standardized symbol (β). A slope is the slope of a line connecting a independent variable with a dependent variable. We have the standardized beta ($\widehat{\beta}$) [26]. It functions similarly to the correlation coefficient. By comparing the beta hat coefficient $\widehat{\beta}$ of each independent variable to the dependent variable, one can estimate the strength of the effect each of these independent variables have on the dependent variable [26]. If the $\widehat{\beta}$ coefficient is higher, then the effect is stronger. Standardized beta coefficients $\widehat{\beta}$ determine effect and the strength of the data with standard deviations. A sample of a population is what we are working with [26]. A data cloud is formed by our sample, we fit the line that minimizes error terms along one dimension that corresponds to the dependent variable. In OLS, based on the column space of the model matrix, It represents a projection of the dependent variable onto that space. $\widehat{\beta}$ symbol is used to denote the estimates of the population parameters, when we have more data points our estimated coefficients $\widehat{\beta_i}$ will be more accurate, For each idealized population coefficient the greater the accuracy estimation can be made, β_i [15].

  The "hat" symbol represents an estimate, not the actual value. $\widehat{\beta}$ is therefore an estimate of β. Symbols have their own conventions: one example is the sample variance, which might be written as. S^2 and not $\widehat{\sigma}^2$. Nevertheless, some people distinguish between biased estimates and unbiased using both. According to the example we mentioned, the $\widehat{\beta}$ values represent parameter estimates for a linear model. According to the linear model, a linear combination of the sample data values x_i generates the outcome variable y. β_i value is assigned to each item (plus some error $\mathcal{E}$) [15].

  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \mathcal{E} \tag{1.9.1}$$

  A linear model can't always determine the "true" β values (possibly, Linear models aren't used to generate the data). From the data, we can still estimate approximate values for y, and these values are called $\widehat{\beta}$ [15].

- WHAT IS BETA [0] AND BETA [1]?
  A regression line's intercept is β_0 while the slope of the regression line is β_1. In practice β_1 does not really exist. β_1 with values above and below it will give us that optimal slope, this slice runs vertically from the dependent variable to the independent variable [26]. If the Gauss-Markov assumption holds, then the residuals will have a nice normal Gaussian distribution. According to the sample. β_1 represents a fit or estimate of $\widehat{\beta}_1$ [15].

In general, Stats can get confusing when different pronumerals are used to mean different things in different contexts! Based on the analysis we discussed earlier $\widehat{\beta}$ means something different in power analysis compared to regression. In regression, the difference between β and $\widehat{\beta}$ relates to whether the coefficients are standardised or unstandardised. β generally refers to the unstandardised coefficient. According to this, we can use the original measurement units to calculate the regression coefficient [3].

For example, imaging we are trying to predict a final exam score based on the number of hours spent studying. If I get $\beta = 2$, this tells me that for every 1 hours study time, I predict an increase in the final exam score of 2. This relationship is in the original units (hours of studying, and exam score). This is useful for predicting things in the real world, but it is difficult to compare different predictors. Predictors might have large *beta* values just because they are measured on a larger scale (compare minutes to hours in the above example).

Standardised regression coefficients do a similar thing, but in a standardised way. The $\widehat{\beta}$ refers to the number of standard deviation changes we would expect in the outcome variable for a 1 standard deviation change in the predictor variable. For example, if I got $\widehat{\beta} = .5$ for hours of study, this would tell me that for every 1 standard deviation increase in hours of study, We can expect .5 standard deviation increase in the exam score. Because this is standardised, $\widehat{\beta}$ make it easier to compare different predictors to see which is more important.

1.10. FITTED VALUES AND RESIDUALS

Significant ideas in regression analysis are the fitted values and residuals . As a rule, the information doesn't fall precisely on a line, so the relapse condition ought to incorporate an express error term $\mathcal{E}$

$$y = \beta_0 + \beta_1 x + \mathcal{E} \tag{1.10.1}$$

We can express the fitted value as the predicted value which typically denoted as $\hat{Y}_i$ (Y-hat). Which represented by this equation

$$\widehat{y_i} = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \tag{1.10.2}$$

$\widehat{\beta}_0$ and $\widehat{\beta}_1$ demonstrates that the coefficients are estimated and known [2].

The "hat" documentation is utilized to separate among gauges and known qualities. therefore the symbol $\hat{\beta}$ (β-hat) is an estimate of the unidentified parameters β. For what reason do an analysts separate between the estimate and the genuine value? The estimate lack certainty, while the genuine value is fixed [2].

The difference between the observation and the predicted values is the residual $\mathcal{E}$

$$\widehat{\mathcal{E}}_{ii} = y_i - \widehat{y_i} \tag{1.10.3}$$

1.11. THE METHOD OF LEAST SQUARES

Simple linear model consist of two parameters β_0 and β_1, they are to be evaluate from the data. Any two data can be utilized to resolve explicitly for the values of the parameters if there are no random error in Y_i. However, the random variation in Y, create individual pair of noticed data points to set out separate outcome. (If the observed data fell precisely on the straight line, then all estimates would be similar). A technique is required that will unite all the information to give out one result that is ”best” by several criterion.

The least squares evaluation procedure utilize the criterion that the result should grant the slightest likely addition of squared deviations of the perceive Y_i from the estimation of the true model given by the results. Let β_0 and β_1 be numerical/statistical evaluation of the parameters β_0 and β_1, individually, then let

$$\widehat{y_i} = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \tag{1.11.1}$$

Be the evaluation mean of y for individual $x_i, i = 1, \ldots, n$. Beware that $\widehat{y_i}$ is acquired by exchanging the evaluation for the parameters in the effective form of the model connecting $\mathcal{E}(y_i)$ to x_i, The least squares theory selected $\widehat{\beta_1}$ and $\widehat{\beta_2}$ that reduce the addition of the residual squares, SS(Res)

$$\begin{aligned} SS(Res) &= \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \\ &= \sum_{i=1}^{n} \mathcal{E}_i^2 \end{aligned} \tag{1.11.2}$$

Whereby $\mathcal{E}_i = (y_i - \widehat{y_i})$ is the noticed residual for the i inspection. The summation which is stipulated by $\sum$ is a general observation in the data place as indicated along the $\sum_{i=1}^{n}$. (The limits of summation are clear from the context when the index of summation is committed). The evaluation for β_1 and β_2 are acquired by utilizing calculus to discover the values that reduce SS(Res).

1.11.1. LEAST SQUARE MODEL IN MATRIX FORM

From the basic knowledge of regression analysis we implement the linear regression model: $\boldsymbol{Y} \sim \mathcal{L}(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$

Linear regression model: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\mathcal{E}}$

The vector $\boldsymbol{X\beta}$ is non-random.

Then

$$\begin{aligned} &E(\boldsymbol{Y}) = E(\boldsymbol{X\beta} + \boldsymbol{\mathcal{E}}) = E(\boldsymbol{X\beta}) + E(\boldsymbol{\mathcal{E}}) = \boldsymbol{X\beta} + \boldsymbol{0} = \boldsymbol{X\beta} \\ &\text{var}(\boldsymbol{Y}) = \text{var}(\boldsymbol{X\beta} + \boldsymbol{\mathcal{E}}) = \text{var}(\boldsymbol{\mathcal{E}}) = \sigma^2 \boldsymbol{I} \end{aligned}$$

$\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T$ is random vector and $\boldsymbol{y} = (y_1, \cdots, y_n)^T$ is its realization.

The parameters $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)^T$ are estimated using the least squares method - the sum of squares is minimized, so we look for the minimum: $\sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{k} x_{i,j}\beta_j \right)^2$.

Then

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in R^k} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{k} x_{i,j}\beta_j \right)^2 = \arg\min_{\substack{i=1 \\ \boldsymbol{\beta} \in R^k}}^{n} \left(Y_i - \boldsymbol{X}^i \boldsymbol{\beta} \right)^2 = \arg\min_{\boldsymbol{\beta} \in R^k}^{k} (\boldsymbol{Y} - \boldsymbol{X\beta})^T (\boldsymbol{Y} - \boldsymbol{X\beta})$$

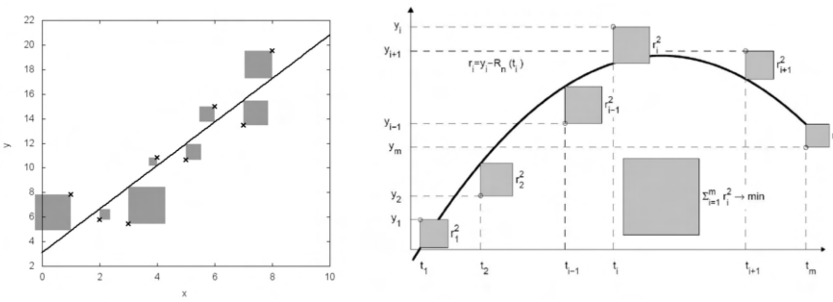


Figure 1.6: Sample diagram

Holds:

The statistics that estimate the parameters $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1, \cdots, \widehat{\beta}_k\right)^T$ are marked: $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1, \cdots, \widehat{\beta}_k\right)^T$ Statistics $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1, \cdots, \widehat{\beta}_k\right)^T$ using least squares method can be expressed in the form:

$$\widehat{\widehat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Now to prove this let's consider the Sum of the Square Error (SSE).

Each data point is subject to some error of prediction due to the coefficients $\widehat{\beta}$, which form a vector:

$$\mathcal{E}(\widehat{\beta}) = \mathrm{y} - \mathrm{x}\widehat{\beta}$$

(By checking this, you can verify that it subtracts the $n \times 1$ matrix from the $n \times 1$ matrix.) Based on the mean squared error, we derived the least squares estimator,

$$SSE(\widehat{\beta}) = \sum_{i=1}^{n} \mathcal{E}_i^2(\widehat{\beta})$$

Using our matrices, how can we express this? Let us claim that the appropriate form would be

$$\mathrm{SSE}(\widehat{\beta}) = \mathcal{E}^T\mathcal{E}$$

This can be seen by looking closely at what matrix multiplication truly entails:

$$SSE = \mathcal{E}_1^2 + \mathcal{E}_1^2 + ... + \mathcal{E}_n^2 = [\mathcal{E}_1\mathcal{E}_2...\mathcal{E}_n]\begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ . \\ . \\ \mathcal{E}_n \end{pmatrix}$$

Where, $\mathcal{E} = y - \widehat{y}$ (i.e Residual vectors = vectors containing the value of independent variable - estimated y vectors contain estimated values)

$$\widehat{y} = X\widehat{\beta}$$

so,

$$\mathcal{E} = y - X\widehat{\beta}$$

Consider

$$SSE = \mathcal{E}^T\mathcal{E} = (y - X\widehat{\beta})^T(y - X\widehat{\beta})$$

implies

$$(y^T - (X\widehat{\beta})^T)(y - X\widehat{\beta})$$

$$SSE = (y^T - X^T\widehat{\beta}^T)(y - X\widehat{\beta})$$

$$= y^T(y - X\widehat{\beta}) - X^T\widehat{\beta}^T(y - X\widehat{\beta})$$

$$y^Ty - y^TX\widehat{\beta} - X^T\widehat{\beta}^Ty + X^T\widehat{\beta}^TX\widehat{\beta} \tag{1.11.3}$$

NB: $y^TX\widehat{\beta}$ is a scalar and any scalar or constant is a matrix of order 1x1 so,

$$(y^TX\widehat{\beta}) = (y^TX\widehat{\beta})^T = \widehat{\beta}^TX^Ty$$

Recall from:

$$y^Ty - y^TX\widehat{\beta} - X^T\widehat{\beta}^Ty + X^T\widehat{\beta}^TX\widehat{\beta}$$

putting

$$(y^TX\widehat{\beta}) = (y^TX\widehat{\beta})^T = \widehat{\beta}^TX^Ty$$

into equ (1.11.3), we get

$$SSE = y^Ty - \widehat{\beta}^TX^Ty - X^T\widehat{\beta}^Ty + X^T\widehat{\beta}^TX\widehat{\beta}$$

$$y^Ty - 2\widehat{\beta}^TX^Ty + X^T\widehat{\beta}^TX\widehat{\beta} \tag{1.11.4}$$

Now, we have to minimize RSS in equ (2) both sides partially with respect to $\hat{\beta}$

$$\frac{\partial}{\partial\widehat{\beta}}(SSE) = \frac{\partial}{\partial\widehat{\beta}}(y^Ty - 2\widehat{\beta}^TX^Ty + X^T\widehat{\beta}^TX\widehat{\beta})$$

$$\frac{\partial}{\partial\widehat{\beta}}y^Ty - \frac{\partial}{\partial\widehat{\beta}}2\widehat{\beta}^TX^Ty + \frac{\partial}{\partial\widehat{\beta}}X^T\widehat{\beta}^TX\widehat{\beta}) \tag{1.11.5}$$

Note that:

$$\frac{\partial}{\partial\widehat{\beta}}y^Ty = 0$$

$$\frac{\partial}{\partial\widehat{\beta}}2(\widehat{\beta}^TX^Ty) = 2X^Ty$$

$$\frac{\partial}{\partial \widehat{\beta}} X^T \widehat{\beta}^T X \widehat{\beta} = 2X^T X \widehat{\beta}$$

putting this values in equ (1.11.5), we get

$$\frac{\partial (SSE)}{\partial \widehat{\beta}} = \frac{\partial}{\partial \widehat{\beta}} y^T y - \frac{\partial}{\partial \widehat{\beta}} (2\beta^T X^T y) + \frac{\partial}{\partial \widehat{\beta}} \beta^T X^T X \widehat{\beta} \tag{1.11.6}$$

$$\frac{\partial}{\partial \widehat{\beta}} (SSE) = 0 - 2X^T y + 2X^T X \widehat{\beta}$$

$$\frac{\partial}{\partial \widehat{\beta}} (SSE) = 0$$

$$-2X^T y + 2X^T X \widehat{\beta} = 0$$

or

$$X^T y = X^T X \widehat{\beta}$$

premultiplying both sides by $(X^T X)^{-1}$

$$\widehat{\beta} = (X^T X)^{-1} X^T y \tag{1.11.7}$$

Therefore

$$(\widehat{\beta}_0, \widehat{\beta}_1)^T = (X^T X)^{-1} X^T y \tag{1.11.8}$$

The matrix equation that we've gotten yields both coefficient estimates. Assuming this is correct, the equation above should in fact reproduce the least-squares estimates we've already obtained, so it follows that

$$\widehat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{1.11.9}$$

and

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} \tag{1.11.10}$$

The slope estimate can also implies that

$$\widehat{beta}_1 = \frac{S_{xy}}{S_{xx}}$$

where $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) \left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}$ and where

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

.

1.12. PROPERTIES OF LEAST SQUARE ESTIMATOR

Least squares estimators are characterized by the ability to reduce total squared residuals. Nevertheless, there are more properties. If we compute these properties in the way just described, they are always true no matter what assumptions are made [33].

From equation (1.9.7)

$$(X^T X)\widehat{\beta} = X^T y \tag{1.12.1}$$

Put $y = X\widehat{\beta} + \mathcal{E}$ for substitution

$$\begin{aligned} \left(X^T X\right)\widehat{\beta} &= X^T(X\widehat{\beta} + \mathcal{E}) \\ \left(X^T X\right)\widehat{\beta} &= \left(X^T X\right)\widehat{\beta} + X^T\mathcal{E} \\ X^T\mathcal{E} &= 0 \end{aligned} \tag{1.12.2}$$

$X^T\mathcal{E}$ seems to be the case of

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} X_{11} \times e_1 + X_{12} \times e_2 + \dots + X_{1n} \times e_n \\ X_{21} \times e_1 + X_{22} \times e_2 + \dots + X_{2n} \times e_n \\ \vdots \\ \vdots \\ X_{k1} \times e_1 + X_{k2} \times e_2 + \dots + X_{kn} \times e_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

Number of properties we can derive from $X^T\mathcal{E} = 0$.

1. Relative to the residuals, X values are uncorrelated:

   $X^T\mathcal{E} = 0$ For all columns, it means x_k of X, $x_k^T\mathcal{E} = 0$ As a result, none of the regressors and residuals exhibit sample correlations. The fact that X is not correlated along the disturbances does not mean it is uncorrelated; we will have to presume that it is uncorrelated [33].

   If there is a constant in X, the topmost column (i.e. X_1) will be a row of ones.

2. Zero is the result of the residuals sum:

   If there is a constant in X (i.e. X_1), then the topmost column is a column of ones. The first element of the $X^T\mathcal{E}$ vector must be zero for $X_{11} \times \mathcal{E}, X_{12} \times \mathcal{E}, ...X_{1n} \times \mathcal{E}_i$ to be true [33].

3. Relative residuals have a sample mean of zero:

   The former property is directly connected to this one $\bar{\mathcal{E}}_i = \frac{\sum_{i=1}^{n_1} \mathcal{E}_i}{n} = 0$

4. In the regression hyperplane, the observed values pass through their means ($\bar{X}$ and $\bar{y}$):

   This statement follows the fact that $\bar{\mathcal{E}}_i = 0$. Recall that $\bar{\mathcal{E}}_i = \bar{y} - \bar{x}\widehat{\beta}$ In other words, we get $\bar{\mathcal{E}}_i = \bar{y} - \bar{x}\widehat{\beta} = 0$ when we multiply by the number of observations. This means that $\bar{y} = \bar{x}\widehat{\beta}$ which display that the regression hyper plane pass through the point of means of the data [33].

5. There is no correlation between y and the residuals:

   This implies $X\widehat{\beta}$ for $\widehat{y} = X\widehat{\beta}$ Through this we obtain

$$y^T\mathcal{E} = (X\widehat{\beta})^T\mathcal{E} = \beta^T X^T\mathcal{E} = 0 \tag{1.12.3}$$

   In conclusion, $X^T\mathcal{E} = 0$ is considered in this final development [33].

6. It is predicted that the mean of the observed Y's will equal the mean of the predicted Y's for the sample i.e. $\bar{\widehat{y}} = y$:

   There is no exception to these properties. We minimize the sum of squared residuals, so you cannot infer the total disturbances or mean disturbances are zero based on the fact that the residuals are zero [33].

   We do not know anything about $\widehat{\beta}$ Besides fulfilling all the characteristics listed above, it also offers the following.

For us to be able to draw any conclusions about β (the true population parameters) from $\widehat{\beta}$ (our estimate of the true parameters), there are some assumptions we need to make about the true model. $\widehat{\beta}$ comes from our sample, but we are interested in learning more about the true parameters.

1.12.1. GAUSS-MARKOV THEOREM

According to the Gauss-Markov Theorem, there is no linear and unbiased estimator of the β coefficients that has a small sampling variance. One of the best linear, unbiased, and efficient estimators is the least squares estimator [33].

- Show that $\widehat{\beta}$ is an unbiased estimator of β:

  We notice from earlier that $\widehat{\beta} = (X^T X)^{-}1X^T y$ implies

$$\widehat{\beta} = (X^T X)^{-1}X^T(X\beta + \mathcal{E}) \tag{1.12.4}$$

$$\widehat{\beta} = \beta + (X^T X)^{-1}X^T\mathcal{E} \tag{1.12.5}$$

  The fact that $(X^T X)^{-1}X^T X = I$ immediately indicates that the least square estimate is unbiased as long as X is fixed (non-stochastic), thus giving as:

$$\begin{aligned} E[\hat{\beta}] &= E[\beta] + E\left[\left(X^T X\right)^{-1} X^T\mathcal{E}\right] \\ &= \beta + \left(X^T X\right)^{-1} X^T E[\mathcal{E}] \end{aligned} \tag{1.12.6}$$

  In other word $E[\mathcal{E}] = 0$ by presumption or X is stochastic however independent of $\mathcal{E}$ so that we have [33]:

- Show that $\widehat{\beta}$ is a linear estimator of β: From Equation. (1.10.12), we posses:

$$\widehat{\beta} = \beta + (X^T X)^{-1} X^T \mathcal{E} \tag{1.12.7}$$

Since we can state $\widehat{\beta} = \beta + A\mathcal{E}$ whereby A $= (X^T X)^{-1} X^T$

Based on the disturbances, $\widehat{\beta}$ is a linear function. By using the explanation that we offer, we can determine that it is a linear estimator [33].

1.12.2. CONFIDENCE INTERVAL

From the regression line: $Y = \beta_0 + \beta_1 X$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{D} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

$$\boldsymbol{X}^T y = \boldsymbol{g} = \begin{pmatrix} \sum_{i=i}^{n} y_i \\ \sum_{i=i}^{n} x_i y_i \end{pmatrix}$$

$$\det(\boldsymbol{D}) = n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2$$

$$\boldsymbol{D}^{-1} = \frac{1}{\det(\boldsymbol{D})} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

$$\beta_1 = \frac{1}{\det(\boldsymbol{D})} \left(n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}, \quad y = \beta_0 + \beta_1 x$$

$$S_{\min}^{*} = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x))^2$$

$$S_{\min}^{*} = \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i$$

$$s^2 = \frac{S_{\min}^{*}}{n-2}$$

- Interval estimate for the mean value of y with reliability $1-\alpha$ for x_0 :

$$\left\langle y_0 - t_{1-\alpha/2} s\sqrt{v^*}; y_0 + t_{1-\alpha/2} s\sqrt{v^*}\right\rangle$$

  where $y_0 = \beta_0 + \beta_1 x_0$,

$$v^* = \frac{1}{n} + \frac{n(x-\bar{x})^2}{\det(\boldsymbol{D})}$$

  and $t_{1-\alpha/2}$ is a quantile of the Student 's distribution with $n-2$ degrees of freedom. [40].

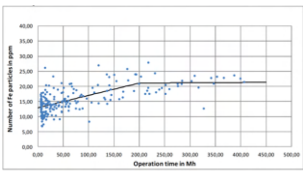
- Confidence interval for individual value of y with reliability $1-\alpha$ for x_0 :

$$\left\langle y^0 - t_{1-\alpha/2} s\sqrt{v^*+1}; y^0 + t_{1-\alpha/2} s\sqrt{v^*+1}\right\rangle$$

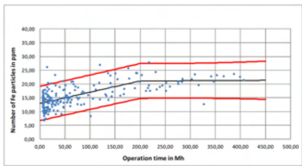
  where $y_0 = \beta_0 + \beta_1 x_0$,

$$v^* = \frac{1}{n} + \frac{n\left(x-\bar{x}\right)^2}{\det(\boldsymbol{D})}$$

  and $t_{1-\alpha/2}$ is a quantile of the Student 's distribution with $n-2$ degrees of freedom [40].



(a) Sample of a Linear regression line



(b) Sample of a Linear regression line and Confidence interval

1.12.3. MODEL ASSUMPTIONS

:

After Linear Regression obtain whether one or more predictor variables describe the dependent variable and thus it include 5 assumptions:

1. Linearity: The relationship between dependent variable, independent variable, and disturbance can be described by a linear function. [19]
2. Random sample: we posses a random sample of size n $(x_i, y_i) : i = 1, .., n)$, Whereby the observations are independent to one another. [19]
3. No perfect collinearity: Any of the independent variables is constant, also there are no exact linear relationships between the independent variables. [19]

4. Exogeneity: Given any value of the independent variable the disturbance term has an expected value of zero. This is the case $E(\mathcal{E}|x_i) = 0$ [19]

5. Homoscedasticity: Given any value of the independent variables the disturbance term has the same variance. That is to say $Var(\mathcal{E}|x_i) = \sigma^2$ [19]

With all these assumptions examined while building the model, the model can be build and we can do our predictions for the dependent factor. For any kind of machine learning model, we must know if the variable examine for the model are accurate and have been analysed by a metric. In the event of Regression analysis, the statistical measure that analyse the model is named the coefficient of determination that is represented as r^2 [25].

Coefficient of determination is the segment of the overall variation in the dependent variable which is described by variation in the dependent variable. A high value of r^2 better is the model along the independent variables being examined for the model [25].

$$r^2 = \frac{SSR}{SST} \tag{1.12.8}$$

Note: r^2 is the range of $0 \leq r^2 \leq 1$

1.13. NON-LINEAR REGRESSION

We have been utilizing the linear least squares method to fit a straight line to data points that are informative, but our data is more focused on Non-linear model. Now and again the relationship that we want to genuinely model is curved rather than flat. For Example: Assuming something is developing dramatically, and that implies developing at a consistent rate, the connection among the X and Y is the curve, similar to that displayed in Figure 2.2 [29].

Building a new variables appropriately, the curved function of a unique variables can be communicated as a linear function of the new variables.To fit something like this, we really want non-linear regression. Frequently, you can adjust straight least squares to do this. The technique is to make new factors from your in. formation. The new factors are nonlinear elements of the variables in your information [19].

Looking into two famous non-linear model that are agreeable to this method:

Equation	Interpretation	Linear Form
$\mathrm{Y} = \mathrm{A}e^{\beta \mathrm{X}}\mathcal{E}$	Y is developing (or contracting at a) consistent relative pace of β.	$\ln(\mathrm{Y}) = \ln(\mathrm{A}) + \beta\mathrm{X} + \ln(\mathcal{E})$
$\mathrm{Y} = \mathrm{AX}^{\beta}\mathcal{E}$	The versatility of Y with deference to to X is a consistent, β.	$\ln(\mathrm{Y}) = \ln(\mathrm{A}) + \beta \cdot \ln(\mathrm{X}) + \ln(\mathcal{E})$

Take into account the primary condition which describe the describes exponential growth

$$\mathrm{Y} = \mathrm{A}e^{\beta \mathrm{X}}\mathcal{E}$$

- β is the rate of growth.
- $\mathcal{E}$ is an unexpected error

Assuming you're taking the logarithm of the 2 aspects of that situation, you get

$$ln(Y) = ln(A) + \beta X + ln(\mathcal{E})$$

.

This circumstance has logarithms in it, but they relate in an instant way. It is located within the structure

$$y = \beta + \beta X + error$$

, then again, surely y, a, and the error are logarithms [19].

Closely, examine the second equation, Y = AX $\mathcal{E}$. This is a constant-elasticity equation (more reason why we call it that after), generally utilized for demand curves. Take the logarithm of the two sides of that equation then you get ln(Y) = ln(A) + β ln(X) + ln($\mathcal{E}$). For this equation, if you construct the variable ln(Y) including a variable for the base-e logarithm of X, written as ln(X), you can utilize the regular least squares method to place the curve Y = AX to your data [19].

The evaluation of β in $Y = Ae^{\beta x}\mathcal{E}$:

β is the parameter you are mostly interested in, regularly. Your evaluation of β is your evaluation of the relative change in Y connected with a unit change in X. Mathematically, if X moves up by 1, Y is multiplied by $e^{\mathcal{E}}$. The reason is $Ae^{\beta(x+1)}$ equals $Ae^{\beta x}e^{\beta}$, which is Y is multiplied by e^{β} . That might not seems to resemble "relative change," however it is, if you are utilizing continuous mix [19].

1.13.1. NON-LINEAR EQUATION IN LINEAR FORM UTILIZING THE NATURAL LOGARITHMS

To change $Y = Ae^{\beta x}\mathcal{E}$ to a linear equation, take the natural log of the two sides:

$ln(Y) = ln(Ae^{\beta x}\mathcal{E})$ Make use of the rules above and we obtain:

$$ln(Y) = ln(A) + \beta x + ln(\mathcal{E})$$

To execute this, construct a new variable y = ln(Y). (The Y inside the actual calculation is the 'big Y.' The current variable is the 'little y.') In addition, interpret v as ln($\mathcal{E}$) and also as ln(A).

Concerning the non-linear model, to employing the least squares method, it is important to presume that using v as an expression for errors and also as an expression for linear regression. One of these presumption is that v's expected value is 0. That is the reason we presume that the mean of $\mathcal{E}$ is 1. That suit because ln(1) = 0. $\mathcal{E}$ will never be 0 or negative, however v may take on positive or negative values, because if $\mathcal{E}$ is lower than 1, v = ln($\mathcal{E}$) is lower than 0 [19].

1.14. CHANGE POINT ANALYSIS

Numerous fields, including medicine, aerospace, finance, business, meteorology, and entertainment rely on time series analysis. Observations of a system's behaviour over time are called time series data. As external events occur, as well as structural changes within dynamics and distribution, these behaviors may change over time. Detecting change points in a time series when one of its properties changes is the concept of change point detection (CPD). Change point detection is similar to segmentation, edge detection, event detection, and anomaly detection, which are occasionally applied. The search for change points is closely related to the problem of change point estimation and change point mining. The emphasis of change point estimates, however, is to describe the nature and degree of known changes in time series instead of identifying the change itself. Change point estimation is concerned with modeling and interpreting known changes rather than identifying that one has occurred and it's also played an role in the model of statistical analysis. Throughout this thesis, we examine recent research in the area of change point detection/analysis [8].

Breakpoints segmentation, structural breaks, regime switches, and detecting disorder are another names for changing points while on the other hand In order to detect whether a change has occurred, change points are analysed on time ordered data. It further provides confidence levels and confidence intervals for changes and for time, and it determines the number of changes [21]. Change point analysis is a technique for identifying a point of entry or beginning in relationships between two variables. An analysis of a distribution of values is intended to identify a point where values before and after the point differ. A change-point analysis can be carried out on the x axis of a stress or response relationship to find the point at which the characteristics of the y axis change - suggesting a shift in variance or a change in slope of the relationship [22].

To put it a bit more mathematically

Let φ be a data set and let m be the point of the data, For data $y_1, \cdots, y_m$, if a change point exists at φ, then $y_1, \cdots, y_\varphi$ differ from $y_{\varphi+1}, \cdots, y_m$ in some way.

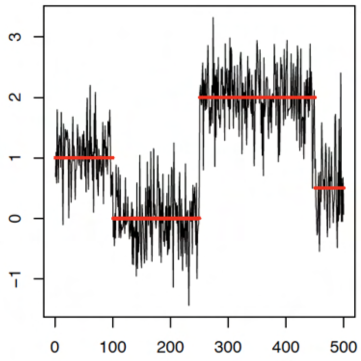
1.14.1. TYPES OF CHANGE POINTS ANALYSIS

Changes are typically detected using control charts. Control charts differ from change-point analyses in that they are meant to be updated as data is gathered for each point. In contrast to a change-point analysis, a control chart is meant to be updated after each data point is collected. Both methods can be used in conjunction with each other [21].

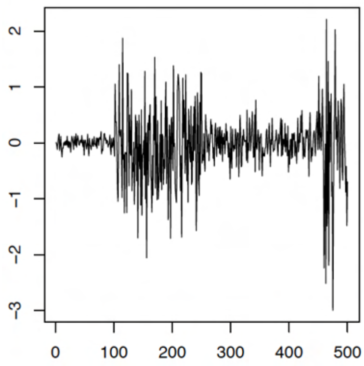
Change point analysis can take many different forms but the most common forms are

- Change in mean
- Change in variance
- Change in trend

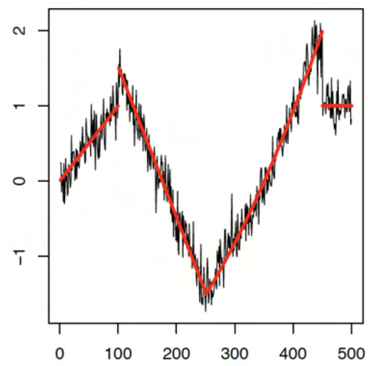
These 3 forms are given in the pictures below



(a) Change in mean [9]



(b) Change in variance [9]



(c) Change in trend [9]

1.14.2. AREAS OF APPLICATION of CHANGE POINT ANALYSIS

Change point detection/analysis (CPD or CPA) has been studied in the fields of data mining, statistics and computer science for several decades. This problem has broad application in many fields. There are many real-world problems covered by this problem [21]. Let's look at a few examples.

1. Speech recognition: This is the process of transcribing utterances/spoken words into text. We are using change point detection methods to recognize silence, sentences, words, and noise boundaries among audio segments [8].

2. Human activity analysis: Based on characteristics of sensor-based data observed by smart homes or mobile devices. It can be formulated as detection of activity transitions or breakpoints. Human interaction can be enhanced by segmenting activities based on these change points , and assessing health status-related behavioral changes [8].

3. Climate change detection: Due to the possibility of climate change and the increase of greenhouse gases in the atmosphere, the use of change point detection to analyze, monitor, and predict climate has gained increasing importance in recent decades.[8].

4. Medical condition monitoring: Physiological variables like electroencephalograms (EEGs), electrocardiograms (ECGs), and heart rate are monitored constantly to identify trends automatically, in real-time. Studies examine changes in specific areas of medicine, such as sleep disorders, epilepsy, magnetic resonance imaging (MRI) interpretations, and understating brain activity. [8].

5. Image analysis: The purpose of video-based surveillance is to collect video data over time, or image data. A change-point problem can be formulated to detect abrupt events such as security breaches. A digital image is encoded at each time point as the observation [8].

1.15. LAGRANGE MULTIPLIER APPLICATION

In this section, we are going to talk briefly on what Lagrange Multiplier is all about. The problem of optimization that see to maximizing or minimizing a real function, play a key role in the physical world. This can be categorize into two which are constrained optimization problems and unconstrained optimization problems. Most practical that are used on economics, engineering, science also in our daily life can be considered as constrained optimization problems, like the minimizing of the energy of a particle in physics [22].

Unconstrained problems, the stationary points theory provides the important condition to get the utmost points of the objective function $f(x_1, \cdots, x_n)$ This stationary points are the point whereby the gradient ∇f Is zero which means each partial derivatives is zero. Every variables in $f(x_1, \cdots, x_n)$ are independent, therefore they can be arbitrarily ready to search for the utmost of f However, the arbitration of the variable is nonexistent when it comes to constrained optimization problems. Optimization can be prepared into an adequate form like [22]

$$\min f\left(x_1, \cdots, x_n\right) \tag{1.15.1}$$

As a result of:

$$\mathrm{G}\left(x_1, \cdots, x_n\right) = 0 \tag{1.15.2}$$

$$\mathrm{H}\left(x_1, \cdots, x_n\right) \leq 0 \tag{1.15.3}$$

Whereby G, H are function vectors. Variables are restricted to the feasible range, based on the constraints satisfied [22].

The use of substitution can be a good approach to solving optimization problems. Nevertheless, it can only be taken advantage of when solving equality constrained optimization problems and can be ineffective sometimes when solving nonlinear constrained optimization problems where it is difficult to get explicit expressions of variables that terminate in the objective functions. A method for solving constrained nonlinear optimization problems is the Lagrange multiplier method, which is named after Joseph Loius Lagrange. It can be used when inequality and equality constraints are present [22].

For nonlinear problems with equality constraints we examined the Lagrange multiplier method. The mathematical proof and geometry explanation are presented. In addition, the method is extended to include inequality constraints. Nonlinear optimization problems without inequality constraints have the standard form of

$$\min f\left(x_1, \cdots, x_n\right) \tag{1.15.4}$$

As a result of:

$$\mathrm{G}\left(x_1, \cdots, x_n\right) = 0 \tag{1.15.5}$$

Suppose, $\mathrm{G} = \left[G_1\left(x_1, \cdots, x_n\right) = 0, \cdots, G_k\left(x_1, \cdots, x_n\right) = 0\right]^T$, be the constraints vector. The Lagrange function F is constructed as:

$$F(\mathrm{X}, \lambda) = f(\mathrm{X}) - \lambda \mathrm{G}(\mathrm{X}) \tag{1.15.6}$$

Supposed, X $= [x_1, \ldots, x_n]$, are the variable vector, $\lambda = [\lambda_1, \cdots, \lambda_k], \lambda_1, \cdots \lambda_k$ are refered to as Lagrange multipliers.

If λ and f satisfy the following extreme points:

$$\nabla F = 0 \tag{1.15.7}$$

then:

$$\frac{\partial f}{\partial x_i} - \sum_{m=1}^{k} \lambda_m \frac{\partial G_m}{x_i} = 0, i = 1, \ldots n \tag{1.15.8}$$

and

$$\mathrm{G}(x_1, \cdots, x_n) = 0 \tag{1.15.9}$$

In the constrained nonlinear optimization problem, the Lagrange multiplier method describes important conditions. Economic, engineering, and scientific problems have been successfully resolved with the Lagrange multiplier method. In situations where the objective function f and constraints G have meaning, there is sometimes an identifiable significance to Lagrange multipliers. In economics, if profit subject is being maximize to a defined resources, the resources marginal value is λ, which is occasionally refer to as shadow price. more specifically, the Lagrange multiplier is the ratio in which the optimal value of the objective function f changes if the constraints are exchanged. Lagrange multiplier method plays a major role in power systems' economy dispatch, or the, or λ dispatch problem, which is a cross between economics and engineering. This problem has the objective function of minimizing the generating costs, and the variables are subject to the constraint of power balance [22].

Nonlinear optimization problems can be dealt with efficiently using the Lagrange multiplier method since it can cope with both inequality constrained and equality constrained nonlinear optimization problems.Computational programming methods include the interior point method, the barrier, augmented Lagrange method, and penalizing.In economics, engineering, science and our daily lives, Lagrange multipliers methods and their extended methods are used widely [22].

1.16. LITERATURE REVIEW

Regression analysis is an important statistical tool to analyze the data and developed a meaningful and optimised relationship between the dependent and independent variables. In this study a relatively new approach is used to analyze the data of COVID-19 deaths in Italy. The purpose of the study is to analyze the data in which the dependence of one variable on the other can not be simply explained or quantified by a simple regression function. The area of interest is to develop a method to quantify relationship between the variables especially when there are change points in the data.

The history of regression analysis development starts with the method of least square approximation which was first mentioned by Legendre in his book [37] published in 1805. The method was further developed by Gauss who published a Gauss-Markov Theorem [32] in 1821. The major development of regression analysis took place in the 19th and the 20th century which revolutionized the analysis of complex and huge data. Despite huge developments, regression analysis is still a growing and active area of research. The change point regression analysis is a relatively new area of regression analysis with on-going research.

Bhattacharya et al. [30] worked on the aspects of change point analysis by dividing the data into homogeneous segments. He tested the concepts of no change, point and interval estimation of a change-point, non-parametric model changes, detection of change in distribution of sequential data and the changes in regression model. Jushan Bai [28] studied the change point estimation for least square method with multiple regressions. The method is used to analyze the response of market interest rates to discount rate changes.The approach is used to investigate the reaction of market interest rates with respect to discount changes in rate. It included the derivation of analytical density function and the cumulative distribution function for the general distribution.

Jie Chen [31] propose a new criteria called Schwarz Information Criterion (SIC), to locate change point within the straightforward simple regression model, further as in the multiple linear regression model. the tactic is then applied to a monetary information set, and a change point is detected with success. Muller et al. [34] considered a smooth regression model and proposed a two-step calculator for locating change point purpose and studied its straight line convergence properties.In a 1st step, initial pilot estimates of the modification purpose and associated asymptotically shrinking intervals that contain actuality change point with chance convergence to one are obtained within the second step, a weighted mean distinction counting on the assumed location of the change point is maximized among these intervals and therefore the maximising argument is then the ultimate change point estimator. Godfrey et al. [36] looked at the properties of various tests regarding logarithm and linear (or log-linear) regression models. The test procedures could also be classified as the tests that exploit the very fact that the 2 models are per se non-nested, tests supported the Box-Cox knowledge transformation and the diagnostic tests of purposeful type mis-specification against an any old alternative. The small-sample properties of many tests are investigated through a Monte Carlo experiment, as is their efficiency to non-normality of the errors. Andrews et al. [27] considered checks for parameter instability and unknown change point. The results applied to a good category of constant quantity models that are appropriate for estimation by generalized technique of moments procedures. The paper considers Wald, Lagrange multiplier, and chance ratio-like tests. every test implicitly uses an estimate of a change point. The change point may be not known and exist between a fixed interval. The assessments were found to perform pretty well in a Monte Carlo test suggested someplace else.

Li et al. [38] used the saddle-point approximations to detect the change point in the data. Mean-shift problem was considered and the probability of change point was calculated for every point of location in the available data. The saddle-point approximation primarily based distribution of the test statistic which was worked out in the paper is of unbiased interest and attractive method. The results were also confirmed by the simulations and the real world data.Julious et al [35] introduced a two-line model for known change point location to detect the change in the coefficient of regression using F-test. He concluded that that when the change point location is not known the resulting parametric distribution from the F-test is not as expected. He proposed the non-parametric bootstrap methods to overcome the shortcomings in the method.

All the above mentioned studies shows the fact that there are continuous advancements in improving the regression analysis methods especially when there are intervals in the data separated by the change point. The focus of this thesis is to present a method of analysing such data where single regression function is not enough to explain the inter-dependencies of the variables involved. In this paper the data of COVID-19 deaths in Italy over a period of time is analyzed by the application of regression analysis. The data is divided into two sections and two separate regression functions were found and then optimised under the condition that the two functions would become equal at the selected change point. This optimisation is achieved using Lagrange multiplier function which is applied in order to minimize the squared error of the two regression lines under the constraint that the two lines would meet at the arbitrary user selected change point.

2. DESCRIPTION OF PROBLEM AND IT'S SOLUTION

2.1. SCOPE OF THE STUDY

When examining our data, we encounter situations where it is not appropriate to use a single expression to describe the dependence between variables, but it is necessary to divide the data into several sections and find an expression of the dependence for each of them. The problem is both to find the points at which the dependence changes and the expressions that describe these individual dependencies.
The study is structured around the application of a change-point analysis methodology on linear regression to study if there is a change in the data as well as modelling the individual dependencies and showing the derived solution on a plot using the specific Covid-19 Italian Data.

2.2. OBJECTIVE OF THE STUDY

The specific objectives of this research are structured about four (4) major tasks which are:

- Find a single line expression/model that describes the data
- Evaluate the point of change in the data using the above stated model.
- Find the individual expression/model that described the individual
- Dependencies as evaluated in the change-point analysis

2.3. DATA DESCRIPTION: The Italy covid-19 data

Italy, is a part condition of the European Union and a famous vacationer location, joined the rundown of Covid impacted nations on 30 January when two COVID-19 positive cases were accounted for in Chinese travelers. Italy COVID cases arrived at 59,138 on 23 March, denoting the greatest Covid episode outside Asia. Italy is additionally the second most impacted Covid country on the planet with the cases expanding at a higher rate than some other nation [13]. Italy was the main Western country to encounter a significant Covid episode and therefore confronted enormous scope well being and financial difficulties.

The Italian government upheld a wide arrangement of homogeneous mediations broadly, in spite of the contrasting occurrences of the infection all through the nation [5]. Liliana, Antonio, Alessandra, & Saverio, (2020). Expounded on the circumstance and in their works, they said in the current environment, there is a lot of talk about "legends". "The legends of this conflict are the Doctors" is a repetitive figure of troop in Italy and the other part of the world these days. However basically as Medical workers, very much like the nurses and all of the other health workers who continue to do their work well aware of the high risk of contagion in healthcare settings [6].

The impacts of the pandemic on Italy and the Italian public overall are huge. Italy is nineteenth among the main 30 nations getting carrier explorers from high-hazard urban communities from Covid in China, as indicated by World Pop's fundamental examination of the nCoV spread. The Italian government went to lengths, for example, screening and suspending significant local area occasions during early seasons of the Covid flare-up, and has at last reported conclusion of instructive foundations and cleanliness/sterilization measures at air terminals. The Italian National Institute of Health (Istituto Superiore di Sanità) suggested social removing and recognized that the country's bigger matured populace represents a test. Numerous different nations including the US have, in the interim, encouraged to briefly keep away from movement to Italy, except if fundamental [13].

The data used for this research is the COVID-19 new death data obtained from the Italian covid-19 outbreak data. The software used for the solutions of this work is the Excel programming tool.

Italy

order	Date	Active Case	Total Deaths	New Deaths	ln(New Deaths)
1	22.02.2020	79	1	1	0
2	23.02.2020	150	2	1	0
3	24.02.2020	229	3	3	1,098612289
4	25.02.2020	322	6	4	1,386294361
5	26.02.2020	445	10	2	0,693147181
6	27.02.2020	650	12	5	1,609437912
7	28.02.2020	888	17	4	1,386294361
8	29.02.2020	1128	21	8	2,079441542
9	01.03.2020	1694	29	5	1,609437912
10	02.03.2020	2036	34	18	2,890371758
11	03.03.2020	2502	52	27	3,295836866
12	04.03.2020	3089	79	28	3,33220451
13	05.03.2020	3858	107	41	3,713572067
14	06.03.2020	4636	148	49	3,891820298
15	07.03.2020	5883	197	36	3,583518938
16	08.03.2020	7375	233	133	4,890349128
17	09.03.2020	9172	366	97	4,574710979
18	10.03.2020	10149	463	168	5,123963979
19	11.03.2020	12462	631	196	5,278114659
20	12.03.2020	15113	827	189	5,241747015
21	13.03.2020	17660	1016	250	5,521460918
22	14.03.2020	21157	1266	175	5,164785974
23	15.03.2020	24747	1441	368	5,908082938
24	16.03.2020	27980	1809	349	5,855071922
25	17.03.2020	31506	2158	345	5,843544417
26	18.03.2020	35713	2503	475	6,163314804
27	19.03.2020	41035	2978	427	6,056784013
28	20.03.2020	47021	3405	627	6,440946541
29	21.03.2020	53578	4032	793	6,675823222
30	22.03.2020	59138	4825	651	6,478509642
31	23.03.2020	63927	5476	601	6,398594935
32	24.03.2020	69176	6077	743	6,610696045
39	31.03.2020	105792	11591	837	6,72982407
40	01.04.2020	110574	12428	727	6,588926478
41	02.04.2020	115242	13155	760	6,633318433
42	03.04.2020	119827	13915	766	6,64118217
43	04.04.2020	124632	14681	681	6,523562306
44	05.04.2020	128948	15362	525	6,263398263
45	06.04.2020	132547	15887	636	6,455198563
46	07.04.2020	135586	16523	604	6,403574198
47	08.04.2020	139422	17127	542	6,295266001
48	09.04.2020	143626	17669	610	6,413458957
49	10.04.2020	147577	18279	570	6,345636361
50	11.04.2020	152271	18849	619	6,428105273
51	12.04.2020	156363	19468	431	6,06610809
52	13.04.2020	159516	19899	566	6,338594078
53	14.04.2020	162488	20465	602	6,400257445
54	15.04.2020	165155	21067	578	6,359573869
55	16.04.2020	168941	21645	525	6,263398263
56	17.04.2020	172434	22170	575	6,354370041
57	18.04.2020	175925	22745	482	6,177944114
58	19.04.2020	178972	23227	433	6,070737728
59	20.04.2020	181228	23660	454	6,118097198
60	21.04.2020	183957	24114	534	6,280395839
61	22.04.2020	187327	24648	437	6,079933195
62	23.04.2020	189973	25085	464	6,139884552
63	24.04.2020	192994	25549	420	6,040254711
64	25.04.2020	195351	25969	415	6,02827852
65	26.04.2020	197675	26384	260	5,560681631
66	27.04.2020	199414	26644	333	5,80814249
67	28.04.2020	201505	26977	382	5,945420609
68	29.04.2020	203591	27359	323	5,777652323
69	30.04.2020	205463	27682	285	5,65248918
70	01.05.2020	207428	27967	269	5,59471138
71	02.05.2020	209328	28236	474	6,161207322
72	03.05.2020	210717	28710	174	5,159055299
73	04.05.2020	211938	28884	195	5,272999559
74	05.05.2020	213013	29079	236	5,463831805
75	06.05.2020	214457	29315	369	5,910796644
79	10.05.2020	219070	30395	165	5,105945474
80	11.05.2020	219814	30560	179	5,187385806
81	12.05.2020	221216	30739	172	5,147494477
82	13.05.2020	222104	30911	195	5,272999559
83	14.05.2020	223096	31106	262	5,568344504
84	15.05.2020	223885	31368	242	5,488937726
85	16.05.2020	224760	31610	153	5,030437921
86	17.05.2020	225435	31763	145	4,976733742
87	18.05.2020	225886	31908	99	4,59511985
88	19.05.2020	226699	32007	162	5,087596335
89	20.05.2020	227364	32169	161	5,081404365
90	21.05.2020	228006	32330	156	5,049856007
91	22.05.2020	228658	32486	130	4,86753445
92	23.05.2020	229327	32616	119	4,779123493
93	24.05.2020	229858	32735	50	3,912023005
94	25.05.2020	230158	32785	92	4,521788577
95	26.05.2020	230555	32877	78	4,356708827
96	27.05.2020	231139	32955	117	4,762173935
97	28.05.2020	231732	33072	70	4,248495242
98	29.05.2020	232248	33142	87	4,465908119
99	30.05.2020	232664	33229	111	4,709530201
100	31.05.2020	232997	33340	75	4,317488114
101	01.06.2020	233197	33415	60	4,094344562
102	02.06.2020	233515	33475	55	4,007333185
103	03.06.2020	233836	33530	71	4,262679877
104	04.06.2020	234013	33601	88	4,477336814
105	05.06.2020	234531	33689	85	4,442651256
106	06.06.2020	234801	33774	72	4,276666119
107	07.06.2020	234998	33846	53	3,970291914
108	08.06.2020	235278	33899	65	4,17438727
109	09.06.2020	235561	33964	79	4,369447852
110	10.06.2020	235763	34043	71	4,262679877
111	11.06.2020	236142	34114	53	3,970291914
112	12.06.2020	236305	34167	56	4,025351691
113	13.06.2020	236651	34223	78	4,356708827
130	30.06.2020	240578	34744	23	3,135494216
131	01.07.2020	240760	34767	21	3,044522438
132	02.07.2020	240961	34788	30	3,401197382
133	03.07.2020	241184	34818	15	2,708050201
134	04.07.2020	241419	34833	21	3,044522438
135	05.07.2020	241611	34854	7	1,945910149
136	06.07.2020	241819	34861	8	2,079441542
137	07.07.2020	241956	34869	30	3,401197382
138	08.07.2020	242149	34899	15	2,708050201
139	09.07.2020	242363	34914	12	2,48490665
140	10.07.2020	242639	34926	12	2,48490665
141	11.07.2020	242827	34938	7	1,945910149
142	12.07.2020	243061	34945	9	2,197224577
143	13.07.2020	243230	34954	13	2,564949357
144	14.07.2020	243344	34967	17	2,833213344
145	15.07.2020	243506	34984	13	2,564949357
146	16.07.2020	243736	34997	20	2,995732274
147	17.07.2020	243967	35017	11	2,397895273
148	18.07.2020	244216	35028	14	2,63905733
149	19.07.2020	244434	35042	3	1,098612289
150	20.07.2020	244624	35045	13	2,564949357
151	21.07.2020	244752	35058	15	2,708050201
152	22.07.2020	245032	35073	9	2,197224577
153	23.07.2020	245338	35082	10	2,302585093
154	24.07.2020	245590	35092	5	1,609437912
155	25.07.2020	245864	35097	5	1,609437912
156	26.07.2020	246118	35102	5	1,609437912
157	27.07.2020	246286	35107	5	1,609437912
158	28.07.2020	246488	35112	11	2,397895273
159	29.07.2020	246776	35123	6	1,791759469
160	30.07.2020	247158	35129	3	1,098612289
161	31.07.2020	247537	35132	9	2,197224577
162	01.08.2020	247832	35141	5	1,609437912
163	02.08.2020	248070	35146	8	2,079441542
164	03.08.2020	248229	35154	12	2,48490665
165	04.08.2020	248419	35166	5	1,609437912

Figure 2.1: Italian Covid'19 Data

The result of the connection of the "Order (22.2.2020 – 4.8.2020)" and "New Death (22.2.2020 – 4.8.2020)" data is a Non linear regression graph which is given below

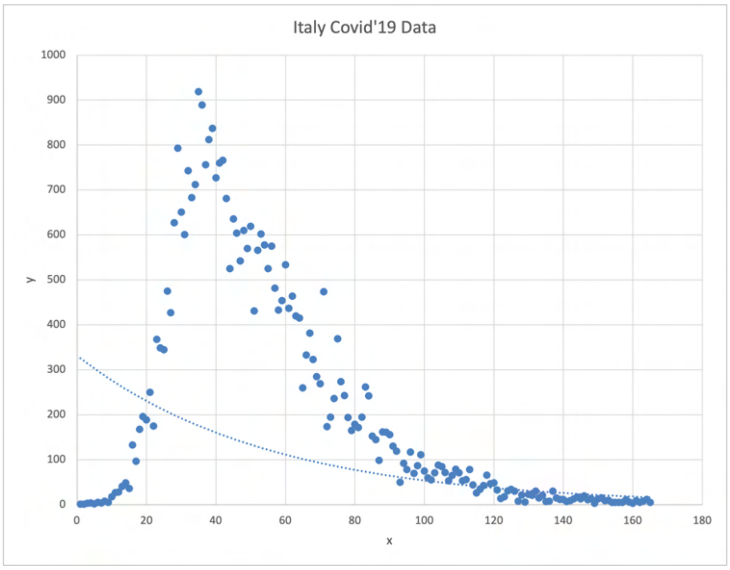


Figure 2.2: Non Fitted Non-Linear Graph

So this graph gave us a lead to talk more about Non liner regression, based on the result of the scattered plot above, we encounter situations where it is not appropriate to use a single expression to describe the dependence between variables. So the data was changed from non-linear to Linear and then both the changed data and the original data was divided into several sections in other to find an expression of dependence for each of them.

2.4. LINEAR REGRESSION FUNCTION USING EXPONENTIAL MODEL

Occasionally linear regression can be utilized with relationships that are not inherently linear, however can be construct to be linear after a transformation. Particularly, we examine the next exponential model:

$$Y = Ae^{\beta x} \tag{2.4.1}$$

Taking the natural log (sight Exponential and Logs) of the two sides of the equation, we have the next equivalent equation:

$$Y = Ae^{\beta x} \tag{2.4.2}$$

Note: $Aimpliese^{\beta}$ This equation has the structure of a linear regression model (where an error term is included).

$$y = \beta + \beta x + error \tag{2.4.3}$$

Now, back to the given data, we transform ”New Death Order (22.2.2020 – 4.8.2020)” data corresponding to this model $\gamma = \beta e$. So that we can use the linear regression form to find the linear relationship between ”Order (22.2.2020 – 4.8.2020)” and ”New Death Order (22.2.2020 – 4.8.2020)”. Taking the natural log of both sides just as we've discussed earlier

$$ln(Y) = ln(A) + ln(e^{\beta}x) \tag{2.4.4}$$

which implies,

$$y = \beta + \beta x \tag{2.4.5}$$

then,

$$y = \beta_0 + \beta x + error \tag{2.4.6}$$

(after introducing the linearization terminologies)

where $ln(Y)$ = y, ln(β) = β_0, x = Order and y = ln(New Death)

order	New Deaths	ln(New Deaths)
1	1	0
2	1	0
3	3	1,098612289
4	4	1,386294361
5	2	0,693147181
6	5	1,609437912
7	4	1,386294361
8	8	2,079441542
9	5	1,609437912
10	18	2,890371758
11	27	3,295836866
12	28	3,33220451
13	41	3,713572067
14	49	3,891820298
15	36	3,583518938
16	133	4,890349128
17	97	4,574710979
18	168	5,123963979
19	196	5,278114659
20	189	5,241747015
21	250	5,521460918
22	175	5,164785974
23	368	5,908082938
24	349	5,855071922
25	345	5,843544417
26	475	6,163314804
27	427	6,056784013
28	627	6,440946541
29	793	6,675823222
30	651	6,478509642
31	601	6,398594935
32	743	6,610696045
33	683	6,52649486
51	431	6,06610809
52	566	6,338594078
53	602	6,400257445
54	578	6,359573869
55	525	6,263398263
56	575	6,354370041
57	482	6,177944114
58	433	6,070737728
59	454	6,118097198
60	534	6,280395839
61	437	6,079933195
62	464	6,139884552
63	420	6,040254711
64	415	6,02827852
65	260	5,560681631
66	333	5,80814249
67	382	5,945420609
68	323	5,777652323
69	285	5,65248918
70	269	5,59471138
71	474	6,161207322
72	174	5,159055299
73	195	5,272999559
74	236	5,463831805
75	369	5,910796644
76	274	5,613128106
77	243	5,493061443
78	194	5,267858159
79	165	5,105945474
80	179	5,187385806
81	172	5,147494477
82	195	5,272999559
83	262	5,568344504
84	242	5,488937726
88	162	5,087596335
89	161	5,081404365
90	156	5,049856007
91	130	4,86753445
92	119	4,779123493
93	50	3,912023005
94	92	4,521788577
95	78	4,356708827
96	117	4,762173935
97	70	4,248495242
98	87	4,465908119
99	111	4,709530201
100	75	4,317488114
101	60	4,094344562
102	55	4,007333185
103	71	4,262679877
104	88	4,477336814
105	85	4,442651256
106	72	4,276666119
107	53	3,970291914
108	65	4,17438727
109	79	4,369447852
110	71	4,262679877
111	53	3,970291914
112	56	4,025351691
113	78	4,356708827
114	44	3,784189634
115	26	3,258096538
116	34	3,526360525
117	43	3,761200116
118	66	4,189654742
119	47	3,850147602
120	49	3,891820298
121	33	3,496507561
132	30	3,401197382
133	15	2,708050201
134	21	3,044522438
135	7	1,945910149
136	8	2,079441542
137	30	3,401197382
138	15	2,708050201
139	12	2,48490665
140	12	2,48490665
141	7	1,945910149
142	9	2,197224577
143	13	2,564949357
144	17	2,833213344
145	13	2,564949357
146	20	2,995732274
147	11	2,397895273
148	14	2,63905733
149	3	1,098612289
150	13	2,564949357
151	15	2,708050201
152	9	2,197224577
153	10	2,302585093
154	5	1,609437912
155	5	1,609437912
156	5	1,609437912
157	5	1,609437912
158	11	2,397895273
159	6	1,791759469
160	3	1,098612289
161	9	2,197224577
162	5	1,609437912
163	8	2,079441542
164	12	2,48490665
165	5	1,609437912

Figure 2.3: Linear Form

Some part of the dataset is shown in the data displayed above, order is represented as x which is the independent variable and ln(New Death) being the dependent natural log of new death. We take the graph which seems to be non-linear as we can peruse in the graph below Assuming that the error in the transformed equation has the desired properties (normal distribution with mean null or 0). When we obtain our estimates from the transformed equation, going back to the original equation can be tricky. Some true-equation parameter evaluation are biased, however consistent, if the parameter was transformed (e.g. A in the models above). Confidence intervals surrounding predicted values are no more symmetrical. It is compulsory for us to get the confidence interval from the transformed equation and then transform the ends back.

Here is the scattered plot below after changing/transforming the original Non-linear equation into linear

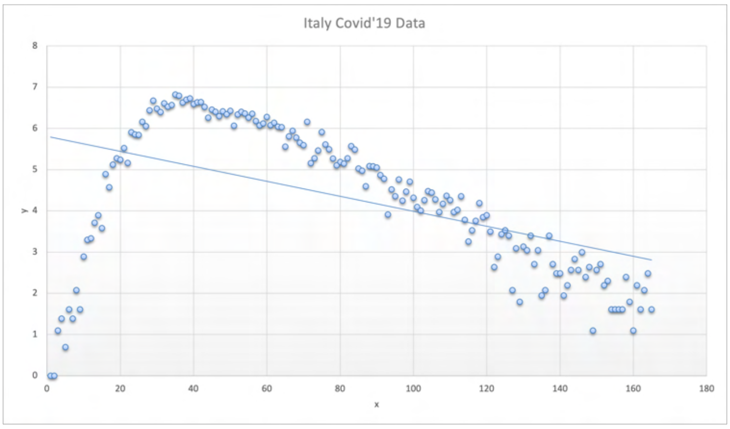


Figure 2.4: Non Fitted Linear Plot

Before we begin to utilize a non-linear regression equation: we should have a better purpose for not utilizing a linear model, like a theory of what way does the process that we are observing works, or a pattern we see on the graph or in the residuals from a linear regression, so, we decide which non-linear equation will be best for our data. Could we construct a reasonable analogy with steady Non-linear? What about along demand or production?

Moreover, the graph Figure 2.4 is not a fitted regression line, so therefore we must find the individual sections that are described by regression function, to find the appropriate regression functions for these sections we need to divide the data (i.e the Order, New deaths and ln(New deaths))into two parts/sections so that we can find the fitted regression line and the confidence interval following each other, then taken an area borders to minimize the model error.

The procedure of least square model regression examine the total of the complete deviation of the observations from the line in the vertical direction in the scatter diagram as in the event of direct regression to get the estimates of β_0 and β_1.

No presumption is needed about the form of the probability distribution of $\mathcal{E}_i$ in obtaining the least squares estimates. For the aim of getting the statistical inferences alone, we presume that $\mathcal{E}_i$ are random variable along $E(\mathcal{E}_i) = 0, \operatorname{Var}(\mathcal{E}_i) = \sigma^2$ and $\operatorname{Cov}(\mathcal{E}_i, \mathcal{E}_j) = 0$ for all $i \neq j (i, j = 1, 2, \ldots, n)$. This assumption is required to look for the mean, variance including more properties of the least-squares estimates. The presumption that $\mathcal{E}_i$'s are usually distributed is utilized while building the tests of hypotheses including confidence intervals of the parameters.

Depending on these approaches, separate estimates of β_0 and β_1 are acquire which include separate statistical properties. Between them, the direct regression approach is more accepted. Commonly, the direct regression estimates are known as the least-squares estimates.

2.4.1. FINDING THE LEAST SQUARE REGRESSION LINE

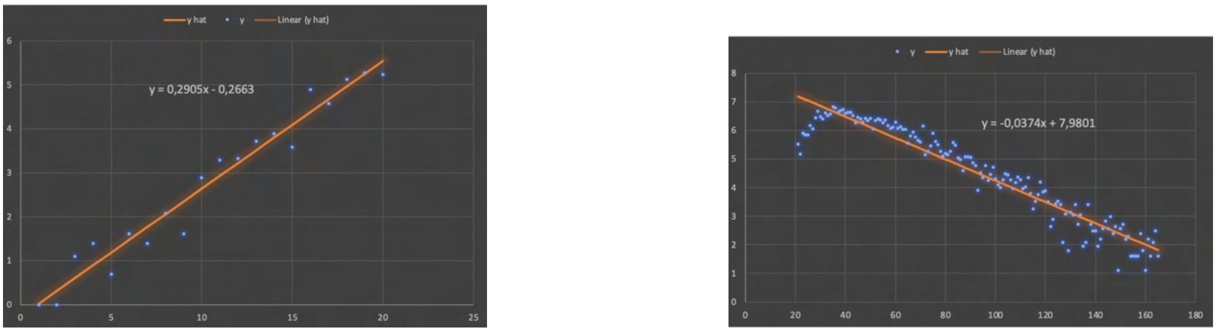


Figure 2.5: Regression line

The scattered diagram and the plot in Figure 2:37 seems to be showing positive relationship between x and y i.e as x is the order so thus y is the ln(New Death), and a fitted straight line.

So, with the linear model, we can describe this relationship between x and y by finding the slope and the y-intercept that defines the line that fits this data perfectly using the sample data. We are going to get the line using the least squares method. what we are going to be doing is find the line that fit the data the best. The $\widehat{y}$ regression line fits the data the best when the distance of each of the data points is at its minimum distance from the line.

We will be using the formula below to minimize the distance of each y_i from each corresponding $\widehat{y}$

$$\begin{aligned} SS(Res) &= \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \\ &= \sum_{i=1}^{n} \mathcal{E}_i^2 \end{aligned} \tag{2.4.7}$$

Recall that we divided our transformed data into two parts. Now, we analyse the first part and the analysis is also applicable to the second part.

Analyses from the first part of our sample data y_i are the observed ln(New Deaths). For example, dividing our data into two at a particular point $X^0 = 20$, the first part is from (1 to 20) while the second part is from (21 to 165). Now, y_i is minimized which formed the regression line and that would be the line that fits the data from the previous formula.

- y_i = observed value for the dependent variable for the i^{th} observation.
- $\widehat{y}$ = predicted value of the dependent variable for the i^{th} observation.

So, from our sample table and diagram

let's take the "order" x_i where $i = 1, ..., 20$. On the graph we look at points in x_i and then look up to the line of regression and over to where all points are on the y axis, them we would get a predicted y value $\widehat{y}$ for the observations. Once we get an equation for the regression line we will be able to predict that value more exactly.

Looking back at the previous study, we have an observation of "order" x_i and that observation have a corresponding observed y value of y_i. In short, what least square method is saying is to define a straight line that minimize the difference or deviations from each of the dots to the line (i.e take each y_i from each $\widehat{y}$ and minimize that squared difference). Now that we understand what the best fitting line to the data would be, so , we need to calculate the slope and the y-intercept.

2.4.2. TO CALCULATE THE SLOPE

We are going to use the following equation to obtain the slope

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.4.8}$$

- x_i = value of independent variable for i^{th} observation (we have 2O observation)
- y_i = value of dependent variable for i^{th} observation
- $\bar{x}$ = mean value of independent variable (i.e we will add up all the x's and divide it by 20)
- $\bar{y}$ = mean value of dependent variable (i.e we will add up all the y's and divide it by 20)

Once we plug in all the numbers and calculate the slope then we can calculate the y-intercept β_0 by using the formula stated above

2.4.3. TO CALCULATE THE INTERCEPT

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}. \tag{2.4.9}$$

Based on this formula, we must calculate the slope before the intercept

Here are the results of the evaluation

- Slope is 0,2905
- Intercept is -0,2664

The general equation is now given as

$$\widehat{\boldsymbol{y}} = -0,2664 + 0,2905x \tag{2.4.10}$$

So, when x = 10 the predicted value is 2,6387.

- How good is this prediction (that turns out how good is the regression line to the data)
- Anyone can draw a straight line through any data points and define it mathematically with a slope and y-intercept but that doesn't mean it's a good fitting model. So we need a measurement that tells us how well the regression line fits the data, that such measurement is called "Coefficient of determination " and it tells us how good a fit regression line is to our data.

2.4.4. COEFFICIENT OF DETERMINATION

How well does the regression line fit the data

$$r^2 = \frac{SSR}{SST} \tag{2.4.11}$$

- r^2 is the coefficient of determination and this is calculated by SSR and SST
- SSR means sum of square due to regression =

$$\sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2 \tag{2.4.12}$$

the way we calculated the SSR is the sum of the squared derivatives of each predicted value of y that's each $\hat{y}$ and subtract $\bar{y}$ which is the average y, so it's between the predicted values and the average the denominator.

- SST means sum of square of the total deviation and we find that value by taking the sum of the squared differences of each yi at each actual observation from $\bar{y}$

$$\sum_{i=1}^{n} (y - \bar{y})^2 \tag{2.4.13}$$

Another means of getting SST is the sum of SSR and SSE

- SSE means square of the error and that is calculated by taking the squared differences of each y_i from each predicted value $\widehat{y}$. This is positioned on the scattered diagram at the deviation of the actual value of y and the predicted value of y and that is called unexplained variation, this is the variation of y that is not explained by the line of regression.

$$\sum_{i=1}^{n} (y - \widehat{y})^2 \tag{2.4.14}$$

So from our sample data

- SSE = 2,4397
- SST = 58,5606

- SSR = 56,1209

Which implies

$$r^2 = \frac{SSR}{SST} = 0,9583 \tag{2.4.15}$$

The coefficients of determination is $0,9554$. r^2 measures the present of variability in y can be explained by the x variable.

Since r^2 is $95,54\%$ of the variability in ln(New Deaths) can be explained by the number of Orders (i.e the $\widehat{y}$ explains 95% of the variation in ln(New Deaths) from the mean but 5% of the variation is unexplained by the line of regression and that is the error.). Another measure of how well our line fits the data need to be discussed and that is the correlation coefficients.

2.4.5. CORRELATION COEFFICIENT

This measure the strength of association between x and y, the correlation coefficient is called r and it's values are between -1 and +1.

- r = +1 means perfect positive linear relationship between x and y, so that means all the data points from the sample lie exactly on the line of regression with no deviation and the data points from the sample lie exactly on the line of regression with no deviation and the line slopes upward.
- r = -1 means perfect negative linear relationship between x and y in this case all the data points line exactly on the line of regression but the line is sloping downward.
- if r = 0, then it means there is no relationship x and y

To calculate r, we simply take the square root of the coefficient of determination and we use the sign of the slope to calculate it.

$$r_{xy} = (\text{sign of } \widehat{\beta}_1)\sqrt{r^2} \tag{2.4.16}$$

We calculated the r has a subscript of x and y, it just tells us that the correlation coefficient is for the values of x and y. From our sample $r^2 = 0,9554$. Taken the square root of the coefficient determination (0,9554). We don't know if it should be positive since the square number always loose their signs. So, in other to know if it's positive or negative number. We have to look at the slope if it's a positive slope or a negative slope and then we use the sign of our slope. In our sample data the slope is $+0,2404$, so we use the positive sign and we get

$$r_{xy} = (\text{sign of } \widehat{\beta}_1)\sqrt{0,9583}$$

$$r_{xy} = +0,9789$$

Recall that +1 would be a perfect linear relationship which is very rare. So, a $+0,9789$ will be a very strong linear relationship between x and y. calculating the $randr^2$ we see that the line is a very good fit to the data.

2.4.6. HYPOTHESIS TEST OF SIGNIFICANCE, T-TEST

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Starting with the null hypothesis H_0 that the slope $(\beta_1) = 0$ and the alternative H_a is to see if we find evidence that the slope is not equal to zero (i.e $\beta_1 \neq 0$). We can conclude that there is a linear relationship between x and y since we do not know the value of sigma for this distribution we will be using a t-test and the test statistics would be:

$$t = \frac{\beta_1}{S_{\beta_1}} \tag{2.4.17}$$

S_{β_1} is the standard error of the slope

$$S_{\beta_1} = \frac{S}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{2.4.18}$$

and

$$S = \sqrt{\frac{SSE}{n-2}}$$

S is the standard deviation

From the previous calculation, SSE $= 2,4397$, so to get S, We have

$$S = \sqrt{\frac{2,4397}{20-2}} = 0,3682$$

Now, to find S_{β_1} (the standard error for the slope) and that is

$$S_{\beta_1} = \frac{0,3682}{\sqrt{665}} = 0,01428$$

Now, we can finally calculate our test statistics as follows:
$\beta_1 = 0,2905$,

$$t = \frac{0,2905}{0,01428} = 20,3431$$

We tested to see if we have enough evidence to support the alternative hypothesis that the slope is not equal to zero, if we find this evidence we will conclude that here is a linear relationship between x and y. We calculated our t test to be $20,3431$.

Now, we are ready to use either the critical value approach or p-value approach to solve this problem. We will begin with the critical value approach and let's use the alpha value $(\alpha) = 0.01$, as seen this in two-tailed test we split alpha in half (i.e $(\alpha)/2 = 0 : 005$). Since this is a t-test we looked up our critical value in t table under $n - 2$ degrees of freedom (i.e we have 18 degrees of freedom based on our sample data). So, with this we find the critical value of $2,1009$.

From the t-distribution the critical value splits the distribution into rejection regions and non-rejection region and the statistic falls around $20,3431$

Now, we are ready to come to a statistical conclusion and this of course will be to reject the null.

2.4.7. STATISTICAL CONCLUSION

There is evidence that the slope is not equal to zero which there is a significant relationship between ln(New Death) y and number of Orders(x). We can also solve this problem using the p-value approach, to use the p-value approach.

2.4.8. PROBABILITY VALUE APPROACH

Using the t-statistic $(20,3431)$, looking up to this number in t-table under $df = 18$ using excel. For a two-tailed test we double the value (i.e for a two-tailed test: Double the area and compare to α).

So the p-value is $8,95835E - 20$

2.4.9. REJECTION RULE

Rejection rule is to reject the null hypothesis if the p-value is less that or equal to α. Since our $\alpha = 0.1$ which is the value for this problem, then our p-value $(8,95835E-20)$ is less than our α value.(i.e $8,95835E - 20 \leq 0.1$). Therefore, we reject the null hypothesis and find evidence that the slope is not equal to 0 which means that ln(New Death) and Order have a linear relationship.

Note: When the p-value is less than the α value, then we have a linear relationship. Reject H_0, there is evidence that β_1 is not equal to zero and that a sigńificant relationship exists between ln(New Death) and Order.

2.4.10. CONFIDENCE INTERVAL

Remember that $\widehat{y}$ is a point estimate. Since we want more realistic estimate value, we would take $\widehat{y} \pm$ the margin of error. A confidence interval would be a more realistic way of expressing the ln(New Death).

So, we obtain the result for the confidence interval for mean and individual value using the following equation:

- Mean value formula

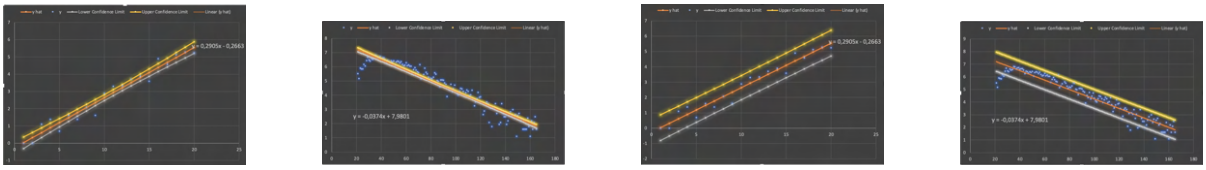
$$\left\langle v_0 - t_1 - \alpha_2 s\sqrt{v^*}; y_0 + t_{1-\alpha} s^2 \sqrt{v}^* \right\rangle$$

- Predicted value formula

$$\left\langle y^0 - t_{1-\alpha/2} s\sqrt{v^* + 1}; y^0 + t_{1-\alpha/2} s\sqrt{v^* + 1} \right\rangle$$

So based on this results, with 95% confidence that for every individual of each Order (x) there predicted ($\widehat{y}$) is be between upper and the lower confidence limit displayed above.

While the confidence interval for the slope is $0,2605$ for the upper limit and $0,3205$ for the lower limit at 95% level of confidence.



(a) a (b) b (c) c (d) d

Figure 2.6: Linear Confidence Interval for Mean and Predicted Value

2.5. FINDING THE MINIMAL

From the sample data (Order (22.2.2020 – 4.8.2020), ln (New Deaths (22.2.2020 – 4.8.2020))) we want to find the minimal point, The lower the SSE, the similar the result.

- we select point $\boldsymbol{x}$ from set $\{3, 4, \ldots, 163\}$, then
- we calculate the regression lines for area $1, .., x$ and for area $x, \ldots, 165$ and their residual sums

$$S_E = \sum_{i=1}^{n} y_i^2 - \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i \tag{2.5.1}$$

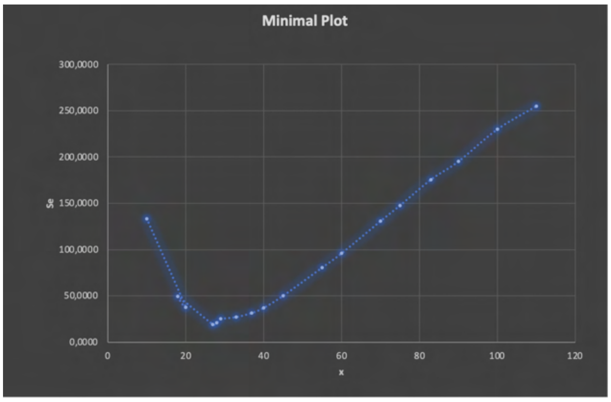
.

- So, we denote: S_E^A - the residual sum for the area $1, .., x$ and S_E^B for the area $x, \ldots, 165$.

So, we can see the results in the figure below

x	Se
10	133,3196
20	37,8550
18	49,4444
27	19,1714
28	20,7336
29	24,9321
33	27,1433
37	31,5923
40	37,0382
45	50,1604
55	80,5164
60	95,9290
70	130,7016
75	147,5689
83	175,7109
90	195,0916
100	229,8477
110	254,7596

(a) Minimal Table



(b) Minimal Plot

Figure 2.7: Some Minimal Points

3. ANALYTICAL SOLUTION OF THE MODEL

3.1. FINDING THE ANALYTIC SOLUTION

3.1.1. BACKGROUND

To analyse the data of the COVID-19 deaths in Italy in detail, the regression analysis is used. The purpose of the regression analysis is to understand the relation of the dependent variable on the independent variable. In our case the dependent variable is the amount of deaths with respect to independent variable time, represented in the form of dates (22.2.2020 – 4.8.2020). The data is converted from non-linear to linear form by taking the natural log of the number of deaths. The data seems to be divided in two parts or areas where the first area sees an increasing trend and then in the second area, there is a decrease in the number of deaths with respect to the independent variable i.e., time. So, there must be two separate expressions to show the dependence of variable on the other and there must be a point where this dependence is changing which can be denoted by x_0.

3.1.2. ANALYTIC SOLUTION

The regression analysis is used to find the expressions of the two separate areas under the condition that the two lines would meet at the point where dependence of the variables is changing. Let's assume the independent variable i.e. the order (22.2.2020 – 4.8.2020) is denoted by x and the dependent variable i.e. the natural log of the number of deaths is denoted by y. If the first area is represented by 1, the independent variable values will be,

$$x_1^1, x_2^1, ..., x_n^1$$

The dependent variable values will be,

$$y_1^1, y_2^1, ..., y_n^1$$

If the second area is represented by 2, the independent variable values will be,

$$x_1^2, x_2^2, ..., x_n^2$$

The dependent variable for second area will be,

$$y_1^2, y_2^2, ..., y_n^2$$

The linear regression model for first area is given by,

$$y = \beta_0 + \beta_1 x_i^1 + \mathcal{E} \tag{3.1.1}$$

The linear regression model for second area is given by,

$$y = \gamma_0 + \gamma_1 x_i^2 + \mathcal{E} \tag{3.1.2}$$

In our case, the two regression lines are not meeting each other. To make the lines meet we need to optimize the values of slopes and y-intercepts of individual regression functions using Lagrange multiplier for squared errors of the combined regression lines. A Lagrange multiplier will be used to minimize the squared error of the regression lines subject to the constraint that both lines would meet at an arbitrary point denoted by x^o. The generalized form of the Lagrange function is given by,

$$L(x, \lambda) = f(x) + \lambda g(x) \tag{3.1.3}$$

In our case the function we want to minimize is the squared error of the two regression lines which is given below,

$$f(x) = \sum_{i=1}^{n1} \left(y_i^1 - \left(\beta_0 + \beta_1 x_i^1\right)\right)^2 + \sum_{j=1}^{n2} \left(y_j^2 - \left(\gamma_0 + \gamma_1 x_j^2\right)\right)^2 \tag{3.1.4}$$

The condition under which this function needs to be optimized is given by,

$$g(x) = \beta_0 + \beta_1 x^0 - \gamma_0 - \gamma_1 x^0$$

So the Lagrange function of the squared error of the regression lines is given by,

$$L(\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda) = \sum_{i=1}^{n1} \left(y_i^1 - \left(\beta_0 + \beta_1 x_i^1\right)\right)^2 + \sum_{j=1}^{n2} \left(y_j^2 - \left(\gamma_0 + \gamma_1 x_j^2\right)\right)^2 + \lambda\left(\beta_0 + \beta_1 x^0 - \gamma_0 - \gamma_1 x^0\right)$$

The Lagrange multiplier estimates of $\beta_0, \beta_1, \gamma_0, \gamma_1$ and λ can be obtained by minimizing $L(\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda)$

The normal equations are obtained by partial differentiation of Lagrange multiplier with respect to $\beta_0, \beta_1, \gamma_0, \gamma_1$ and λ and equating them to zero as follows to obtain

$$\frac{\partial L}{\partial \beta_0} = 2\left(\sum_{i=1}^{n_1} \left(y_i^1 - \beta_0 - \beta_1 x_i^1\right)\right)(-1) + \lambda = -2\sum_{i=1}^{n_1} y_i^1 + 2n_1\beta_0 + \left(2\sum_{i=1}^{n_1} x_i^1\right)\beta_1 + \lambda = 0$$

Denote

$$b_1 := 2\sum_{i=1}^{n_1} y_i^1, \quad a_{11} = 2n_1, \quad a_{12} = 2\sum_{i=1}^{n_1} x_i^1 or 2n_1\bar{x}^1, a_{13} = 0, a_{14} = 0, a_{15} = 1$$

Then, we have

$$a_{11}\beta_0 + a_{12}\beta_1 + a_{13}\gamma_0 + a_{14}\gamma_1 + a_{15}\lambda = b_1$$

Similarly, we have

$$\frac{\partial L}{\partial \beta_1} = 2\left(\sum_{i=1}^{n_1}\left(y_i^1 - \beta_0 - \beta_1 x_i^1\right)\right)\left(-x_i^1\right) + \lambda x^0$$

$$= -2\sum_{i=1}^{n_1} y_i^1 x_i^1 + 2\left(\sum_{i=1}^{n_1} x_i^1\right)\beta_0 + \left(2\sum_{i=1}^{n_1}\left(x_i^1\right)^2\right)\beta_1 + \lambda x^0 = 0$$

Denote

$$b_2 := 2\sum_{i=1}^{n_1} y_i^1 x_i^1, a_{21} = 2\left(\sum_{i=1}^{n_1} x_i^1\right), a_{22} = \left(2\sum_{i=1}^{n_1}\left(x_i^1\right)^2\right), a_{23} = 0, a_{24} = 0, a_{25} = x^0.$$

Then, we have

$$a_{21}\beta_0 + a_{22}\beta_1 + a_{23}\gamma_0 + a_{24}\gamma_1 + a_{25}\lambda = b_2.$$

Similarly, we have

$$\frac{\partial L}{\partial \gamma_1} = 2\left(\sum_{j=1}^{n_2}\left(y_j^2 - \gamma_1 - \gamma_2 x_j^2\right)\right)(-1) - \lambda = -2\sum_{j=1}^{n_2} y_j^2 + 2n_2\gamma_1 + \left(2\sum_{j=1}^{n_2} x_j^2\right)\gamma_2 - \lambda = 0$$

Denote

$$b_3 := 2\sum_{j=1}^{n_2} y_j^2, \quad a_{33} = 2n_2, \quad a_{34} = \left(2\sum_{j=1}^{n_2} x_j^2\right), a_{31} = 0, a_{32} = 0, a_{35} = -1$$

Then, we have

$$a_{31}\beta_0 + a_{32}\beta_1 + a_{33}\gamma_0 + a_{34}\gamma_1 + a_{35}\lambda = b_3.$$

Similary, we obtain

$$\frac{\partial L}{\partial \gamma_1} = 2\left(\sum_{j=1}^{n_2}\left(y_j^2 - \gamma_0 - \gamma_1 x_j^2\right)\right)\left(-x_j^2\right) - \lambda x^0$$

$$= -2\sum_{j=1}^{n_2} y_j^2 x_j^2 + 2\left(\sum_{j=1}^{n_2} x_j^2\right)\gamma_0 + \left(2\sum_{j=1}^{n_2}\left(x_j^2\right)^2\right)\gamma_1 - \lambda x^0 = 0$$

Denote

$$b_4 := 2\sum_{j=1}^{n_2} y_j^2 x_j^2, \quad a_{43} = 2\left(\sum_{j=1}^{n_2} x_j^2\right), \quad a_{44} = \left(2\sum_{j=1}^{n_2}\left(x_j^2\right)^2\right), a_{41} = 0, a_{42} = 0, a_{45} = -x^0.$$

Then, we have

$$a_{41}\beta_0 + a_{42}\beta_1 + a_{43}\gamma_0 + a_{44}\gamma_1 + a_{45}\lambda = b_4.$$

Finally, we have

$$\frac{\partial L}{\partial \lambda} = \beta_0 + \beta_1 x^0 - \gamma_0 - \gamma_1 x^0 \tag{3.1.5}$$

Denote

$$a_{51} = 1, a_{52} = x^0, a_{53} = -1, a_{54} = -x^0, a_{55} = 0, b_5 = 0.$$

Then, we have

$$a_{51}\beta_0 + a_{52}\beta_1 + a_{53}\gamma_0 + a_{54}\gamma_1 + a_{55}\lambda = b_5.$$

Now, we have

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}, x = \begin{bmatrix} \beta_0, \\ \beta_1 \\ \gamma_0, \\ \gamma_1, \\ \lambda \end{bmatrix}, b = \begin{bmatrix} b_1, \\ b_2, \\ b_3, \\ b_4, \\ b_5, \end{bmatrix}$$

Which implies

$$A = \begin{bmatrix} 2n_1 & 2n_1\bar{x}^1 & 0 & 0 & 1 \\ 2n_1\bar{x}^1 & 2\sum_{i=1}^{n1}\left(x_i^1\right)^2 & 0 & 0 & x^0 \\ 0 & 0 & 2n_2 & 2n_2\bar{x}^2 & -1 \\ 0 & 0 & 2n_2\bar{x}^2 & 2\sum_{j=1}^{n_2}\left(x_j^2\right)^2 & -x^0 \\ 1 & x^0 & -1 & -x^0 & 0 \end{bmatrix}, x = \begin{bmatrix} \beta_0, \\ \beta_1 \\ \gamma_0, \\ \gamma_1, \\ \lambda \end{bmatrix}, b = \begin{bmatrix} b_1, \\ b_2, \\ b_3, \\ b_4, \\ b_5, \end{bmatrix}$$

Then, we can solve for $\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda$ by solving the following linear system of equations

$$Ax = b \tag{3.1.6}$$

The above linear system of equations has a solution if an only if $AA^T b = b$. Let's assume that $AA^T b = b$ holds, then let x^* be such that $Ax^* = b$. The vector x^* is given by

$$x^* = A^T b + \left(I - A^T A\right) y \tag{3.1.7}$$

where y is any arbitrary vector in $\mathbb{R}^5$.

We further denote

$$x^* = \left(\beta_0^*, \beta_1^*, \gamma_0^*, \gamma_1^*, \lambda^*\right). \tag{3.1.8}$$

From the above calculations it is easy to deduce that

$$\nabla L\left(\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda\right) = Ax - b \tag{3.1.9}$$

With $\lambda = \lambda^*$, the Hessian is given by

$$\nabla^2 L\left(\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda^*\right) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \tilde{A}$$

Note that $\tilde{A}$ is constant, as all the entries are constant. So, based on Sylvester Criterion application; If $\tilde{A}$ is positive (i.e greater than 0) semi-definite then we can conclude that $(\beta_0^*, \beta_1^*, \gamma_0^*, \gamma_1^*)$ is the minimizer of the primal problem, which is

$$\min_{\beta_0+\beta_1 x^0=\gamma_0+\gamma_1 x^0} \sum_{i=1}^{n_1} \left(y_i^1 - \beta_0 - \beta_1 x_i^1\right)^2 + \sum_{j=1}^{n_2} \left(y_j^2 - \gamma_0 - \gamma_1 x_j^2\right)^2. \tag{3.1.10}$$

But if it's negative (i.e less than 0) then it's a saddle point, indefinite and strict local maximizer.

A minima or maxima value of zero is necessary for all partial derivatives. if Gradient is zero at a minima or maxima.The function always increases as we move away from the minimum which makes the Hessian matrix to be a positive definite.The function decreases as we move away from the maxima which makes the Hessian matrix to be a negative definite. In the situation where the Hessian has neither positive nor negative definite points, then the point is neither a minima nor a maxima. but It's more like a saddle (moving in some directions increases the function, while moving in others reduces it).During our discussion about the Sylvester Criterion, we will elaborate on this further [10].

3.1.3. APPLICATION OF SYLVESTER CRITERIA

Sylvester criteria is an important method to find the local extrema of a function. The criteria is applied to the hessian matrix created from the function L. The matrix is given by,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}$$

Let $\Delta_k = \det\left(A^{(k)}\right)$. (So $\Delta_n = \det(A)$.) By examining A's eigenvalues, we should be able to determine its determinant [20].

Since

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x),$$

As a result of setting $x = 0$ then $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$. When $A \succ 0$, Each eigenvalue is positive, so $\det(A) > 0$ Likewise.

According to the Sylvester Criteria, the function to have a local minima then all of its principal minors of its hessian matrix have to be positive.

According to Sylvester's criterion, such matrices are actually positive definite: From some of Sylvester's criterion theorems: suppose A is an $n \times n$ symmetric matrix [20]. Then:

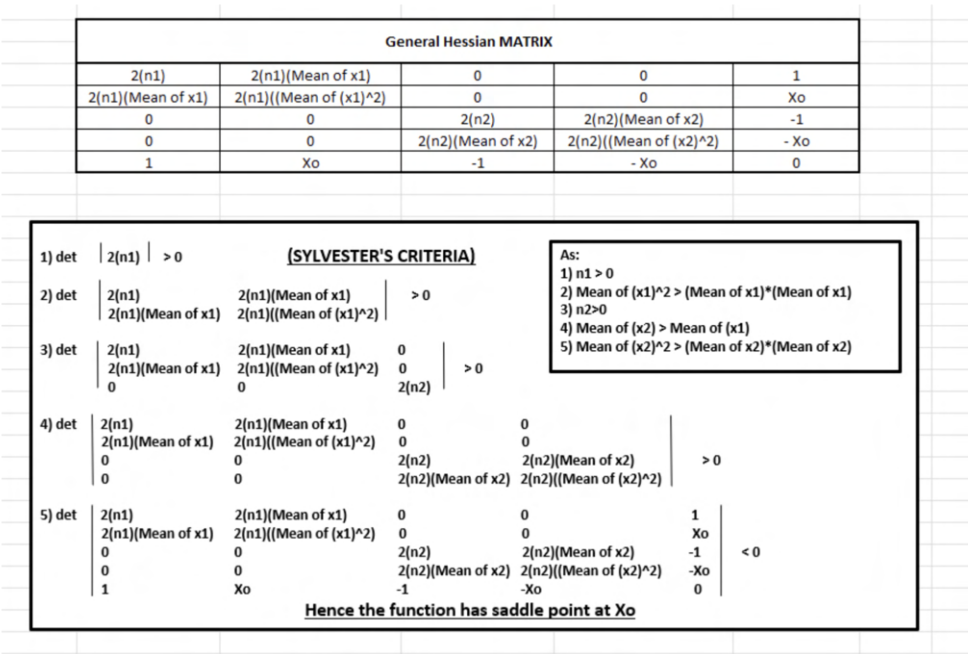
- $A \succ 0$ if and only if $\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_n > 0$ [20].
- $A \prec 0$ if and only if $(-1)^1\Delta_1 > 0, (-1)^2\Delta_2 > 0, \ldots, (-1)^n\Delta_n > 0$ [20].
- A is indefinite if the first Δ_k that breaks both patterns is the wrong sign [20].
- A can be either negative semidefinite or positive, or indefinite, so we can say that Sylvester's criterion is not conclusive when the result of the Δ_k is 0 [20].

Another Sylvester's criterion theorem also state that; if $f : \mathbb{R}^n \to \mathbb{R}$ is a function with continuous Hf, and suppose $\mathrm{x}^* \in \mathbb{R}^n$ is a critical point of f [20].

- Assuming $Hf\,(\mathrm{x}^*) \succ 0$, in that case x^* is a strict local minimizer [20].
- Assuming $Hf\,(\mathrm{x}^*) \prec 0$, in that case x^* is a strict local maximizer [20].
- The result from the previous theorem is further enhanced by this result: suppose $\mathrm{Hf}\,(\mathrm{x}^*)$ is indefinite, in that case x^* is a saddle point [20].

Applying the criteria to the above matrix of our data of COVID-19 deaths,

As,



General Hessian MATRIX

2(n1)	2(n1)(Mean of x1)	0	0	1
2(n1)(Mean of x1)	2(n1)((Mean of (x1)^2)	0	0	Xo
0	0	2(n2)	2(n2)(Mean of x2)	-1
0	0	2(n2)(Mean of x2)	2(n2)((Mean of (x2)^2)	- Xo
1	Xo	-1	- Xo	0

Figure 3.1: Sylvester Criterion

$$n_1 > 0$$

So,

$$det[\ a_{11}\ ] > 0$$

As

$$n_1 > 0$$

$$\left(\sum_{i=1}^{n_1} \left(x_i^1\right)^2\right) > 0$$

$$\left(\sum_{i=1}^{n_1} x_i^1\right) \cdot \left(\sum_{i=1}^{n_1} x_i^1\right) > 0$$

So,

$$det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} > 0$$

As

$$n_1 > 0$$

$$\left(\sum_{i=1}^{n_1} \left(x_i^1\right)^2\right) > 0$$

$$\left(\sum_{i=1}^{n_1} x_i^1\right) \cdot \left(\sum_{i=1}^{n_1} x_i^1\right) > 0$$

$$n_2 > 0$$

So,

$$det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} > 0$$

As

$$n_1 > 0$$

$$\left(\sum_{i=1}^{n_1} \left(x_i^1\right)^2\right) > 0$$

$$\left(\sum_{i=1}^{n_1} x_i^1\right) \cdot \left(\sum_{i=1}^{n_1} x_i^1\right)$$

$$n_2 > 0$$

$$\left(\sum_{i=1}^{n_2} x_i^2\right) > \left(\sum_{i=1}^{n_1} x_i^1\right) > 0$$

$$\left(\sum_{i=1}^{n_2} \left(x_i^2\right)^2\right) > \left(\sum_{i=1}^{n_2} x_i^2\right) \cdot \left(\sum_{i=1}^{n_2} x_i^2\right)$$

So,

$$det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} > 0$$

As

$$n_1 > 0$$

$$\left(\sum_{i=1}^{n_1} (x_i^1)^2\right) > 0$$

$$\left(\sum_{i=1}^{n_1} x_i^1\right) \cdot \left(\sum_{i=1}^{n_1} x_i^1\right)$$

$$n_2 > 0$$

$$\left(\sum_{i=1}^{n_2} x_i^2\right) > \left(\sum_{i=1}^{n_1} x_i^1\right) > 0$$

$$\left(\sum_{i=1}^{n_2} (x_i^2)^2\right) > \left(\sum_{i=1}^{n_2} x_i^2\right) \cdot \left(\sum_{i=1}^{n_2} x_i^2\right)$$

So,

$$det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} < 0$$

For $x^0 > 0$, this determinant is negative. Hence it is proved that the function L has the saddle point at x_0. It means that the function L has local minimum in one direction and has a local maximum in other direction.By looking at the data, it is clear that is the case.

3.1.4. CHANGE POINT APPLICATION ON THE COVID-19 DATA

The data of COVID-19 deaths in Italy over a period of time shows the initial trend of increase till a point after that the decreasing trend can be observed. The point this change of behaviour occurs is the change point or in mathematical terms it is the saddle point. The significance of that point becomes clear when the squared error of the two regression lines needs to be minimized and the condition under which it can be minimized that the two lines would meet at an arbitrary point i.e. x^o where the two regression functions become equal. The condition is defined as,

$$\beta_0 + \beta_1 x^o = \gamma_0 + \gamma_1 x^o$$

The Lagrange multiplier function is used to minimize the squared error under the constraint that the two functions would become equal at the change point x^o. The Lagrange function helps optimize the values of the required regression parameters under a defined condition by introducing the Lagrange variable i.e.λ whose value varies with the variation in the value of the change point i.e. x^o.

3.1.5. MODIFICATION OF SLOPES AND y-INTERCEPTS

As the function L has the saddle point at x^0, so for every selected value of x_0, there will be new values of slopes and y-intercepts of the two regression lines under the condition that the two lines meet each other at x^0. By solving the system of linear equations for different values of x^0, the different values of $\beta_0, \beta_1, \gamma_0, \gamma_1, \lambda$ will be obtained. The system of linear equations is given by,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}, x = \begin{bmatrix} \beta_0, \\ \beta_1 \\ \gamma_0, \\ \gamma_1, \\ \lambda \end{bmatrix}, b = \begin{bmatrix} b_1, \\ b_2, \\ b_3, \\ b_4, \\ b_5, \end{bmatrix}$$

Solving for x is given by,

$$x = A^{-1}b \tag{3.1.11}$$

where,

$$b_1 := 2\sum_{i=1}^{n_1} y_i^1, \quad a_{11} = 2n_1, \quad a_{12} = 2\sum_{i=1}^{n_1} x_i^1, a_{13} = 0, a_{14} = 0, a_{15} = 1$$

$$b_2 := 2\sum_{i=1}^{n_1} y_i^1 x_i^1, a_{21} = 2\left(\sum_{i=1}^{n_1} x_i^1\right), a_{22} = \left(2\sum_{i=1}^{n_1} \left(x_i^1\right)^2\right), a_{23} = 0, a_{24} = 0, a_{25} = x^0.$$

$$b_3 := 2\sum_{j=1}^{n_2} y_j^2, \quad a_{33} = 2n_2, \quad a_{34} = \left(2\sum_{j=1}^{n_2} x_j^2\right), a_{31} = 0, a_{32} = 0, a_{35} = -1$$

$$b_4 := 2\sum_{j=1}^{n_2} y_j^2 x_j^2, \quad a_{43} = 2\left(\sum_{j=1}^{n_2} x_j^2\right), \quad a_{44} = \left(2\sum_{j=1}^{n_2} \left(x_j^2\right)^2\right), a_{41} = 0, a_{42} = 0, a_{45} = -x^0.$$

$$a_{51} = 1, a_{52} = x^0, a_{53} = -1, a_{54} = -x^0, a_{55} = 0, b_5 = 0.$$

By solving this system of equations for a specific value of x^0, the two transformed regression lines will be obtained which are optimized under the condition that the two regression functions become equal at x^0.

3.1.6. CONFIDENCE INTERVAL FOR THE MODIFIED REGRESSION LINES

The confidence interval will be re-defined for the two transformed regression functions.The confidence interval for the mean of the two regression functions $\phi(x)$ is expressed as follows,

$$\varphi(x) = \begin{cases} b_0 + \beta_1 x & x < x_0 \\ \gamma_0 + \gamma_1 x & x > x_0 \end{cases} \tag{3.1.12}$$

- The mean value is given by:

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^*}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^*} \right\rangle \tag{3.1.13}$$

while

- The Predicted value is given by

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^*}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^*} \right\rangle \tag{3.1.14}$$

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^* + 1}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^* + 1} \right\rangle \tag{3.1.15}$$

where,

$$h^* = \begin{bmatrix} 1, x, 1, x \end{bmatrix} \begin{bmatrix} X^T X \end{bmatrix} \begin{bmatrix} 1 \\ x \\ 1 \\ x \end{bmatrix}$$

For variable estimation, the formula will be modified as,

$$S^*_{min} = \sum_{i=1}^{n} (y - \phi(x_i))^2 \tag{3.1.16}$$

where,

$$s^2 = S^*_{min}/n - m \tag{3.1.17}$$

where,
m = number of estimated parameters

4. NUMERICAL SOLUTION OF THE MODEL

4.1. BACKGROUND

This chapter deals with the application of the analytical solution to our data and finding the numerical solution of the problem. This chapter will apply the solution to the data of COVID-19 deaths in Italy.In Chapter 2 the data was transformed from non-linear to linear form by the use of natural logarithm function. The data seems to be divided into two parts for which separate regression functions and confidence intervals were found. The two separate regression lines were not meeting each other even by minimizing the squared errors of the two regression functions.In order to meet these two regression lines, modification of the regression analysis was needed. It was achieved by introducing the Lagrange multiplier under the constraint that the two regression functions become equal at an arbitrary point which was later proved to be the saddle point of the Lagrange multiplier function. In Chapter 3, the analytical solution is derived for optimization of the regression functions of the two separate lines. The Lagrange multiplier function was used to minimize the squared error of the two lines under the constraint that the two lines will meet at a saddle point i.e.x^0. The system of linear equations was found and by solving it, the new modified values of slopes and y-intercepts can be calculated. In this chapter these values of slopes and y-intercepts of each area will be calculated for the data under consideration i.e. COVID-19 deaths in Italy over a period of time.

4.2. NUMERICAL SOLUTION

The data of COVID-19 deaths in Italy is given in Figure 2.1. The transformed data in the natural logarithm is given in Figure 2.3. Figure 2.2 and Figure 2.4 shows the non-linear data and linear data respectively. To optimize these two regression lines under the constraint that the two would meet at an arbitrary user selected point, the equation 3.1.3 will be solved for the data of COVID-19 deaths to find the modified values of slopes and y-intercepts of the regression lines.

4.2.1. OPTIMIZED SOLUTION OF THE PROBLEM

In this section we optimized different points of x^0, some of the points we optimized are x^o=20 x^o=29 x^o=40, and x^o= 100 ,

Taken the expression on point x^o=29 as follows

Using Equation 2.4.8 and 2.4.9,

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}. \qquad (4.2.1)$$

- x_i = value of independent variable for i^{th} observation (we have 29 observation)
- y_i = value of dependent variable for i^{th} observation
- $\bar{x}$ = mean value of independent variable (i.e we will add up all the x's and divide it by 29)
- $\bar{y}$ = mean value of dependent variable (i.e we will add up all the y's and divide it by 29)

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4.2.2}$$

The values of the slopes and y-intercepts of first regression line are calculated,

		n1* (because of row gap)
Enter n1 =	29	31
mean of x =	15	

n2 =	136
mean of x =	97,5

SUMS		MEANS
∑ X(1)	435	15,0
∑ Y(1)	109,3087	3,769265190
∑ XY(1)	2127,6101	73,36586489
∑ X^2(1)	8555	295,0
β1	0,1635	
β2	0,2404	

SUMS		MEANS
∑ X(2)	13260	97,5
∑ Y(2)	599,6330	4,4091
∑ XY(2)	49911,716	366,997912
∑ X2(2)	1502460	11047,5
γ1	8,38726001	
γ2	-0,04080199	

Figure 4.1: Sum,mean, slope and intercept computation

$$\beta_0 = 0,1635$$

$$\beta_1 = 0,2404$$

Here is the regression function,

$$\widehat{\boldsymbol{y}} = 0,1635 + 0,2404x \tag{4.2.3}$$

The values of the slope and y-intercept of second regression line are calculated as,

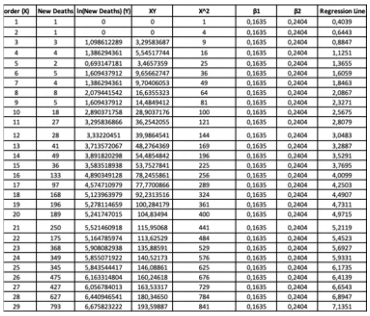
$$\gamma_0 = 8,3873$$

$$\gamma_1 = -0,0408$$

Here is the regression function for the second line,

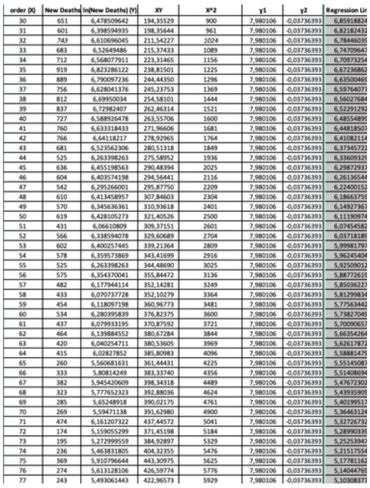
$$\widehat{\boldsymbol{y}} = 8,3873 + (-0,040801986)x \tag{4.2.4}$$

The result of this is shown in the table below



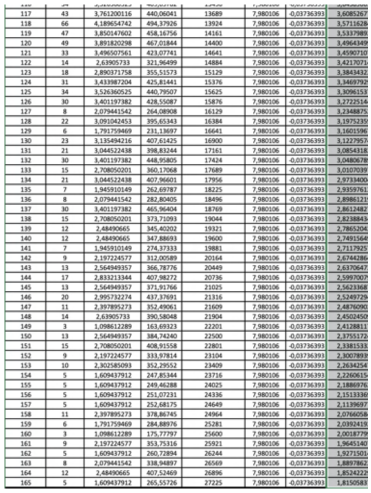
order (X)	New Deaths	ln(New Deaths) (Y)	XY	X^2	β1	β2	Regression Line
1	1	0	0	1	0,1635	0,2404	0,4039
2	1	0	0	4	0,1635	0,2404	0,6443
3	3	1,098612289	3,29583687	9	0,1635	0,2404	0,8847
4	4	1,386294361	5,54517744	16	0,1635	0,2404	1,1251
5	2	0,693147181	3,4657359	25	0,1635	0,2404	1,3655
6	5	1,609437912	9,65662747	36	0,1635	0,2404	1,6059
7	4	1,386294361	9,70406053	49	0,1635	0,2404	1,8463
8	8	2,079441542	16,6355323	64	0,1635	0,2404	2,0867
9	5	1,609437912	14,4849412	81	0,1635	0,2404	2,3271
10	18	2,890371758	28,9037176	100	0,1635	0,2404	2,5675
11	27	3,295836866	36,2542055	121	0,1635	0,2404	2,8079
12	28	3,33220451	39,9864541	144	0,1635	0,2404	3,0483
13	41	3,713572067	48,2764369	169	0,1635	0,2404	3,2887
14	49	3,891820298	54,4854842	196	0,1635	0,2404	3,5291
15	36	3,583518938	53,7527841	225	0,1635	0,2404	3,7695
16	133	4,890349128	78,2455861	256	0,1635	0,2404	4,0099
17	97	4,574710979	77,7700866	289	0,1635	0,2404	4,2503
18	168	5,123963979	92,2313516	324	0,1635	0,2404	4,4907
19	196	5,278114659	100,284179	361	0,1635	0,2404	4,7311
20	189	5,241747015	104,83494	400	0,1635	0,2404	4,9715
21	250	5,521460918	115,95068	441	0,1635	0,2404	5,2119
22	175	5,164785974	113,62529	484	0,1635	0,2404	5,4523
23	368	5,908082938	135,88591	529	0,1635	0,2404	5,6927
24	349	5,855071922	140,52173	576	0,1635	0,2404	5,9331
25	345	5,843544417	146,08861	625	0,1635	0,2404	6,1735
26	475	6,163314804	160,24618	676	0,1635	0,2404	6,4139
27	427	6,056784013	163,53317	729	0,1635	0,2404	6,6543
28	627	6,440946541	180,34650	784	0,1635	0,2404	6,8947
29	793	6,675823222	193,59887	841	0,1635	0,2404	7,1351

(a) a



order (X)	New Deaths	ln(New Deaths) (Y)	XY	X^2	y1	y2	Regression Lin
30	651	6,478509642	194,35529	900	7,980106	-0,03736393	6,85918824
31	601	6,398594935	198,35644	961	7,980106	-0,03736393	6,82182432
32	743	6,610696045	211,54227	1024	7,980106	-0,03736393	6,78446039
33	683	6,52649486	215,37433	1089	7,980106	-0,03736393	6,74709647
34	712	6,568077911	223,31465	1156	7,980106	-0,03736393	6,70973254
35	919	6,823286122	238,81501	1225	7,980106	-0,03736393	6,67236862
36	889	6,790097236	244,44350	1296	7,980106	-0,03736393	6,63500469
37	756	6,628041376	245,23753	1369	7,980106	-0,03736393	6,59764077
38	812	6,69950034	254,58101	1444	7,980106	-0,03736393	6,56027684
39	837	6,72982407	262,46314	1521	7,980106	-0,03736393	6,52291292
40	727	6,588926478	263,55706	1600	7,980106	-0,03736393	6,48554899
41	760	6,633318433	271,96606	1681	7,980106	-0,03736393	6,44818507
42	766	6,64118217	278,92965	1764	7,980106	-0,03736393	6,41082114
43	681	6,523562306	280,51318	1849	7,980106	-0,03736393	6,37345722
44	525	6,263398263	275,58952	1936	7,980106	-0,03736393	6,33609329
45	636	6,455198563	290,48394	2025	7,980106	-0,03736393	6,29872937
46	604	6,403574198	294,56441	2116	7,980106	-0,03736393	6,26136544
47	542	6,295266001	295,87750	2209	7,980106	-0,03736393	6,22400152
48	610	6,413458957	307,84603	2304	7,980106	-0,03736393	6,18663759
49	570	6,345636361	310,93618	2401	7,980106	-0,03736393	6,14927367
50	619	6,428105273	321,40526	2500	7,980106	-0,03736393	6,11190974
51	431	6,06610809	309,37151	2601	7,980106	-0,03736393	6,07454582
52	566	6,338594078	329,60689	2704	7,980106	-0,03736393	6,03718189
53	602	6,400257445	339,21364	2809	7,980106	-0,03736393	5,99981797
54	578	6,359573869	343,41699	2916	7,980106	-0,03736393	5,96245404
55	525	6,263398263	344,48690	3025	7,980106	-0,03736393	5,92509012
56	575	6,354370041	355,84472	3136	7,980106	-0,03736393	5,88772619
57	482	6,177944114	352,14281	3249	7,980106	-0,03736393	5,85036227
58	433	6,070737728	352,10279	3364	7,980106	-0,03736393	5,81299834
59	454	6,118097198	360,96773	3481	7,980106	-0,03736393	5,77563442
60	534	6,280395839	376,82375	3600	7,980106	-0,03736393	5,73827049
61	437	6,079933195	370,87592	3721	7,980106	-0,03736393	5,70090657
62	464	6,139884552	380,67284	3844	7,980106	-0,03736393	5,66354264
63	420	6,040254711	380,53605	3969	7,980106	-0,03736393	5,62617872
64	415	6,02827852	385,80983	4096	7,980106	-0,03736393	5,58881479
65	260	5,560681631	361,44431	4225	7,980106	-0,03736393	5,55145087
66	333	5,80814249	383,33740	4356	7,980106	-0,03736393	5,51408694
67	382	5,945420609	398,34318	4489	7,980106	-0,03736393	5,47672302
68	323	5,777652323	392,88036	4624	7,980106	-0,03736393	5,43935909
69	285	5,65248918	390,02175	4761	7,980106	-0,03736393	5,40199517
70	269	5,59471138	391,62980	4900	7,980106	-0,03736393	5,36463124
71	474	6,161207322	437,44572	5041	7,980106	-0,03736393	5,32726732
72	174	5,159055299	371,45198	5184	7,980106	-0,03736393	5,28990339
73	195	5,272999559	384,92897	5329	7,980106	-0,03736393	5,25253947
74	236	5,463831805	404,32355	5476	7,980106	-0,03736393	5,21517554
75	369	5,910796644	443,30975	5625	7,980106	-0,03736393	5,17781162
76	274	5,613128106	426,59774	5776	7,980106	-0,03736393	5,14044769
77	243	5,493061443	422,96573	5929	7,980106	-0,03736393	5,10308377

(b) b



[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
117	43	3,761200116	440,06041	13689	7,980106	-0,03736393	3,6085267
118	66	4,189654742	494,37926	13924	7,980106	-0,03736393	3,5711628
119	47	3,850147602	458,16756	14161	7,980106	-0,03736393	3,5337989
120	49	3,891820298	467,01844	14400	7,980106	-0,03736393	3,4964349
121	33	3,496507561	423,07741	14641	7,980106	-0,03736393	3,4590710
122	14	2,63905733	321,96499	14884	7,980106	-0,03736393	3,4217071
123	18	2,890371758	355,51573	15129	7,980106	-0,03736393	3,3843432
124	31	3,433987204	425,81441	15376	7,980106	-0,03736393	3,3469792
125	34	3,526360525	440,79507	15625	7,980106	-0,03736393	3,3096153
126	30	3,401197382	428,55087	15876	7,980106	-0,03736393	3,2722514
127	8	2,079441542	264,08908	16129	7,980106	-0,03736393	3,2348875
128	22	3,091042453	395,65343	16384	7,980106	-0,03736393	3,1975235
129	6	1,791759469	231,13697	16641	7,980106	-0,03736393	3,1601596
130	23	3,135494216	407,61425	16900	7,980106	-0,03736393	3,1227957
131	21	3,044522438	398,83244	17161	7,980106	-0,03736393	3,0854318
132	30	3,401197382	448,95805	17424	7,980106	-0,03736393	3,0480678
133	15	2,708050201	360,17068	17689	7,980106	-0,03736393	3,0107039
134	21	3,044522438	407,96601	17956	7,980106	-0,03736393	2,9733400
135	7	1,945910149	262,69787	18225	7,980106	-0,03736393	2,9359761
136	8	2,079441542	282,80405	18496	7,980106	-0,03736393	2,8986121
137	30	3,401197382	465,96404	18769	7,980106	-0,03736393	2,8612482
138	15	2,708050201	373,71093	19044	7,980106	-0,03736393	2,8238843
139	12	2,48490665	345,40202	19321	7,980106	-0,03736393	2,7865204
140	12	2,48490665	347,88693	19600	7,980106	-0,03736393	2,7491564
141	7	1,945910149	274,37333	19881	7,980106	-0,03736393	2,7117925
142	9	2,197224577	312,00589	20164	7,980106	-0,03736393	2,6744286
143	13	2,564949357	366,78776	20449	7,980106	-0,03736393	2,6370647
144	17	2,833213344	407,98272	20736	7,980106	-0,03736393	2,5997007
145	13	2,564949357	371,91766	21025	7,980106	-0,03736393	2,5623368
146	20	2,995732274	437,37691	21316	7,980106	-0,03736393	2,5249729
147	11	2,397895273	352,49061	21609	7,980106	-0,03736393	2,4876090
148	14	2,63905733	390,58048	21904	7,980106	-0,03736393	2,4502450
149	3	1,098612289	163,69323	22201	7,980106	-0,03736393	2,4128811
150	13	2,564949357	384,74240	22500	7,980106	-0,03736393	2,3755172
151	15	2,708050201	408,91558	22801	7,980106	-0,03736393	2,3381533
152	9	2,197224577	333,97814	23104	7,980106	-0,03736393	2,3007893
153	10	2,302585093	352,29552	23409	7,980106	-0,03736393	2,2634254
154	5	1,609437912	247,85344	23716	7,980106	-0,03736393	2,2260615
155	5	1,609437912	249,46288	24025	7,980106	-0,03736393	2,1886976
156	5	1,609437912	251,07231	24336	7,980106	-0,03736393	2,1513336
157	5	1,609437912	252,68175	24649	7,980106	-0,03736393	2,1139697
158	11	2,397895273	378,86745	24964	7,980106	-0,03736393	2,0766058
159	6	1,791759469	284,88976	25281	7,980106	-0,03736393	2,0392419
160	3	1,098612289	175,77797	25600	7,980106	-0,03736393	2,0018779
161	9	2,197224577	353,75316	25921	7,980106	-0,03736393	1,9645140
162	5	1,609437912	260,72894	26244	7,980106	-0,03736393	1,9271501
163	8	2,079441542	338,94897	26569	7,980106	-0,03736393	1,8897862
164	12	2,48490665	407,52469	26896	7,980106	-0,03736393	1,8524222
165	5	1,609437912	265,55726	27225	7,980106	-0,03736393	1,8150583

(c) c

Figure 4.2: Non-optimized Table

- Table (a)is X=1 to 29
- Table (b and c) is X=30 to 169

In Figure 4.1 and Figure 4.2, it can be seen that the two regression lines for each area of the data.

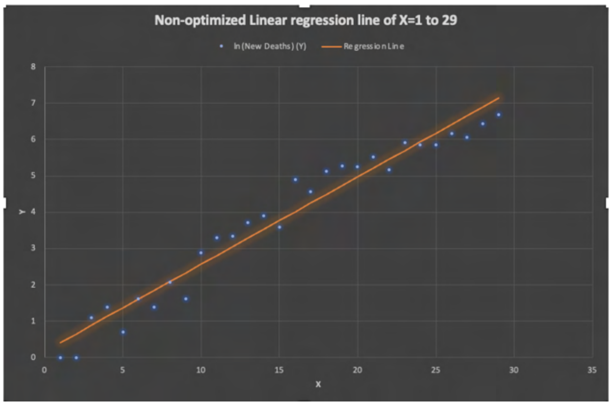


Figure 4.3: Non-optimized Scattered plot A, x = 1 to 29

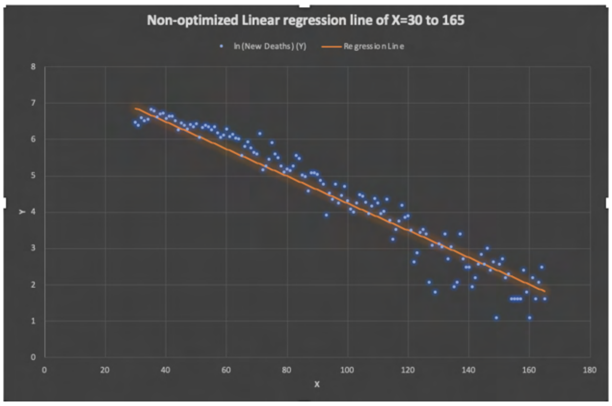


Figure 4.4: Non-optimized Scattered plot B, x = 29 to 165

4.2.2. MODIFICATION OF SLOPES AND y-INTERCEPTS

To optimize these two regression lines under the constraint that the two would meet at an arbitrary user selected point, the Equation 3.1.3 will be solved for the data of COVID-19 deaths to find the modified values of slopes and y-intercepts of the regression lines.

For x^o=29,

The system of linear equations is given by,

$$A = \begin{bmatrix} 58 & 870 & 0 & 0 & 1 \\ 870 & 17110 & 0 & 0 & 29 \\ 0 & 0 & 272 & 26520 & -1 \\ 0 & 0 & 26520 & 3004920 & -29 \\ 1 & 29 & -1 & -29 & 0 \end{bmatrix}, x = \begin{bmatrix} \beta_0, \\ \beta_1 \\ \gamma_0, \\ \gamma_1, \\ \lambda \end{bmatrix}, b = \begin{bmatrix} 218.62, \\ 4255.22, \\ 1199.27, \\ 99823.43, \\ 0, \end{bmatrix}$$

From the excel output

AX=B, X=A^-1 * B

MATRIX A of variables				
58	870	0	0	1
870	17110	0	0	29
0	0	272	26520	-1
0	0	26520	3004920	-29
1	29	-1	-29	0

INVERSE OF MATRIX A				
0,057868307	-0,0022154	-0,00841102	7,0092E-05	-0,42896194
-0,002215402	9,8388E-05	0,0008411	-7,0092E-06	0,04289619
-0,008411018	0,0008411	0,02156978	-0,00019272	-0,24391953
7,00918E-05	-7,0092E-06	-0,00019272	2,0532E-06	0,00203266
-0,42896194	0,04289619	-0,24391953	0,00203266	-12,4398963

B
218,617
4255,220
1199,266
99823,432
0

Figure 4.5: Matrix Table

Continuation from the excel output in Figure 4.1 we have,
Solving for x is given by,

$$x = A^{-1}b \tag{4.2.5}$$

The values of the modified slopes and y-intercepts are calculated from the computation of the excel out in Figure 4.1 and Figure 4.5 and the results are

CORRECTED VALUES	
β1*	0,1338
β2*	0,2434
γ1*	8,3703
γ2*	-0,0407
λ	-0,8628

Figure 4.6: Optimized slope and y-intercept for the two separate regression lines

of first regression line are calculated as,

$$\beta_0 = 0,1338$$

$$\beta_1 = 0,2434$$

The values of the modified slope and y-intercept of second regression line are calculated as,

$$\gamma_0 = 8,3703$$

$$\gamma_1 = -0,0407$$

Here is the modified regression function for the first line,

$$\widehat{\boldsymbol{y}} = 0,1338 + 0.2434x \tag{4.2.6}$$

β1*	β2*	R L
0,1338	0,2434	0,3772
0,1338	0,2434	0,6206
0,1338	0,2434	0,864
0,1338	0,2434	1,1074
0,1338	0,2434	1,3508
0,1338	0,2434	1,5942
0,1338	0,2434	1,8376
0,1338	0,2434	2,081
0,1338	0,2434	2,3244
0,1338	0,2434	2,5678
0,1338	0,2434	2,8112
0,1338	0,2434	3,0546
0,1338	0,2434	3,298
0,1338	0,2434	3,5414
0,1338	0,2434	3,7848
0,1338	0,2434	4,0282
0,1338	0,2434	4,2716
0,1338	0,2434	4,515
0,1338	0,2434	4,7584
0,1338	0,2434	5,0018
0,1338	0,2434	5,2452
0,1338	0,2434	5,4886
0,1338	0,2434	5,732
0,1338	0,2434	5,9754
0,1338	0,2434	6,2188
0,1338	0,2434	6,4622
0,1338	0,2434	6,7056
0,1338	0,2434	6,949
0,1338	0,2434	7,1924

Figure 4.7: Optimized table for x=1 to 29

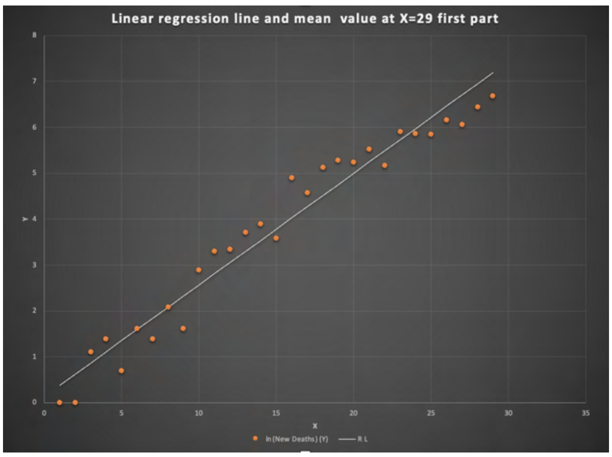


Figure 4.8: Optimized table for x=1 to 29

Here is the modified regression function for the second line,

$$\widehat{\boldsymbol{y}} = 8.3703 + -0.0407x \tag{4.2.7}$$

y1*	y2*	R L
8,3703	-0,0407	7,1493
8,3703	-0,0407	7,1086
8,3703	-0,0407	7,0679
8,3703	-0,0407	7,0272
8,3703	-0,0407	6,9865
8,3703	-0,0407	6,9458
8,3703	-0,0407	6,9051
8,3703	-0,0407	6,8644
8,3703	-0,0407	6,8237
8,3703	-0,0407	6,783
8,3703	-0,0407	6,7423
8,3703	-0,0407	6,7016
8,3703	-0,0407	6,6609
8,3703	-0,0407	6,6202
8,3703	-0,0407	6,5795
8,3703	-0,0407	6,5388
8,3703	-0,0407	6,4981
8,3703	-0,0407	6,4574
8,3703	-0,0407	6,4167
8,3703	-0,0407	6,376
8,3703	-0,0407	6,3353
8,3703	-0,0407	6,2946
8,3703	-0,0407	6,2539
8,3703	-0,0407	6,2132
8,3703	-0,0407	6,1725
8,3703	-0,0407	6,1318
8,3703	-0,0407	6,0911
8,3703	-0,0407	6,0504
8,3703	-0,0407	6,0097
8,3703	-0,0407	5,969
8,3703	-0,0407	5,9283
8,3703	-0,0407	5,8876
8,3703	-0,0407	5,8469
8,3703	-0,0407	5,8062
8,3703	-0,0407	5,7655
8,3703	-0,0407	5,7248
8,3703	-0,0407	5,6841
8,3703	-0,0407	5,6434
8,3703	-0,0407	5,6027
8,3703	-0,0407	5,562
8,3703	-0,0407	5,5213
8,3703	-0,0407	5,4806
8,3703	-0,0407	5,4399
8,3703	-0,0407	5,3992
8,3703	-0,0407	5,3585
8,3703	-0,0407	5,3178
8,3703	-0,0407	5,2771
8,3703	-0,0407	5,2364

(a) a

8,3703	-0,0407	3,6084
8,3703	-0,0407	3,5677
8,3703	-0,0407	3,527
8,3703	-0,0407	3,4863
8,3703	-0,0407	3,4456
8,3703	-0,0407	3,4049
8,3703	-0,0407	3,3642
8,3703	-0,0407	3,3235
8,3703	-0,0407	3,2828
8,3703	-0,0407	3,2421
8,3703	-0,0407	3,2014
8,3703	-0,0407	3,1607
8,3703	-0,0407	3,12
8,3703	-0,0407	3,0793
8,3703	-0,0407	3,0386
8,3703	-0,0407	2,9979
8,3703	-0,0407	2,9572
8,3703	-0,0407	2,9165
8,3703	-0,0407	2,8758
8,3703	-0,0407	2,8351
8,3703	-0,0407	2,7944
8,3703	-0,0407	2,7537
8,3703	-0,0407	2,713
8,3703	-0,0407	2,6723
8,3703	-0,0407	2,6316
8,3703	-0,0407	2,5909
8,3703	-0,0407	2,5502
8,3703	-0,0407	2,5095
8,3703	-0,0407	2,4688
8,3703	-0,0407	2,4281
8,3703	-0,0407	2,3874
8,3703	-0,0407	2,3467
8,3703	-0,0407	2,306
8,3703	-0,0407	2,2653
8,3703	-0,0407	2,2246
8,3703	-0,0407	2,1839
8,3703	-0,0407	2,1432
8,3703	-0,0407	2,1025
8,3703	-0,0407	2,0618
8,3703	-0,0407	2,0211
8,3703	-0,0407	1,9804
8,3703	-0,0407	1,9397
8,3703	-0,0407	1,899
8,3703	-0,0407	1,8583
8,3703	-0,0407	1,8176
8,3703	-0,0407	1,7769
8,3703	-0,0407	1,7362
8,3703	-0,0407	1,6955
8,3703	-0,0407	1,6548

(b) b

Figure 4.9: Optimized table for x=30 to 165

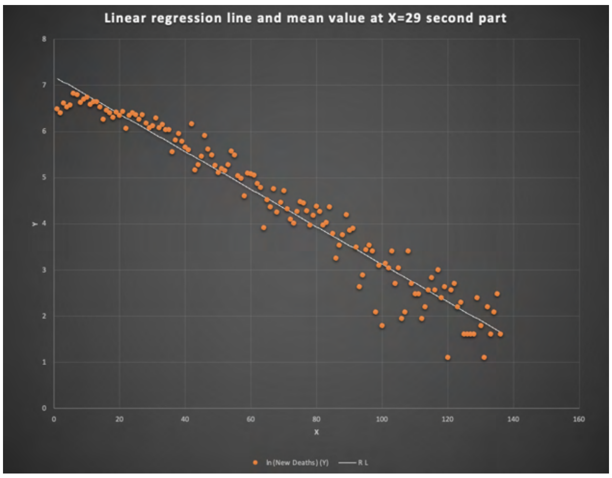


Figure 4.10: Optimized linear regression line for x=30 to 165

By drawing these regression lines, It can be observed that the two regression lines are meeting at x^o=29. we can observed this in Figure 4.11

In Figure 4.11, it can be seen that the two modified regression lines for each area of the data are meeting at the point x^o=29.

Figure 4.11: Joined optimized linear regression line

4.2.3. CONFIDENCE INTERVAL ESTIMATE OF THE MODIFIED REGRESSION FUNCTIONS

A confidence interval would be a more realistic way of expressing the ln(New Death).

So, we calculated the result for the confidence interval for mean and individual value using the following equation:

- Mean value formula

$$\left\langle v_0 - t_1 - \alpha_2 s\sqrt{v^*}; y_0 + t_{1-\alpha} s^2 \sqrt{v}^* \right\rangle$$

- Predicted value formula

$$\left\langle y^0 - t_{1-\alpha/2} s\sqrt{v^* + 1}; y^0 + t_{1-\alpha/2} s\sqrt{v^* + 1} \right\rangle$$

So based on these results, with 95% confidence that for every individual of each Order (x) there predicted $(\widehat{y})$ is between upper and the lower confidence limit displayed above.

For x^o=29

The Figure 4.3 and Figure 4.4 shows the confidence interval of each area.

- Plot for the mean and predicted value

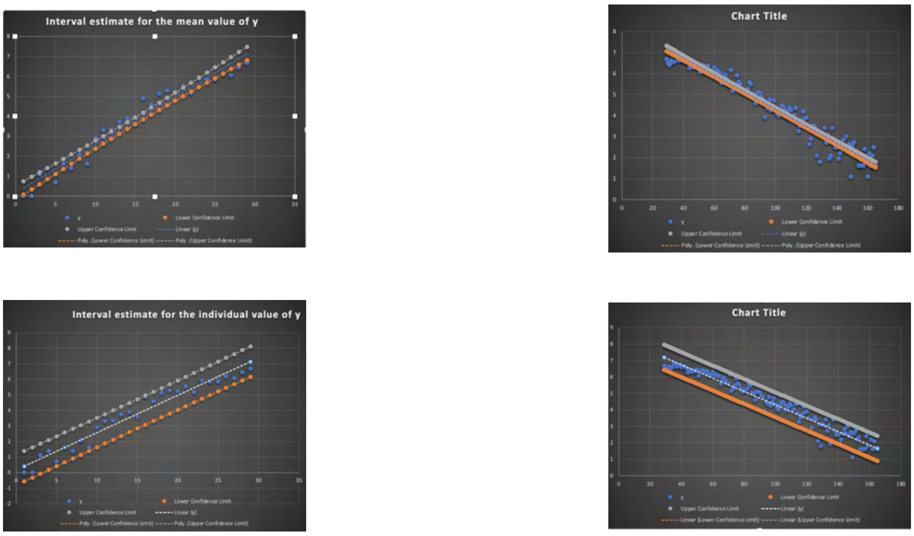


Figure 4.12: Mean and Predicted confidence Interval for Non-optimized linear regression lines

For the modified regression functions, the formula for the confidence interval will be modified using Equation 3.1.4,

$$\varphi(x) = \begin{cases} \beta_0 + \beta_1 x & x < x_0 \\ \gamma_0 + \gamma_1 x & x > x_0 \end{cases} \tag{4.2.8}$$

– The mean value is given by:

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^*}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^*} \right\rangle \tag{4.2.9}$$

while

– The Predicted value is given by

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^*}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^*} \right\rangle \tag{4.2.10}$$

$$\left\langle \phi(x) - t_{1-\alpha/2} s\sqrt{h^* + 1}; \phi(x) + t_{1-\alpha/2} s\sqrt{h^* + 1} \right\rangle \tag{4.2.11}$$

where,

$$h^* = \begin{bmatrix} 1, x, 1, x \end{bmatrix} \begin{bmatrix} X^T X \end{bmatrix} \begin{bmatrix} 1 \\ x \\ 1 \\ x \end{bmatrix}$$

For variable estimation, the formula will be modified as,

$$S^*_{min} = \sum_{i=1}^{n} (y - \phi(x_i))^2 \tag{4.2.12}$$

where,

$$s^2 = S^*_{min}/n - m \tag{4.2.13}$$

where,

m = number of estimated parameters

So calculating the values for x^o=29 from the excel computation

n =	165
m	4
df =	161
t table =	1,9748
S* =	24,69551
S =	0,3916481
Se	0,0462635
t table *Se	0,0913615

Se	0,3943711
t table *Se	0,7788073

Figure 4.13: Confidence Interval computation

which give the following table,

LC	UC	LC	UC
0,2858385	0,4685615	-0,4016073	1,1560073
0,5292385	0,71196148	-0,15820727	1,39940727
0,77263852	0,95536148	0,08519273	1,64280727
1,01603852	1,19876148	0,32859273	1,88620727
1,25943852	1,44216148	0,57199273	2,12960727
1,50283852	1,68556148	0,81539273	2,37300727
1,74623852	1,92896148	1,05879273	2,61640727
1,98963852	2,17236148	1,30219273	2,85980727
2,23303852	2,41576148	1,54559273	3,10320727
2,47643852	2,65916148	1,78899273	3,34660727
2,71983852	2,90256148	2,03239273	3,59000727
2,96323852	3,14596148	2,27579273	3,83340727
3,20663852	3,38936148	2,51919273	4,07680727
3,45003852	3,63276148	2,76259273	4,32020727
3,69343852	3,87616148	3,00599273	4,56360727
3,93683852	4,11956148	3,24939273	4,80700727
4,18023852	4,36296148	3,49279273	5,05040727
4,42363852	4,60636148	3,73619273	5,29380727
4,66703852	4,84976148	3,97959273	5,53720727
4,91043852	5,09316148	4,22299273	5,78060727
5,1538385	5,33656148	4,46639273	6,02400727
5,39723852	5,57996148	4,70979273	6,26740727
5,64063852	5,82336148	4,95319273	6,51080727
5,88403852	6,06676148	5,19659273	6,75420727
6,12743852	6,31016148	5,43999273	6,99760727
6,37083852	6,55356148	5,68339273	7,24100727
6,61423852	6,79696148	5,92679273	7,48440727
6,85763852	7,04036148	6,17019273	7,72780727
7,10103852	7,28376148	6,41359273	7,97120727

Figure 4.14: Optimized confidence interval table for mean and predicted values from x=1 to x=29

LC	UC	LC	UC
7,0579385	7,2406615	6,3704927	7,9281073
7,01723852	7,19996148	6,32979273	7,88740727
6,97653852	7,15926148	6,28909273	7,84670727
6,93583852	7,11856148	6,24839273	7,80600727
6,89513852	7,07786148	6,20769273	7,76530727
6,85443852	7,03716148	6,16699273	7,72460727
6,81373852	6,99646148	6,12629273	7,68390727
6,77303852	6,95576148	6,08559273	7,64320727
6,73233852	6,91506148	6,04489273	7,60250727
6,69163852	6,87436148	6,00419273	7,56180727
6,65093852	6,83366148	5,96349273	7,52110727
6,61023852	6,79296148	5,92279273	7,48040727
6,56953852	6,75226148	5,88209273	7,43970727
6,52883852	6,71156148	5,84139273	7,39900727
6,48813852	6,67086148	5,80069273	7,35830727
6,44743852	6,63016148	5,75999273	7,31760727
6,40673852	6,58946148	5,71929273	7,27690727
6,36603852	6,54876148	5,67859273	7,23620727
6,32533852	6,50806148	5,63789273	7,19550727
6,28463852	6,46736148	5,59719273	7,15480727
6,24393852	6,42666148	5,55649273	7,11410727
6,20323852	6,38596148	5,51579273	7,07340727
6,16253852	6,34526148	5,47509273	7,03270727
6,12183852	6,30456148	5,43439273	6,99200727
6,08113852	6,26386148	5,39369273	6,95130727
6,04043852	6,22316148	5,35299273	6,91060727
5,99973852	6,18246148	5,31229273	6,86990727
5,95903852	6,14176148	5,27159273	6,82920727
5,91833852	6,10106148	5,23089273	6,78850727
5,87763852	6,06036148	5,19019273	6,74780727
5,83693852	6,01966148	5,14949273	6,70710727
5,79623852	5,97896148	5,10879273	6,66640727
5,75553852	5,93826148	5,06809273	6,62570727
5,71483852	5,89756148	5,02739273	6,58500727
5,67413852	5,85686148	4,98669273	6,54430727
5,63343852	5,81616148	4,94599273	6,50360727
5,59273852	5,77546148	4,90529273	6,46290727
5,55203852	5,73476148	4,86459273	6,42220727
5,51133852	5,69406148	4,82389273	6,38150727
5,47063852	5,65336148	4,78319273	6,34080727
5,42993852	5,61266148	4,74249273	6,30010727
5,38923852	5,57196148	4,70179273	6,25940727
5,34853852	5,53126148	4,66109273	6,21870727
5,30783852	5,49056148	4,62039273	6,17800727
5,26713852	5,44986148	4,57969273	6,13730727
5,22643852	5,40916148	4,53899273	6,09660727
5,18573852	5,36846148	4,49829273	6,05590727
5,14503852	5,32776148	4,45759273	6,01520727

5,91833852	6,10106148	5,23089273	6,78850727
5,87763852	6,06036148	5,19019273	6,74780727
5,83693852	6,01966148	5,14949273	6,70710727
5,79623852	5,97896148	5,10879273	6,66640727
5,75553852	5,93826148	5,06809273	6,62570727
5,71483852	5,89756148	5,02739273	6,58500727
5,67413852	5,85686148	4,98669273	6,54430727
5,63343852	5,81616148	4,94599273	6,50360727
5,59273852	5,77546148	4,90529273	6,46290727
5,55203852	5,73476148	4,86459273	6,42220727
5,51133852	5,69406148	4,82389273	6,38150727
5,47063852	5,65336148	4,78319273	6,34080727
5,42993852	5,61266148	4,74249273	6,30010727
5,38923852	5,57196148	4,70179273	6,25940727
5,34853852	5,53126148	4,66109273	6,21870727
5,30783852	5,49056148	4,62039273	6,17800727
5,26713852	5,44986148	4,57969273	6,13730727
5,22643852	5,40916148	4,53899273	6,09660727
5,18573852	5,36846148	4,49829273	6,05590727
5,14503852	5,32776148	4,45759273	6,01520727
5,10433852	5,28706148	4,41689273	5,97450727
5,06363852	5,24636148	4,37619273	5,93380727
5,02293852	5,20566148	4,33549273	5,89310727
4,98223852	5,16496148	4,29479273	5,85240727
4,94153852	5,12426148	4,25409273	5,81170727
4,90083852	5,08356148	4,21339273	5,77100727
4,86013852	5,04286148	4,17269273	5,73030727
4,81943852	5,00216148	4,13199273	5,68960727
4,77873852	4,96146148	4,09129273	5,64890727
4,73803852	4,92076148	4,05059273	5,60820727
4,69733852	4,88006148	4,00989273	5,56750727
4,65663852	4,83936148	3,96919273	5,52680727
4,61593852	4,79866148	3,92849273	5,48610727
4,57523852	4,75796148	3,88779273	5,44540727
4,53453852	4,71726148	3,84709273	5,40470727
4,49383852	4,67656148	3,80639273	5,36400727
4,45313852	4,63586148	3,76569273	5,32330727
4,41243852	4,59516148	3,72499273	5,28260727
4,37173852	4,55446148	3,68429273	5,24190727
4,33103852	4,51376148	3,64359273	5,20120727
4,29033852	4,47306148	3,60289273	5,16050727
4,24963852	4,43236148	3,56219273	5,11980727
4,20893852	4,39166148	3,52149273	5,07910727
4,16823852	4,35096148	3,48079273	5,03840727
4,12753852	4,31026148	3,44009273	4,99770727
4,08683852	4,26956148	3,39939273	4,95700727
4,04613852	4,22886148	3,35869273	4,91630727
4,00543852	4,18816148	3,31799273	4,87560727
3,96473852	4,14746148	3,27729273	4,83490727

3,55773852	3,74046148	2,87029273	4,42790727
3,51703852	3,69976148	2,82959273	4,38720727
3,47633852	3,65906148	2,78889273	4,34650727
3,43563852	3,61836148	2,74819273	4,30580727
3,39493852	3,57766148	2,70749273	4,26510727
3,35423852	3,53696148	2,66679273	4,22440727
3,31353852	3,49626148	2,62609273	4,18370727
3,27283852	3,45556148	2,58539273	4,14300727
3,23213852	3,41486148	2,54469273	4,10230727
3,19143852	3,37416148	2,50399273	4,06160727
3,15073852	3,33346148	2,46329273	4,02090727
3,11003852	3,29276148	2,42259273	3,98020727
3,06933852	3,25206148	2,38189273	3,93950727
3,02863852	3,21136148	2,34119273	3,89880727
2,98793852	3,17066148	2,30049273	3,85810727
2,94723852	3,12996148	2,25979273	3,81740727
2,90653852	3,08926148	2,21909273	3,77670727
2,86583852	3,04856148	2,17839273	3,73600727
2,82513852	3,00786148	2,13769273	3,69530727
2,78443852	2,96716148	2,09699273	3,65460727
2,74373852	2,92646148	2,05629273	3,61390727
2,70303852	2,88576148	2,01559273	3,57320727
2,66233852	2,84506148	1,97489273	3,53250727
2,62163852	2,80436148	1,93419273	3,49180727
2,58093852	2,76366148	1,89349273	3,45110727
2,54023852	2,72296148	1,85279273	3,41040727
2,49953852	2,68226148	1,81209273	3,36970727
2,45883852	2,64156148	1,77139273	3,32900727
2,41813852	2,60086148	1,73069273	3,28830727
2,37743852	2,56016148	1,68999273	3,24760727
2,33673852	2,51946148	1,64929273	3,20690727
2,29603852	2,47876148	1,60859273	3,16620727
2,25533852	2,43806148	1,56789273	3,12550727
2,21463852	2,39736148	1,52719273	3,08480727
2,17393852	2,35666148	1,48649273	3,04410727
2,13323852	2,31596148	1,44579273	3,00340727
2,09253852	2,27526148	1,40509273	2,96270727
2,05183852	2,23456148	1,36439273	2,92200727
2,01113852	2,19386148	1,32369273	2,88130727
1,97043852	2,15316148	1,28299273	2,84060727
1,92973852	2,11246148	1,24229273	2,79990727
1,88903852	2,07176148	1,20159273	2,75920727
1,84833852	2,03106148	1,16089273	2,71850727
1,80763852	1,99036148	1,12019273	2,67780727
1,76693852	1,94966148	1,07949273	2,63710727
1,72623852	1,90896148	1,03879273	2,59640727
1,68553852	1,86826148	0,99809273	2,55570727
1,64483852	1,82756148	0,95739273	2,51500727
1,60413852	1,78686148	0,91669273	2,47430727

(a) a (b) b (c) c

Figure 4.15: Optimized confidence interval table for mean and predicted value from x=29 to x=165

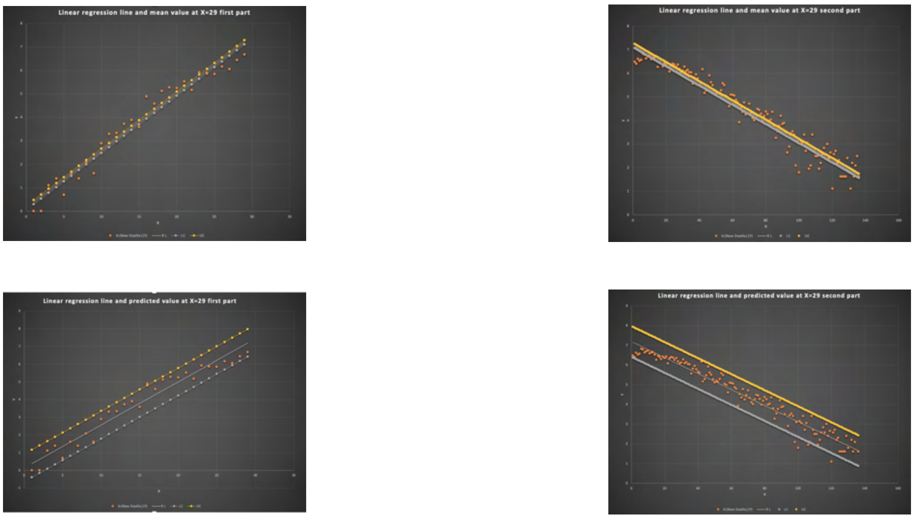


Figure 4.16: Mean and Predicted confidence Interval for optimized linear regression lines

- Joined confidence interval for optimized linear regression lines for mean value

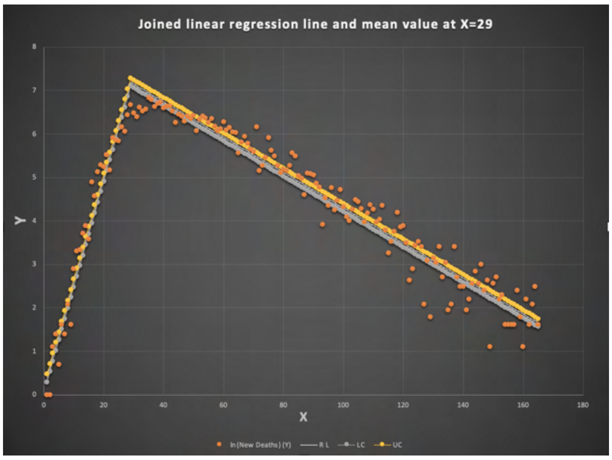


Figure 4.17: Scattered plot of the optimized confidence interval of two linear regression line for the mean value

- Joined confidence interval for optimized linear regression lines for predicted value

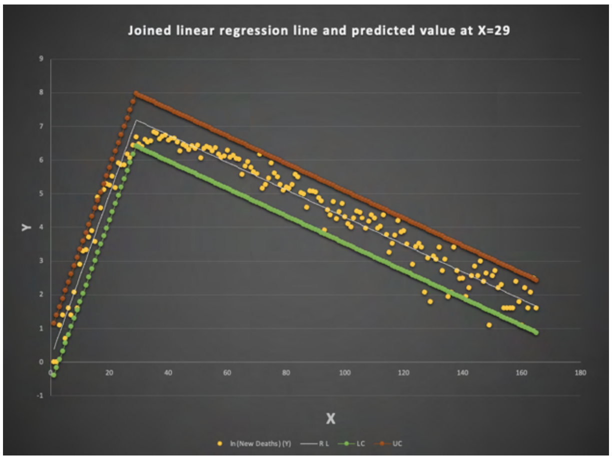


Figure 4.18: Scattered plot of the optimized confidence interval of two linear regression line for the predicted value

Below is the diagram of the scattered plot and the joined optimized regression lines

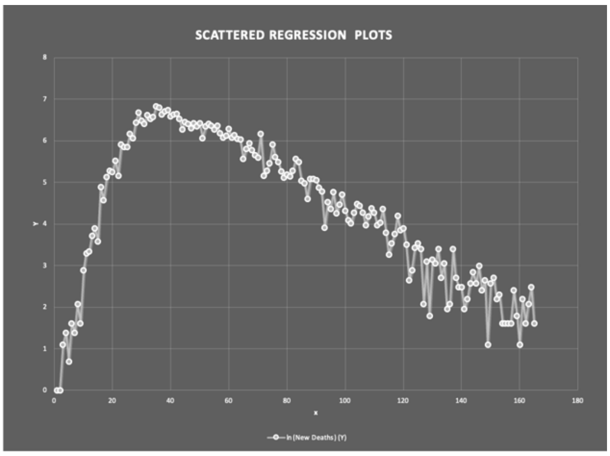


Figure 4.19: Scattered linear regression plot

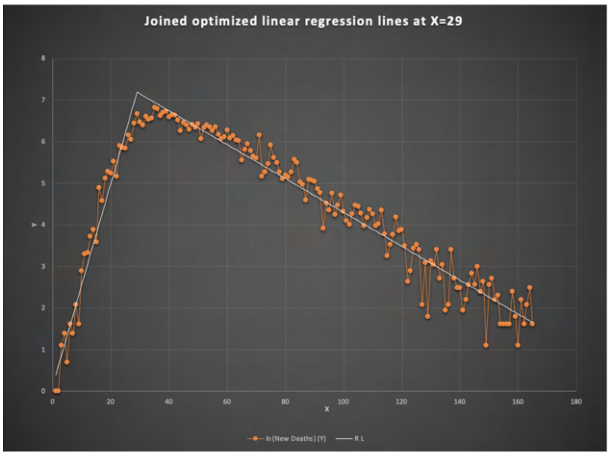
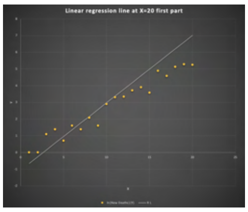
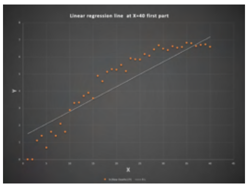


Figure 4.20: Scattered plot and the joined optimized linear regression lines

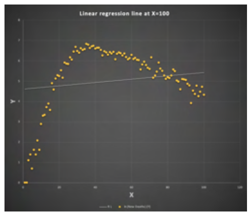
Figure 4.20-4.28 shows the results of modified regression lines and their corresponding confidence intervals for different values of x^o ranging from x^o=20 to x^o=100.



(a) At $X^0 = 20$

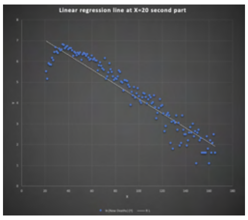


(b) At $X^0 = 40$

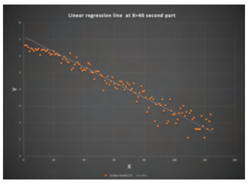


(c) At $X^0 = 100$

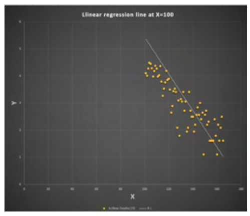
Figure 4.21: Linear regression line at different X^0 for first part



(a) At $X^0 = 20$



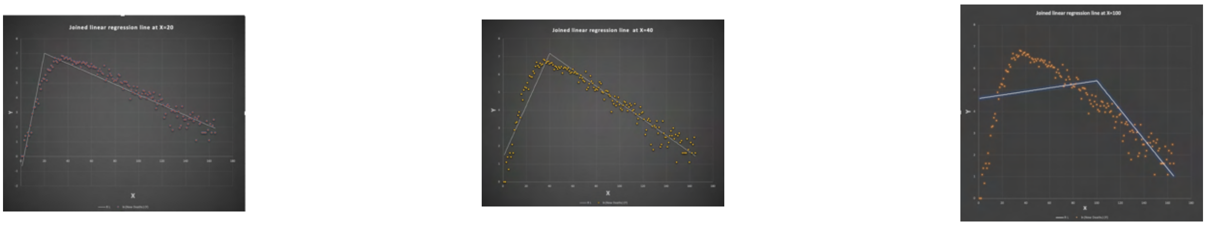
(b) At $X^0 = 40$



(c) At $X^0 = 100$

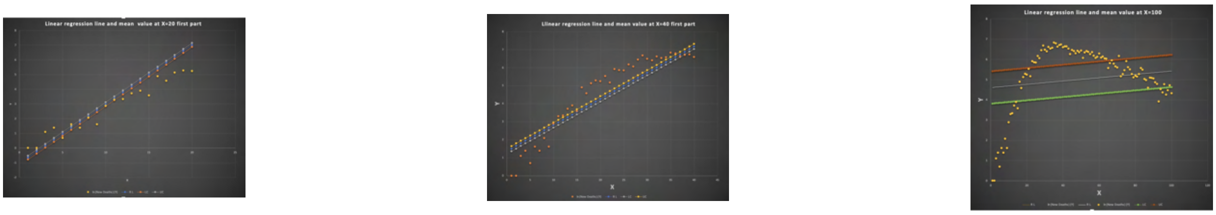
Figure 4.22: Linear regression line at different X^0 for second part

Below are the joined linear regression line at $X^0 = 20, X^0 = 40$ and $X^0 = 100$



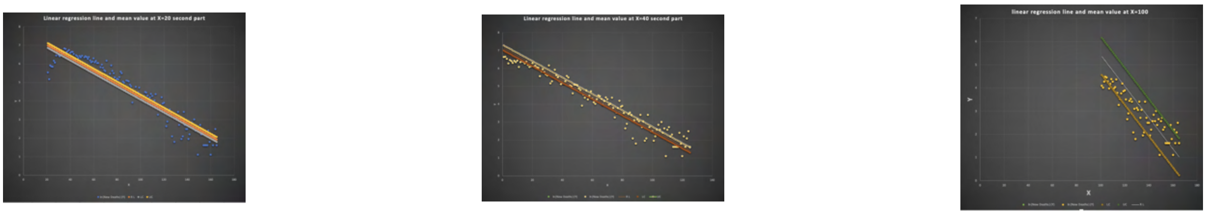
(a) $X^0 = 20$ (b) $X^0 = 40$ (c) $X^0 = 100$

Figure 4.23: Joined Linear regression line at different X^0



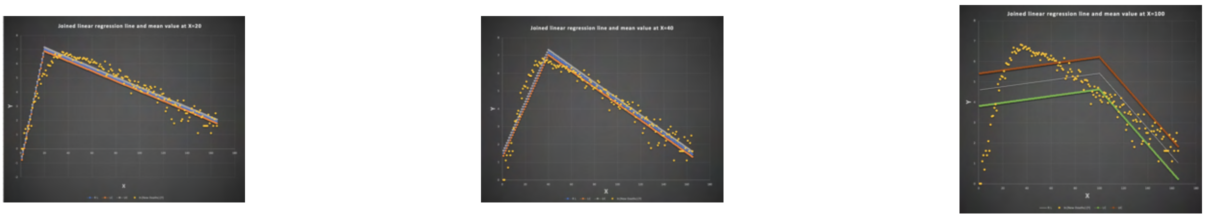
(a) At $X^0 = 20$ (b) At $X^0 = 40$ (c) At $X^0 = 100$

Figure 4.24: Mean Value Linear Confidence Interval at different X^0 for first part



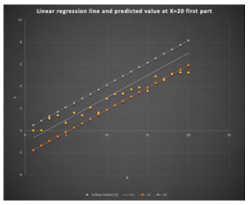
(a) At $X^0 = 20$ (b) At $X^0 = 40$ (c) At $X^0 = 100$

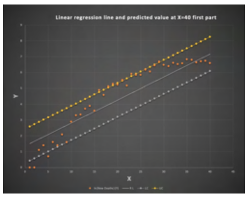
Figure 4.25: Mean Value Linear Confidence Interval at different X^0 for second part



(a) $X^0 = 20$ (b) $X^0 = 40$ (c) $X^0 = 100$

Figure 4.26: Joined Linear Confidence Interval for Mean Value at different X^0


(a) At $X^0 = 20$


(b) At $X^0 = 40$

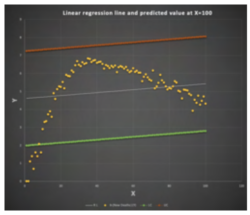
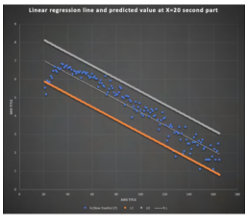
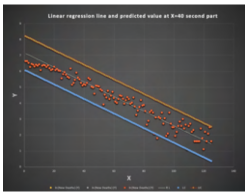

(c) At $X^0 = 100$

Figure 4.27: Linear Confidence Interval for Predicted Value at different X^0 for first part


(a) At $X^0 = 20$


(b) At $X^0 = 40$

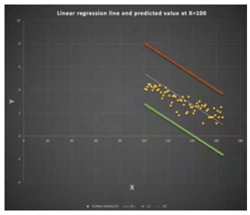
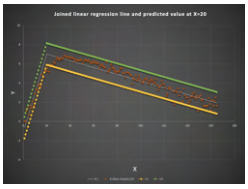
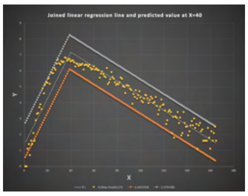

(c) At $X^0 = 100$

Figure 4.28: Predicted Value Linear Confidence Interval at different X^0 for second part


(a) $X^0 = 20$


(b) $X^0 = 40$

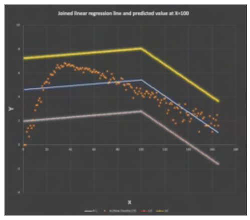

(c) $X^0 = 100$

Figure 4.29: Joined Linear Confidence Interval for Predicted Value at different X^0

4.3. MODIFIED NON-LINEAR REGRESSION LINES

In this section we will transform the optimized regression lines to an exponential form in order to have our original data of COVID-19 deaths in Italy exactly the way it was explained in chapter 2.4. The transformation model will be in the form.

$$y = e^{\beta_0} e^{\beta_1 x} \tag{4.3.1}$$

Transforming the optimized regression lines for x^0=29 by taking their exponential form, the excel computation output is shown below:

Enter Xo (n1) =	29
(n2) =	136

β1*	0,1338
β2*	0,2434
γ1*	8,3703
γ2*	-0,0407
λ	-0,8628

e^β1* =	1,143164164
e^γ1* =	4316,930948
y = (e^β1*)(e^β2*x)	

Figure 4.30: Computation of the transformed model

The model equation from the output of the first part can be written mathematically as follows

$$\widehat{\boldsymbol{y}} = 1,1432e^{0,2434x} \tag{4.3.2}$$

and the second part can be written as

$$\widehat{\boldsymbol{y}} = 1,1432e^{0,2434x} \tag{4.3.3}$$

The results of the model are given in the table below

order (X)	New Deaths	e^β2*x	y
1	1	1,275578754	1,458195919
2	1	1,627101157	1,860043733
3	3	2,075495665	2,372632267
4	4	2,647458174	3,02647931
5	2	3,377041398	3,860512706
6	5	4,307682257	4,924387986
7	4	5,494787964	6,281444689
8	8	7,009034782	8,012477387
9	5	8,940575851	10,22054592
10	18	11,4044086	13,03711122
11	27	14,54722131	16,62986208
12	28	18,55612642	21,21269875
13	41	23,66980062	27,05846783
14	49	30,19269477	34,51520667
15	36	38,51315996	44,02686431
16	133	49,12656858	56,15973269
17	97	62,66480711	71,63616183
18	168	79,93389655	91,37756602
19	196	101,9619801	116,5592818
20	189	130,0605355	148,6805434
21	250	165,9024558	189,6537422
22	175	211,6216478	241,9182841
23	368	269,9400777	308,5858233
24	349	344,3298279	393,6255198
25	345	439,2198127	502,1003499
26	475	560,2594612	640,4685386
27	427	714,6550652	816,9680601
28	627	911,5988173	1042,1071
29	793	1162,816083	1329,289676

(a) a

order (X)	New Deaths	e^γ2*x	y
30	651	0,294935084	1273,214393
31	601	0,283172224	1222,434939
32	743	0,271878501	1173,680715
33	683	0,261035204	1126,87095
34	712	0,250624369	1081,928093
35	919	0,240628748	1038,777687
36	889	0,231031781	997,3482432
37	756	0,221817568	957,5711247
38	812	0,212970845	919,3804323
39	837	0,204476955	882,7128946
40	727	0,196321825	847,5077638
41	760	0,188491946	813,7067149
42	766	0,180974345	781,2537492
43	681	0,173756567	750,0951011
44	525	0,166826655	720,1791496
45	636	0,160173128	691,4563324
46	604	0,153784962	663,8790638
47	542	0,147651576	637,401656
48	610	0,141762806	611,9802434
49	570	0,136108897	587,57271
50	619	0,130680483	564,1386192
51	431	0,125468569	541,6391474
52	566	0,120464521	520,0370193
53	602	0,115660049	499,2964462
54	578	0,111047194	479,3830669
55	525	0,106618312	460,2638905
56	575	0,102366067	441,9072419
57	482	0,098283413	424,2827092
58	433	0,094363588	407,3610936
59	454	0,090600097	391,1143608
60	534	0,086986704	375,5155944
61	437	0,083517424	360,5389517

(b) b

132	30	[illegible]	[illegible]
133	15	0,0044578	19,24401301
134	21	0,00428001	18,47650639
135	7	0,004109311	17,73961013
136	8	0,003945419	17,03210343
137	30	0,003788065	16,35281412
138	15	0,003636986	15,70061683
139	12	0,003491932	15,07443104
140	12	0,003352664	14,47321934
141	7	0,00321895	13,8959857
142	9	0,003090569	13,3417738
143	13	0,002967308	12,80966546
144	17	0,002848964	12,29877913
145	13	0,002735339	11,80826842
146	20	0,002626246	11,33732069
147	11	0,002521503	10,88515571
148	14	0,002420939	10,45102438
149	3	0,002324385	10,03420744
150	13	0,002231681	9,634014372
151	15	0,002142675	9,249782151
152	9	0,002057219	8,880874217
153	10	0,001975172	8,526679393
154	5	0,001896396	8,186610878
155	5	0,001820762	7,860105274
156	5	0,001748145	7,546621653
157	5	0,001678424	7,245640662
158	11	0,001611484	6,956663658
159	6	0,001547213	6,67921189
160	3	0,001485506	6,412825696
161	9	0,001426259	6,15706375
162	5	0,001369376	5,911502327
163	8	0,001314761	5,6757346
164	12	0,001262325	5,449369969
165	5	0,00121198	5,232033411

(c) c

Figure 4.31: Transformed model

- Table (a)is X=1 to 29
- Table (b and c) is X=30 to 169

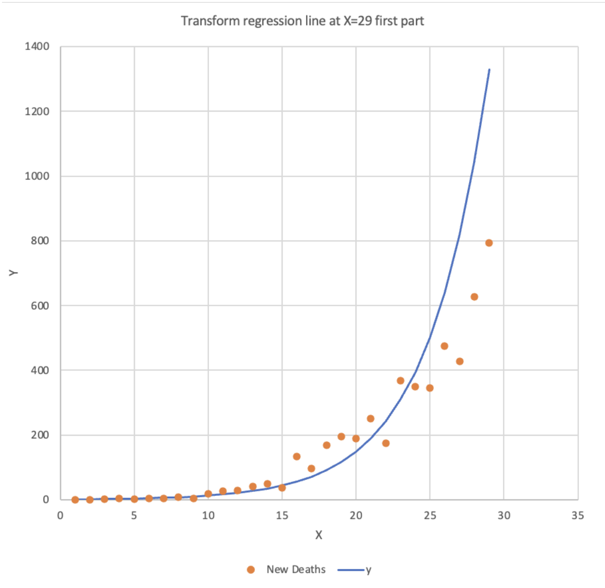


Figure 4.32: Optimized Non-linear Scattered plot A, x = 1 to 29

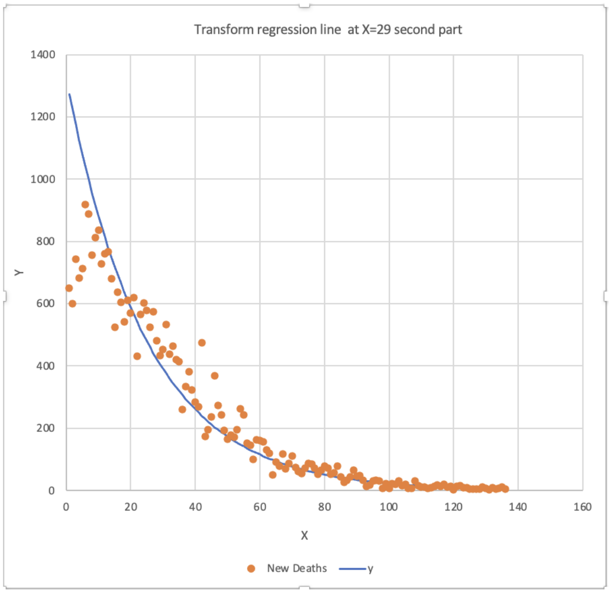


Figure 4.33: Optimized Non-linear Scattered plot B, x = 29 to 165

In Figure 4.34, it can be seen that the two modified Non-linear regression lines for each area of the data are meeting at the point x^o=29.

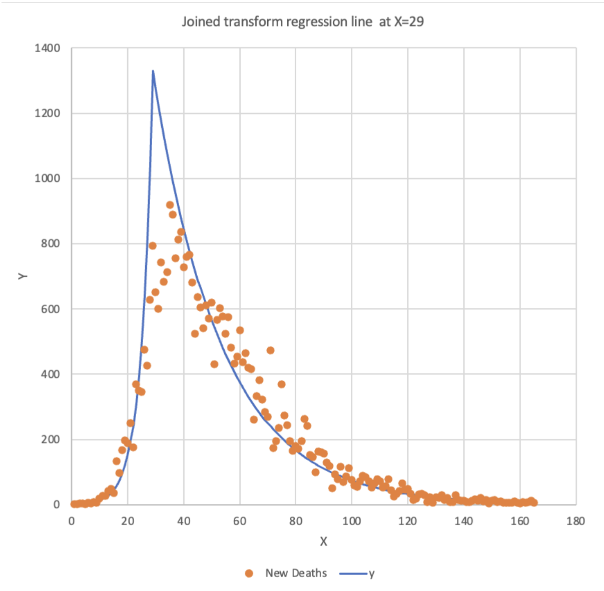


Figure 4.34: Joined Non-linear Scattered plot B, x = 29 to 165

4.3.1. MODIFIED INTERVAL ESTIMATE FOR NON-LINEAR REGRESSION LINES

The optimized confidence interval will be transformed in an exponential form. (i.e taking the exponential form of lower and upper optimized linear confidence interval for the mean value and also applying the same computation to lower and upper optimized linear confidence interval for the predicted value). The results of each transformation is given in the following tables

LC	UC	LC	UC
1,330877524	1,597694229	0,669243529	3,177222114
1,697639093	2,037984814	0,853672826	4,052797024
2,165472358	2,599610128	1,08892692	5,169661777
2,762230531	3,316007447	1,389012043	6,594310725
3,523442578	4,229828646	1,771794251	8,411562656
4,494428492	5,395479553	2,260063102	10,72961061
5,732997494	6,882359083	2,882888474	13,68646333
7,312889798	8,778991021	3,677351287	17,45816183
9,328166854	11,19829442	4,690751171	22,26926031
11,89881145	14,28430644	5,983422532	28,40619531
15,17787108	18,22075781	7,632326656	36,2343392
19,36056987	23,24201154	9,735633722	46,21975324
24,69593158	29,64701611	12,41856753	58,95693523
31,50160563	37,81710385	15,84086089	75,20421395
40,18277884	48,23869419	20,20626559	95,9288975
51,25629895	61,53225342	25,77468308	122,3648635
65,38144593	78,48923512	32,87763811	156,0860201
83,39918331	100,1192007	41,93801664	199,1000109
106,3822263	127,7099252	53,495243	253,9677438
135,6989076	162,9040673	68,23739539	323,9558581
173,0946435	207,7969671	87,04217176	413,2312096
220,7958495	265,0613963	111,029145	527,1089513
281,6424946	338,1066855	141,6264183	672,3689792
359,2571822	431,2817044	180,6556502	857,6595844
458,2608286	550,133779	230,4405091	1094,012344
584,5477766	701,7389601	293,9450173	1395,498902
745,6367243	895,123308	374,9500189	1780,06875
951,1183634	1141,800274	478,2782777	2270,617877
1213,226377	1456,45617	610,0816093	2896,351921

Figure 4.35: Non-linear confidence interval table for mean and predicted values from x=1 to x=29

LC	UC	LC	UC
1162,047154	1395,016446	584,345685	2774,171064
1115,701369	1339,379175	561,0402972	2663,529138
1071,203988	1285,960878	538,6643954	2557,29993
1028,481289	1234,673057	517,180909	2455,307448
987,4624956	1185,430742	496,5542459	2357,382721
948,0796492	1138,152352	476,7502334	2263,363513
910,2675041	1092,759561	457,7360619	2173,094062
873,9634162	1049,177165	439,4802304	2086,424816
839,1072398	1007,33296	421,9524939	2003,21219
805,641228	967,1576219	405,123814	1923,318322
773,5099371	928,5845924	388,9663103	1846,610852
742,6601346	891,5499663	373,4532144	1772,962696
713,0407109	855,9923876	358,5588253	1702,251841
684,6025952	821,8529475	344,2584674	1634,361138
657,2986733	789,0750866	330,5284489	1569,178112
631,0837104	757,6045011	317,3460231	1506,594773
605,9142757	727,3890531	304,6893503	1446,507437
581,7486705	698,3786842	292,5374621	1388,816557
558,5468592	670,5253322	280,8702262	1333,426556
536,2704029	643,7828521	269,6683132	1280,245667
514,8823958	618,1069391	258,9131647	1229,185785
494,3474039	593,4550554	248,5869625	1180,162318
474,6314067	569,7863598	238,672599	1133,094049
455,7017401	547,0616398	229,1536488	1087,902997
437,5270432	525,2432471	220,0143418	1044,514294
420,0772054	504,2950347	211,2395366	1002,856058
403,3233175	484,1822973	202,8146959	962,8592719
387,2376227	464,8717138	194,7258621	924,4576729
371,7934718	446,3312919	186,9596343	887,5876402
356,9652781	428,5303153	179,503146	852,1880905
342,7284754	411,439293	172,3440439	818,2003768
329,0594774	395,0299098	165,4704674	785,5681909
315,9356384	379,2749801	158,8710289	754,2374705

(a) a

178,705378	214,5325517	89,86357923	426,625793
171,5780932	205,9763761	86,27956105	409,6107285
164,735065	197,7614454	82,83848383	393,2742737
158,1649565	189,8741498	79,53464668	377,5893638
151,8568828	182,3014222	76,36257606	362,5300132
145,8003933	175,0307168	73,31701675	348,0712729
139,985454	168,0499881	70,3929231	334,1891887
134,4024312	161,3476709	67,58545073	320,8607621
129,0420754	154,9126614	64,88994844	308,0639114
123,895506	148,7342986	62,30195053	295,777436
118,9541966	142,8023467	59,81716943	283,9809805
114,2099609	137,1069781	57,43148856	272,6550017
109,654939	131,6387572	55,1409555	261,7807355
105,2815844	126,3886247	52,94177549	251,3401663
101,0826518	121,3478826	50,83030511	241,3159971
97,05118472	116,5081798	48,80304625	231,6916206
93,18050415	111,8614982	46,8566403	222,4510919
89,46419746	107,4001397	44,98786263	213,5791022
85,89610778	103,1167131	43,19361719	205,0609529
82,47032378	99,00412175	41,47093142	196,8825319
79,18116991	95,05555243	39,81695131	189,0302898
76,02319696	91,26446342	38,2289367	181,4912178
72,99117306	87,62457395	36,70425667	174,2528257
70,080075	84,12985375	35,24038528	167,3031215
67,28507991	80,77451304	33,83489729	160,6305915
64,60155727	77,55299298	32,4854642	154,2241812
62,02506124	74,45995642	31,18985039	148,073277
59,55132328	71,49027905	29,94590939	142,1676885
57,17624511	68,63904096	28,75158034	136,4976319
54,8958919	65,90151846	27,60488456	131,0537135
52,70648573	63,27317623	26,50392232	125,8269142
50,60439938	60,74965986	25,44686961	120,8085747
48,58615029	58,32678858	24,43197521	115,9903811
46,64839478	56,00054838	23,45755773	111,3643509

(b) b

18,29336667	21,96085352	9,198981156	43,67200443
17,56377455	21,08499148	8,832099312	41,9302392
16,86328067	20,24406133	8,479849772	40,25794058
16,19072451	19,4366699	8,141648957	38,65233805
15,54499181	18,66147957	7,816936564	37,11077156
14,9250128	17,91720605	7,505174635	35,63068718
14,32976033	17,20261631	7,20584667	34,20963283
13,75824825	16,51652646	6,918456766	32,84525421
13,20952971	15,85779985	6,642528798	31,53529094
12,68269565	15,22534515	6,377605631	30,27757277
12,17687324	14,61811456	6,123248363	29,07001603
11,69122449	14,03510208	5,879035595	27,91062012
11,22494481	13,47534182	5,644562735	26,79746426
10,7772617	12,9379064	5,419441327	25,72870426
10,34743349	12,42190546	5,20329841	24,70256949
9,93474806	11,92648412	4,995775893	23,71735992
9,538521714	11,4508216	4,796529972	22,77144334
9,158098015	10,99412988	4,605230552	21,86325264
8,792846708	10,55565234	4,421560703	20,9912832
8,442162675	10,13466254	4,245216137	20,15409041
8,10546493	9,730463032	4,075904699	19,35028728
7,78219566	9,342384161	3,913345888	18,57854213
7,471819299	8,969782992	3,757270391	17,8375764
7,173821641	8,61204223	3,607419634	17,12616252
6,887708986	8,2685692	3,463545357	16,44312186
6,613007328	7,938794862	3,325409199	15,78732284
6,349261563	7,622172874	3,19278231	15,15767897
6,096034738	7,318178682	3,065444962	14,5531471
5,852907328	7,026308654	2,943186194	13,97272571
5,619476538	6,746079242	2,825803458	13,41545319
5,39535564	6,477026186	2,713102283	12,88040631
5,180173329	6,21870374	2,604895955	12,36669863
4,973573107	5,970683937	2,501005207	11,8734791
4,775212697	5,732555877	2,401257922	11,39993058

(c) c

Figure 4.36: Non-linear confidence interval table for mean and predicted values from x=29 to x=165

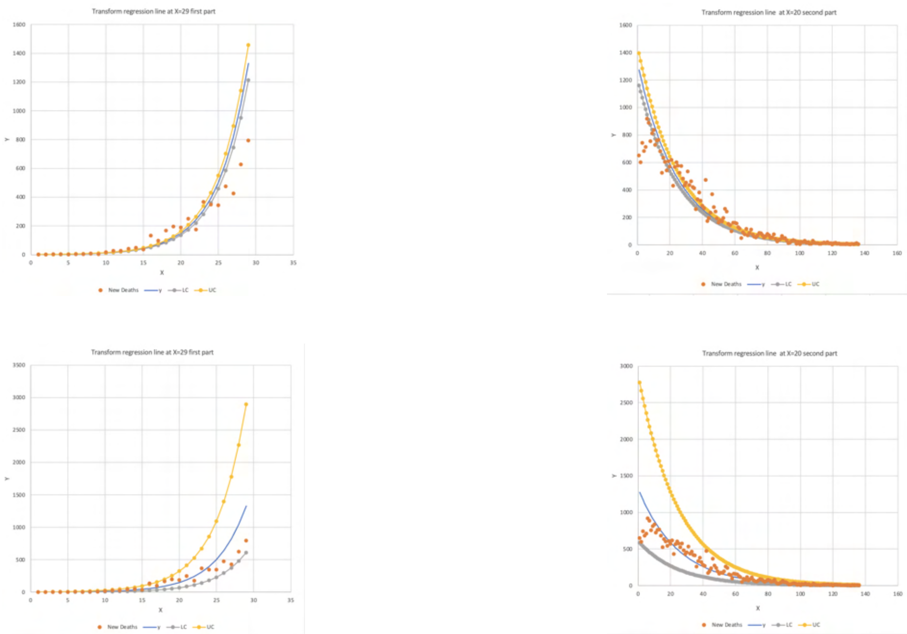


Figure 4.37: Mean and Predicted confidence Interval for Non-linear regression lines

- Joined confidence interval for Non-linear regression lines for mean value

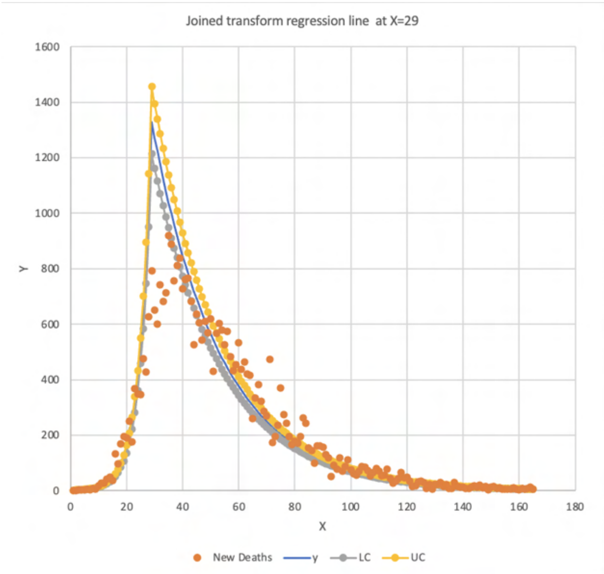


Figure 4.38: Scattered plot of the confidence interval of two Non-linear regression line for the mean value

- Joined confidence interval for Non-linear regression lines for predicted value

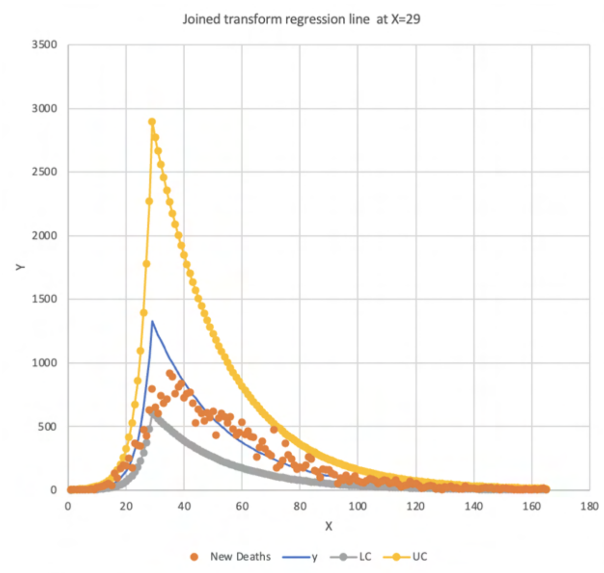


Figure 4.39: Scattered plot of the confidence interval of two Non-linear regression line for the predicted value

Below is the graph of the scattered plot and the joined optimized non-linear regression lines

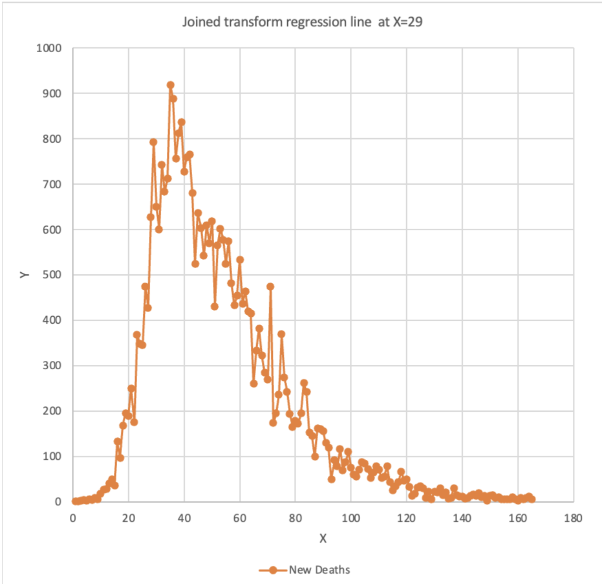


Figure 4.40: Scattered Non-linear regression plot

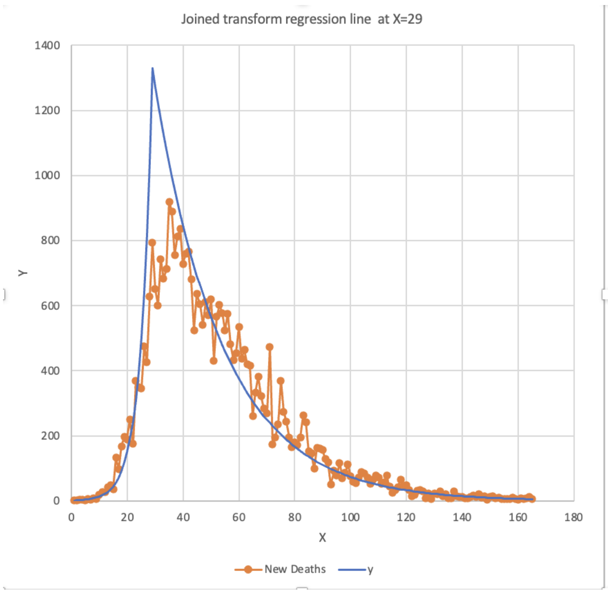
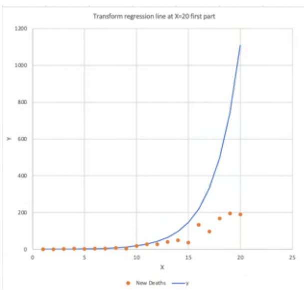
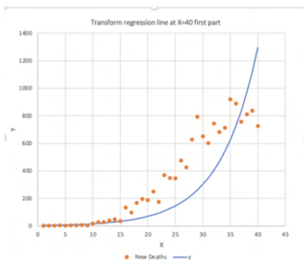


Figure 4.41: Scattered plot and the joined optimized Non-linear regression lines

Figure 4.20-4.28 shows the results of modified Non-linear regression lines and their corresponding confidence intervals for different values of x^o ranging from x^o=20 to x^o=100.

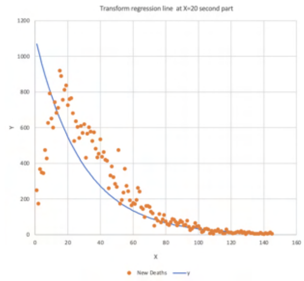


(a) At $X^0 = 20$



(b) At $X^0 = 40$

Figure 4.42: Non-linear regression line at different X^0 for first part

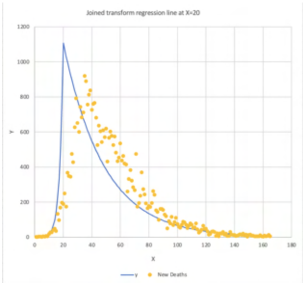


(a) At $X^0 = 20$

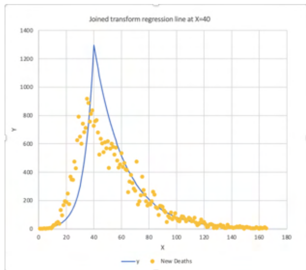
(b) At $X^0 = 40$

Figure 4.43: Non-linear regression line at different X^0 for second part

Below are the joined Non-linear regression line at $X^0 = 20, X^0 = 40$ and $X^0 = 100$

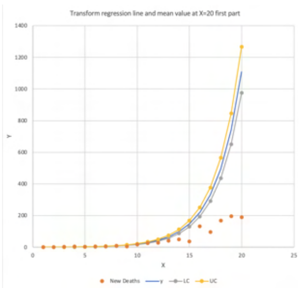


(a) $X^0 = 20$

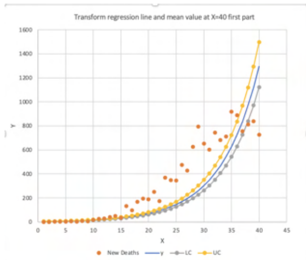


(b) $X^0 = 40$

Figure 4.44: Joined Non-linear regression line at different X^0

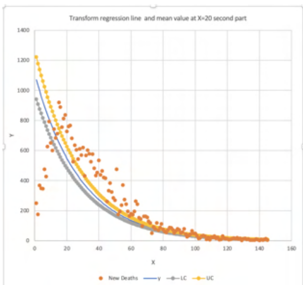


(a) At $X^0 = 20$

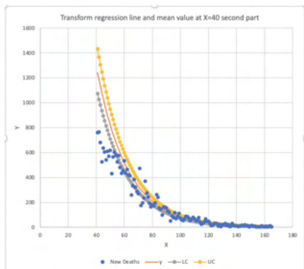


(b) At $X^0 = 40$

Figure 4.45: Mean Value Non-inear Confidence Interval at different X^0 for first part



(a) At $X^0 = 20$



(b) At $X^0 = 40$

Figure 4.46: Mean Value Non-linear Confidence Interval at different X^0 for second part

(a) $X^0 = 20$ (b) $X^0 = 40$

Figure 4.47: Joined Non-linear Confidence Interval for Mean Value at different X^0

(a) At $X^0 = 20$ (b) At $X^0 = 40$

Figure 4.48: Non-linear Confidence Interval for Predicted Value at different X^0 for first part

(a) At $X^0 = 20$ (b) At $X^0 = 40$

Figure 4.49: Predicted Value Non-linear Confidence Interval at different X^0 for second part

(a) $X^0 = 20$ (b) $X^0 = 40$

Figure 4.50: Joined Non-linear Confidence Interval for Predicted Value at different X^0

5. CONCLUSION

The starting point of our analysis is the following expression at which the data was examined using the linear and non-linear model because we encounter situations where it is not appropriate to use a single expression to describe the dependence between variables, so the data was divided into several sections and the expression of the dependence for each of them was obtained. Regression is used in a broader sense, but it is mainly based on quantifying the amount of change in the dependent variable (regression variable) due to the change in the independent variable using the data of the dependent variable. This is because all regression models, whether linear or nonlinear, simple or multiple, involve dependent and independent variables. We found the points at which the dependence changes and the expressions that describe these individual dependencies in the data and therefore, a segmented or break-point analysis is appropriate for the data as it's been analyzed in Chapter 3 & 4.

The data was then sectioned into two parts accordingly and regression lines were developed for each of the sections by minimizing the squared error of each of those regression lines under the condition that the two separated regression functions become equal at a certain arbitrary point i.e. the change point. For that the squared error function was modified using Lagrange Multiplier under the constraint that the two regression lines shall meet at the change point. It does not only minimized the squared error but also fulfils our required condition i.e. meeting of the two regression lines at the change point. These modified lines were plotted using the regression parameters calculated from the Lagrange multiplier function on a graph to show the complete relationship of the entities under study.

The data used for this research is a type of data that was observed over time, hence, it is safe to call it a time series data. Time series data are generally with auto correlation factor which is a disadvantage in change-point regression. We propose that in the future, further researches should be carried out using time series methods with the integration of change-point analysis. The integration of change-point analysis will help identify the break or change in relationship in the data, while the time series analysis will model the data with the inclusion of its auto-correlated factor for optimal relationship establishment.

Bibliography

[1] 13 types of regression analysis (plus when to use them) | indeed.com. https://www.indeed.com/career-advice/career-development/regression-types. (Accessed on 05/20/2022).

[2] 4. regression and prediction - practical statistics for data scientists [book]. https://www.oreilly.com/library/view/practical-statistics-for/9781491952955/ch04.html. (Accessed on 02/19/2022).

[3] (7) can anyone explain what is the difference between b and , in multiple regression? https://www.researchgate.net/post/can_anyone_explain_what_is_the_difference_between_B_and_b_in_multiple_regression. (Accessed on 04/15/2022).

[4] Chapter 7: Correlation and simple linear regression – natural resources biometrics. https://milnepublishing.geneseo.edu/natural-resources-biometrics/chapter/chapter-7-correlation-and-simple-linear-regression/. (Accessed on 05/16/2022).

[5] Chiara, b., marcello, a., mesfin, g., giovann, i., alessandro, l., adrian, m., et al. (2020, august 19). the covid-19 pandemic in italy: Policy and technology impact on health and non-health outcomes. retrieved may 7, 2021, from science direct: https://www.sciencedirect.com/science/article/pii/s2211883720300915 - google search. https://www.google.com/search?q=Chiara%2C+B.%2C+Marcello%2C+A.%2C+Mesfin%2C+G.%2C+Giovann%2C+i.%2C+Alessandro%2C+L.%2C+Adrian%2C+M.%2C+et+al.+%282020%2C+August+19%29.+The+COVID-19+pandemic+in+Italy%3A+Policy+and+technology+impact+on+health+and+non-health+outcomes.+Retrieved+May+7%2C+2021%2C+from+Science+Direct%3A+https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS2211883720300915&hl=en&sxsrf=ALeKk02EUx2y66IhZKjaDB5RqafrWgfaJQ%3A1621763449683&ei=eSWqYMGNKdf3kgXbsJ3gAQ&oq=Chiara%2C+B.%2C+Marcello%2C+A.%2C+Mesfin%2C+G.%2C+Giovann%2C+i.%2C+Alessandro%2C+L.%2C+Adrian%2C+M.%2C+et+al.+%282020%2C+August+19%29.+The+COVID-19+pandemic+in+Italy%3A+Policy+and+technology+impact+on+health+and+non-health+outcomes.+Retrieved+May+7%2C+2021%2C+from+Science+Direct%3A+https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS2211883720300915&gs_lcp=Cgdnd3Mtd2l6EAwyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwg
sclient=gws-wiz&ved=0ahUKEwiBrs_Sw9_wAhXXu6QKHVtYBxwQ4dUDCA4. (Accessed on 05/23/2021).

[6] The covid-19 pandemic in italy and the world: To be or not to be? that is the real problem | health security. https://www.liebertpub.com/doi/full/10.1089/hs.2020.0076. (Accessed on 05/23/2021).

[7] Estimating non-linear models with brms. https://cran.r-project.org/web/packages/brms/vignettes/brms_nonlinear.html. (Accessed on 05/16/2022).

[8] extension://elhekieabhbkpmcefcoobjddigjcaadp/https://eecs.wsu.edu/~cook/pubs/kais16.2.pdf. https://www.qo-goal.com/service/search?q=extension://elhekieabhbkpmcefcoobjddigjcaadp/https://eecs.wsu.edu/~cook/pubs/kais16.2.pdf&qoid=02a6553f0633839e13bdb704a4a22bc3&source=cd0a90ee81234c3e9c15ca1b66365d3f. (Accessed on 05/22/2022).

[9] intro-changepoint-course/introcptworkshop.pdf at master · rkillick/intro-changepoint-course · github. https://github.com/rkillick/intro-changepoint-course/blob/master/IntroCptWorkshop.pdf. (Accessed on 05/21/2022).

[10] Modern regression. https://www.stat.cmu.edu/~cshalizi/mreg/. (Accessed on 05/23/2022).

[11] Multiple linear regression — seaborn 0.11.2 documentation. https://seaborn.pydata.org/examples/multiple_regression.html. (Accessed on 05/16/2022).

[12] Non-linear regression in r - implementation, types and examples - techvidvan. https://techvidvan.com/tutorials/nonlinear-regression-in-r/. (Accessed on 02/18/2022).

[13] Praveen, d. (2020, march 23). coronavirus in italy: Outbreak, measures and impact. retrieved may 8, 2021, from phamaceutical technology: https://www.pharmaceutical-technology.com/features/covid-19-italy-coronavirus-deaths-measures-airports-tourism/ - google search. https://www.google.com/search?hl=en&q=Praveen,+D.+(2020,+March+23).+Coronavirus+in+Italy:+Outbreak,+measures+and+impact.+Retrieved+May+8,+2021,+from+Phamaceutical+Technology:+https://www.pharmaceutical-technology.com/features/covid-19-italy-coronavirus-deaths-measures-airports-tourism/. (Accessed on 05/23/2021).

[14] A refresher on regression analysis. https://hbr.org/2015/11/a-refresher-on-regression-analysis. (Accessed on 02/17/2022).

[15] regression - what is the difference between eta_1 and $hat{eta}_1$? - cross validated. https://stats.stackexchange.com/questions/210543/what-is-the-difference-between-beta-1-and-hat-beta-1. (Accessed on 04/15/2022).

[16] Regression analysis - wikipedia. https://en.wikipedia.org/wiki/Regression_analysis. (Accessed on 02/17/2022).

[17] Regression definition. https://www.investopedia.com/terms/r/regression.asp. (Accessed on 02/16/2022).

[18] Regression table - statistics solutions. https://www.statisticssolutions.com/regression-table/. (Accessed on 04/15/2022).

[19] (simple) linear regression and ols: Introduction to the theory | by maurizio sluijmers | towards data science. https://towardsdatascience.com/simple-linear-regression-and-ols-introduction-to-the-theory-1b48f7c69867#:~:text=Linear%20regression%20is%20used%20t,an%20assumption%20of%20the%20model.&text=The%20equation%20is%20called%20the%20regression%20equation. (Accessed on 02/20/2022).

[20] sylvester criterion. https://www.qo-goal.com/service/search?q=sylvester+criterion&qoid=02a6553f0633839e13bdb704a4a22bc3. (Accessed on 05/23/2022).

[21] What is a change-point analysis? https://variation.com/wp-content/change-point_analyzer_help/hs2.htm#:~:text=A%20change%2Dpoint%20analysis%20is,the%20time%20of%20each%20change. (Accessed on 05/22/2022).

[22] What is a change-point analysis? chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.epa.gov/sites/default/files/2016-07/documents/change-point-analysis.pdf. (Accessed on 05/22/2022).

[23] What is regression analysis and why should i use it? | alchemer blog. https://www.alchemer.com/resources/blog/regression-analysis/. (Accessed on 05/16/2022).

[24] What is regression analysis? types and applications | analytics steps. https://www.analyticssteps.com/blogs/what-regression-analysis-types-and-applications. (Accessed on 05/16/2022).

[25] What is regression and how it works | definition of regression. https://www.mygreatlearning.com/blog/what-is-regression/. (Accessed on 02/20/2022).

[26] What is the difference between b and beta in regression? https://philosophy-question.com/library/lecture/read/69919-what-is-the-difference-between-b-and-beta-in-regression. (Accessed on 04/15/2022).

[27] Donald W. K. Andrews. Tests for parameter instability and structural change with unknown change point. Econometrica, 61(4):821–856, 1993.

[28] Jushan Bai. Estimation of a change point in multiple regression models. The Review of Economics and Statistics, 79(4):51–563, 1997.

[29] Samuel L Baker. Non-linear regression course. University of South Carolina, 2008.

[30] P. K. Bhattacharya. Some aspects of change-point analysis. IMS Lecture Notes-Monograph Series-Change-Point Problems, 23:28–56, 1994.

[31] Jie Chen. Testing for a change point in linear regression models. Communications in Statistics - Theory and Methods, 27(10):2481–2493, 1998.

[32] Carl Friedrich Gauss. Theoria combinationis observationum erroribus minimis obnoxiae, Volume 2. H. Dieterich, 1821.

[33] William H Greene. Econometric analysis. Pearson Education India, 2003.

[34] Kai Sheng Song Hans Georg Müller. Two-stage change-point estimators in smooth regression models. Statistics Probability Letters, 34(4):323–335, 1997.

[35] Steven A. Julious. Inference and estimation in a changepoint regression problem. Journal of the Royal Statistical Society: Series D (The Statistician), 50(1):51–61, 2001.

[36] Michael McAleer L. G. Godfrey and C. R. McKenzie. Variable addition and lagrange multiplier tests for linear and logarithmic regression models. The Review of Economics and Statistics, 70(3):492–503, 1988.

[37] Adrien Marie Legendre. Nouvelles méthodes pour la détermination des orbites des comètes. F.Didot, 1805.

[38] Zhaoyuan Li and Maozai Tian. Detecting change-point via saddlepoint approximations. Journal of Systems Science and Information, 5(1):48–73, 2017.

[39] John F Monahan. A primer on linear models. CRC Press, 2008.

[40] IIT Kanpur Shalabh. Simple linear regression. In Regression Analysis.

[41] Helge Toutenburg, S Shalabh, and Heumann Shalabh. Statistical analysis of designed experiments. 2002.