

Multiple Regression(with Polynomial Regression)

Muhammad Moaz Amin
dept. Electronic Engineering
Hochschule Hamm Lippstadt
Hamm, Germany
muhammad-moaz.amin@stud.hshl.de

Abstract—Estimating the relationship among different variable which have reason and result relation holds great importance, to estimate that Regression analysis is a statistical technique that is adapted. Main focus of this uni-variate regression is to analyse the relationship between a non linear and a linear variable and to formulate a linear equation between the two. a regression model which contain one linear and multiple non-linear independent variables is most often called multi linear regression. This paper is concentrated on the polynomial regression model, which is useful when there is reason to believe that relationship between two variables is curvilinear. The polynomial regression model has been applied using the characterisation of the connection between strains and drilling depth. Parameters of the model were estimated employing a least square method. After fitting, the model was evaluated using a number of the common indicators wont to evaluate accuracy of regression model.

I. INTRODUCTION

Regression analysis involves identifying the link between a variable quantity and one or more independent variables. It's one amongst the foremost important statistical tools which is extensively utilized in the majority sciences. It's specially used in business and economics to check the link between two or more variables that are related causally. A model of the relationship is hypothesized, and estimates of the parameter values are accustomed develop an estimated equation.

Common questions which are generally asked in this research of Multiple regression analysis generally revolve around "are there any relations between dependent and independent variable?", and "if there are some relations that exist, what is total power of the relation?," is there any possibility to predict about orientation regarding the dependent variable?, and "if certain conditions are controlled, what influences does a special variable or a group of variables have over another variable or variables?". (Alpar, 2003).

Uni-variate analysis is the regression which uses a single independent variable while multivariate regression analysis uses more than one independent variable (Tabachnick, 1996, Buyukozturk, 2002). The relation between dependent variable and an independent variable is analysed through uni-variate regression analysis, and the equation which represents the linear relationship between the both is formulated.

In Multivariate regression analysis, an attempt is made to account for the variation between the dependent variable and the independent variable synchronically (Unver & Gamgan,

1999). Formulation of Multivariate model is represented in Figure 1.

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

y - dependent variable
 X_i - independent variable
 β_i - parameter
 ε - error

Fig. 1. Multivariate Model

The assumptions of multivariate regression analysis are normal distribution, linearity, freedom from extreme values and having no multiple ties between independent variable (Buyukozturk, 2002).

II. MULTIPLE REGRESSION WITH PYTHON

Linear Regression generally is a supervised method for machine learning rooted in statistics. From this method numeric values are forecasted using a combination of predictors which can be both numeric or binary variables. The condition for getting the variable depends on a certain relation at hand (a linear, measurable by a correlation) with the target variable.

However, in reality there are a number of factors which alter the results of the working predictive model. There are usually more than variables that work together to achieve better and reliable results from a prediction. This causes more complexity in our model and hence representing it on a two- dimensional plot is not easy. All the predictors will constitute their own unique dimension and we would have to assume that our predictors apart from being related to the response are also related among themselves and this characteristic of data is called multicollinearity.

A. Multiple Regression Formulation

The basic Multiple Regression is described in Figure 1 and in depth detail is shown in Figure 2 where dependent(response) variable Y on a set of k independent(predictor) variables X_1, X_2, \dots, X_k can be expressed as

where

- y_i = value of dependent variable, Y is for i th case.
- x_{ij} = value of j th independent variable, X_j for i th case.

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{1k} + e_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_k x_{2k} + e_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_k x_{nk} + e_n \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i, \text{ for } i=1,2,\dots,n$$

Fig. 2. In depth Multiple Regression Model showing relation of independent(k) and dependent variable(Y)

- β_0 is the Y -intercept of the regression surface.
- each $\beta_j, j=1,2,\dots,k$, is the slope of the regression surface w.r.t. variable X_j and e_i is the random error component for the i th case.

In the first equation we have n observations and k predictors ($n > k+1$) The assumptions of the multiple regression model are similar to those for the simple linear regression model. Model assumptions [1]:

B. Observations

- errors e_i are normally distributed with their mean as zero and their standard deviation σ are independent of the error terms associated with all other observations. Errors are not related to each other.
- variables X_j in the context of regression analysis are considered as fixed quantities, whereas they are random variables in the context of correlation analysis. But in both the cases X_j are totally independent of the error term. If X_j are assumed as fixed quantities, then we are assuming that we have realizations of k variables X_j and the only randomness in Y is coming from the error term.

In matrix notation, we can rewrite the regression model as described in Figure. 3.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Fig. 3. Matrix notation of Model

- response vector \mathbf{Y} and error vector \mathbf{e} are the column vectors of length n .
- vector of parameters $\boldsymbol{\beta}$ is the column vector of length $k+1$ and its design matrix (where all elements in the first column are equal to 1, and second column's values are filled by the observed values of X_1 ,etc).

Here, values of $\boldsymbol{\beta}$ and \mathbf{e} are unknown and assumed.

III. POLYNOMIAL REGRESSION

Polynomial Regression is a technique of Multiple Regression in which we regress a dependent variable on the powers of independent variable.

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

IV. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections IV-A–IV-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads— \LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”.)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. *LaTeX-Specific Advice*

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in *LaTeX* will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

BibTeX does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use *BibTeX* to produce a bibliography you must send the .bib files.

LaTeX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. *Some Common Mistakes*

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).

- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [?].

F. *Authors and Affiliations*

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. *Identify the Headings*

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. *Figures and Tables*

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	<i>Table column subhead</i>	<i>Subhead</i>	<i>Subhead</i>
copy	More table copy ^a		

^aSample of a Table footnote.



Fig. 4. Example of a figure caption.

figures and tables after they are cited in the text. Use the abbreviation “Fig. 4”, even at the beginning of a sentence.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

REFERENCES

- [1] Aczel, A. D., 1989. Complete Business Statistics. Irwin, p. 1056. ISBN 0-256-05710-8.
- [2] L. Massaron and A. Boschetti, Regression analysis with Python: learn the art of regression analysis with Python. Birmingham Mumbai: Packt Publishing, 2016.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.