Multiple Regression(with Polynomial Regression)

Muhammad Moaz Amin
dept. Electronic Engineering
Hocschule 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 (Tabachinick, 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.

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

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,.....X_k can be expressed as

where

- y_i= value of dependent variable, Y is for ith case.
- x_{ij} = value of jth independent variable,X_j for ith case.

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

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

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 β_j, j=1,2,....k, is the slope of the regression surface w.r.t. variable X_j and e_i is the random error component for the ith 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 **Y** and error vector **e** are the column vectors of length n.
- vector of parameters $beta$ 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 β and 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. *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/m^2" or "webers per square meter", not "webers/m^2". 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 "cm^3", not "cc".)

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

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 4. Example of a figure caption.

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.