Journal of Nonlinear Modeling and Analysis
Volume 7, Number 1, February 2025, 20–42

https://journal.global-sci.org/jnma
DOI:10.12150/jnma.2025.20

A Study on Demonstrating the Consistency of the Formulation of the Parameters of 5 or Less Independent Variables of Multiple Linear Regression with Data Examples

Mehmet Korkmaz[1,†]

Abstract In this study, it is aimed to demonstrate the consistency of the previously given formulations of the parameters of 5 or less independent variables of multiple linear regression with data examples. All the values in the parameter formulas of 5 or less independent variables with respect to one dependent variable were found to be the same as all the parameter values obtained from the related equation system. In other words, the values found with the parameter formulas and the parameter values found with the computer program were found to be the same.

Keywords Linear regression, multiple linear regression, formulas of parameters, 5 or less independent variables

MSC(2010) 62J05, 62J12

1. Introduction

In mathematics, statistics and many sciences, regression is one of the important topics. In this study, we will first work with the example given for simple linear regression. After that we will work with the example given for multiple linear regression, MLR.

Regression analysis is an important statistical tool for analyzing the relationships between dependent, and independent variables. The main goal of regression analysis is to determine and estimate parameters of a function that describes the best fit for given data sets. There are many linear types of regression analysis models such as simple and multiple regression models. Regression analysis is the widely used statistical tool for understanding relationships among variables. It is used when there is a continuous dependent variable which can be predicted by independent variables [1]

Seber defined linear regression analysis as a common technique of estimating the relationship between any two random variables, the explanatory variable X, and the dependent variable Y [11]. The simple relationship among dependent and explanatory variables can be defined as follows:

$$Y = \mathrm{f}(X_1, X_2, \ldots, X_N) + \varepsilon,$$

[†] the corresponding author.
Email address:mkorkmaz52@yahoo.com (M. Korkmaz)
[1] Department of Mathematics, Faculty of Arts and Sciences, Ordu University,
52200, Ordu, Turkey

where a random error representing the discrepancy in the approximation is assumed to be ε. It accounts for the failure of the model to fit the data exactly. The function f(X_1, X_2, ... X_N) describes the relationship between the dependent variable Y, and the explanatory variables X_1, X_2, ... X_N.

When the relationship is linear, it may be represented mathematically using a straight line equation. The regression coefficient describes the change in Y that is associated with a unit change in X. This line is frequently computed using the least square procedure [5].

Linear regression is one of the fundamental techniques in the statistical analysis of data. We assume a straight-line model for a response variable Y as a function of one or more predictor (or explanatory) variables X [4]. In this study, first, we look at exactly one predictor variable and then we will look at two or more predictor variables.

Multiple linear regression is as follows when the error is omitted

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + ... + \hat{\beta}_N X_N,$$

where $\hat{Y}$ is the equation of the regression and $\hat{\beta}_i$ ($i = 0, 1, 2, ..., N$) are unknown parameters of the regression, X_i ($i = 1, 2, ..., N$) are the independent variables of the regression.

The unknown parameters, $\hat{\beta}$, are estimated by using the method of least squares in multiple linear regression. The estimates of the $\hat{\beta}$ coefficients are the values that minimize the sum of squared errors for the sample. The obtained formula for this is given in this study on matrix notation.

Equations like this can easily be handled by any computer program that performs ordinary multiple regression. But in this study to get the parameters of multiple linear regression without using a computer program, the general formula of the parameters will be given. After that the parameter values obtained by the formula will be compared with the parameter values obtained by the linear equation system.

Multiple regression analysis is one of the most widely used statistical procedures. Multiple linear regression using the least square procedure is extensively used in astronomy to model observational data, analyze simulated data, and compare empirical data with theoretical models [3]. Its popularity is fostered by its applicability to varied types of data and problems, ease of interpretation, robustness to violations of the underlying assumptions, and widespread availability [7].

If a simple linear regression model with one predictor variable, X_1, is started, then a second predictor variable, X_2, is added, Error sum of squares (SSE) will decrease (or stay the same) while Total sum of squares (SST) remains constant, and thus R-squared (R^2) will increase (or stay the same). In other words, R^2 always increases (or stays the same) as more predictors are added to a multiple linear regression model, even if the predictors added are unrelated to the response variable. Thus, by itself, R^2 cannot be used to help us identify which predictors should be included or excluded in a model. But an alternative measure, adjusted R^2, does not necessarily increase as more predictors are added, and can be used to help us identify which predictors should be included or excluded in a model. Due to the malfunctioning of R^2, the researchers preferred to use adjusted R^2 [2].

In fact, the adjusted R^2 statistic does not change by adding variables to the model. In addition, the adjusted R^2 will often decrease by adding excessive parameters. This is the best way to add unnecessary variables to the model without changing the R^2 significantly [15].