Estimation Parameters Of The Multiple Regression Using Bayesian Approach Based On The Normal Conjugate Function

Adel Abbood Najm 1, Ahmed Hamza Abood 2, Shrook A S AL-Sabbah 3

1 College of Administration and Economics/ University of Sumer/ Iraq
adel.abood6969@gmail.com

2 Department of Anatomy, Histology and Embryology
Factuality of Medicine, Karbala University, Iraq.
dr.ahmedalbaghdadi@gmail.com

3 Prof. Dr. Department of Statistics, Administration and Economic College, Karbala University, Iraq
shrookalsabah1968@gmail.com

Abstract: In this paper, we have been used Bayes Technique depending on the normal conjugate function to estimate parameters of the multiple regression model, and we have been tested significance of this model. The test showed in the application that the mean square error (MSE) for the used model was decreasing, also it showed that the determinant coefficient is increasing highly. In the same time, value of the computed F-test was significant according to the above we can consider that the model is significant

1- Introduction:

The regression analysis is considered one of the most important statistical techniques which are used by the researchers to analyze the data in their fields such as the industry, biology, social, production, etc. for the sake of reaching to the best results, and this issue is done by forming a correct formula for the relationship between the different phenomena, which are represented by the variables, and these variables are subjected to the regression formula in its different forms. The regression model formulas is very useful in case of knowing direction of the explanatory variables which are dealing with it by the researcher, also knowing the effect range which is showed by these variables on the response variables, besides the interpretation ratio of the regression model contribution in explaining the relationship between the response variable and the explanatory variables and all of that is done by process of estimation parameters of the model.[3],[8]

Our goal in this paper is to estimate the multiple linear regression model by using Baysian approach based on the normal conjugate prior function, and then showing the model significance using F-test and we have been showed that by using an applicable real data.
2- Concept of the regression analysis Bayes Approach:

Regression analysis is dealing with studing and estimation a phenomenon by quantity way through collection and analysis the data and determination the relationship figure between these data, where the estimation and prediction for future for this phenomenon are achieved according to certain statistical methods after getting an equation or a curve explain the mathematical relationship between the response variables Y and an explanatory variable \( X_i \) and this so- called the simple linear regression.[7],[11]

If the relationship between the response (dependent) variable Y with several explanatory variables \( (X_1, X_2, ..., X_n) \), in this case we call it multiple linear regression, and we can express the relationship between the dependent variable and one variable or more than one variable mathematically by the following regression equation:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in} \quad \cdots (1)
\]

Note that eq.(1) shows that the dependent variable Y is explained by several independent (explanatory) variables, but if we want to explain this relationship correctly, it must be adding the error term to eq.(1), where this term represents or denotes to the information which the model doesn't involve it:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in} + U_i \quad \cdots (2)
\]

an easy way for representing the multiple regression model in (2) is to re-write it using the matrices form and as follows:

\[
Y = X\beta + U \quad \cdots (3)
\]

Where \( Y \): is a vector of \( n \times 1 \) for the observations of response variable.

\( \beta \): is the parameters vector of \( n \times p \) (which must be estimated).

\( X \): a matrix of order \( n \times p \) for observations of explanatory variables.

\( U \): error term of \( n \times 1 \).

To estimate the regression model, it must be and first of all, the model should have some characteristics depending on a certain assumptions to increase sobriety of the regression model, and if these assumptions is not available, then the model will suffer from the problems which are making the estimation process very weak and imprecise, hence the estimated parameters will be inefficient.[4],[3]

So, these assumptions are:

1) \( E(u_i) = 0 \)
2) \( V(u_i) = \sigma_u^2 \)
3) \( u_i \sim N(0, \sigma_u^2) \)
4) \( \text{cov}(u_i, u_j) = 0 \quad \forall \ i \neq j , i = 1,2, ..., n \)

2.1 Bayes Approach:

The idea of Bayesian approach is dealing with the unknow parameters as a random variable having a distribution function and these parameters have a prior information differ quantively and qualitively depending on size of the available information which the researcher had through the experiences, the pervious experiments or which to be identical or close to the work. The difficulty of this approach lies in collection the information about the unknown parameters and determine its prior probability distribution precisely because the difficulty of getting the prior or previous information or lack of accuracy for these information, however, these information are formulated in a prior probability distribution, which so- called prior probability density function \( P(\theta) \) or prior p.d.f where this function is the difference point between Bayes approach and the other classical approaches.[8],[12]
The prior p.d.f is combined with the maximum likelihood $P(Y|\theta)$ for the current observations of $Y$ by using Bayes Inversion Formula to get a good information or close to the actual information about the unknown parameter. All of these information are putting in a probability distribution form which is called the posterior probability density function $P(\theta|Y)$ or (posterior p.d.f) where this distribution is supposed to be a good description about the unknown parameter together with existence of the sample information.

$$P(\theta|Y) = \frac{\prod_{i=1}^{n} P(Y_{i}|\theta) P(\theta)}{\int_{\theta} \prod_{i=1}^{n} P(Y_{i}|\theta) P(\theta) d\theta} \ldots \ldots (4)$$

Bayes estimation approach needs to use so-called the loss function where the Bayes estimator yields by minimizing the expected loss function for the posterior distribution of the unknown parameter $\theta$, given that the sample data $Y$ is known, and the loss function must satisfying the following two conditions:

1. $L(\hat{\theta}, \theta) \geq 0 \quad \forall \hat{\theta}, \theta$
2. $L(\hat{\theta}, \theta) = 0 \quad \forall \hat{\theta} = \theta$

There are different kinds for the loss function and according to this difference and to the kinds of the prior distribution of the parameter, Bayes estimators will be different too, but our goal is to get a Bayesian estimator $\hat{\theta}_{Bayes}$, which at it the posterior expected loss will be as small as possible, so we will adopt the weighted squared Error Loss Function.[5],[10]

2.2 Prior Probability Density Functions:

In Bayes Technique, we must have an information about the unknown parameters where these information is considered the main support point for this technique. We have two ways to get these information, the first one is the scientific way because these information are came from the same used data attribute or quality and in this case we call it Data Based on prior p.d.f, the second one is achieved by collecting the information using the causal observations and the theoretical assumptions, in other words, not from the used or real data, or these information is coming from sources have not any relationship with the data, here we call it Non data Based on prior p.d.f. [6]

Sometimes we get the information by using a mixture between the two ways. However, these information which we get it and whatever its source have a remarkable rule in choosing the prior pdf.

There are four priors p.d.f, uninformative prior, information prior, prior p.d.f depending on previous sample, and a normal conjugate prior p.d.f. In our paper, we depend on the last one.[1]

2.3 Normal Conjugate Prior p.d.f :

This function has a good properities comparing with the other functions mentioned above which is making it more widely using because it considered a p.d.f of well-known parameters besides it is being explicit, determined and proper, note that the prior p.d.f is built on the likelihood function of the current sample observations by considering it as a function in the parameter $\theta$. And it is worthy to mention that the prior p.d.f $P(\theta)$, likelihood function $P(Y|\theta)$, and the posterior p.d.f $P(\theta|Y)$ are characterized that it have the same functional formulas, but with new different parameters.[9]

This type of functions is preffered to use instead of the uninformative prior p.d.f because it is being improper function, also we preffer it because it treats the multicollinearity problem. One of the famous normal conjugate prior p.d.f is the prior joint p.d.f of Normal – Gamma.

3. Methodology:

Here, we have two unknown parameters ($\beta, \sigma$) and each parameter has a certain distribution according to a certain conditions and rules, that is when we do not have information about these parameters, then
we follow what Jeffery is said in determination the prior p.d.f. He said if interval of the parameter which we want to estimate was in \((-\infty, \infty)\), then the prior p.d.f. will be a uniform distribution:

\[
P(\beta) \, d\beta \propto d\beta, \quad -\infty < \beta < \infty \quad \text{or} \quad P(\beta) \, d\beta \propto \text{constant}, \quad -\infty < \beta < \infty \ldots (5)
\]

and if the parameter interval was in \((0, \infty)\) then the prior p.d.f will be a log – uniform distribution:

\[
P(\sigma) \, d\sigma \propto \frac{1}{\sigma} \, d\sigma \ldots (6)
\]

As a result, the vector \((\beta)\) has a multivariate normal distribution given by the following p.d.f. \(P\left(\frac{\beta}{\sigma}\right) = \frac{1}{\sigma^m} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} \beta - \beta_p \end{bmatrix} \right)^T Q \left( \begin{bmatrix} \beta - \beta_p \end{bmatrix} \right) \right] \ldots (7)\)

Where \(-\infty < \beta < \infty\)

\(\beta\): denotes to mean of the prior distribution.

\(\sigma^2 Q^{-1}\): denotes to variance – covariance matrix and furthermore, \(Q\) is a positive definite matrix. To get the joint p.d.f. of \((\beta, \sigma)\), we use the following equation:

\[
P(\beta, \sigma) = P(\sigma)P\left(\frac{\beta}{\sigma}\right) \ldots (8)
\]

\[
P(\beta, \sigma) \propto \frac{1}{\sigma^n} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right)^T Q \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right) \right] \ldots (9)
\]

Where \(P\) = number of parameters.

The likelihood function for the observations is given by:

\[
P(Y/\beta, \sigma) \propto \frac{1}{\sigma^n} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right)^T Q \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right) \right] \ldots (10)
\]

combining Equation (9) which represents the prior p.d.f for the two parameters with eq(10) yields:

\[
P(\beta, \sigma/Y) \propto \frac{1}{\sigma^{n+m+1}} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right)^T Q \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right) \right] \ldots (11)
\]

or

\[
\frac{1}{\sigma^{n+m+1}} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} Y - X \beta \end{bmatrix} \right)^T \left( \begin{bmatrix} \sigma^2 Q^2 \beta_p - Q^2 \beta \end{bmatrix} \right) \right] \ldots (11)
\]

Now, by letting:

\[
w = \left[ Y Q^2 \beta_p \right], Z = \left[ X Q^2 \right]
\]

then, the posterior distribution formula becomes:

\[
P(\beta, \sigma/Y) \propto \frac{1}{\sigma^{n+m+1}} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} Y - Z \beta \end{bmatrix} \right)^T \left( \begin{bmatrix} Y - Z \beta \end{bmatrix} \right) \right] \ldots (12)
\]

and by letting: \(\beta_{BC} = (ZZ)^{-1}ZW \ldots (13)\)

where: \(\beta_{BC}\) = Bayes estimator based on a normal conjugate prior function then, the posterior p.d.f. becomes:

\[
P(\beta, \sigma/Y) \propto \frac{1}{\sigma^{n+m+1}} \exp \left[ - \frac{1}{2\sigma^2} \left( \begin{bmatrix} W - Z \beta_{BC} \end{bmatrix} \right)^T \left( \begin{bmatrix} W - Z \beta_{BC} \end{bmatrix} \right) \right] \ldots (14)
\]

Now, integrate eq.(14) with respect to \(\sigma\), then we'll get the marginal p.d.f for the parameters vector \((\beta)\).

\[
P(\beta/Y) \propto \left( \begin{bmatrix} \beta - \beta_{BC} \end{bmatrix} \right)^T Z^T Z \left( \begin{bmatrix} \beta - \beta_{BC} \end{bmatrix} \right)^{-1} \ldots (15)
\]
The formula (15) represents a p.d.f of m-variate t distribution with mean equals to \( \beta \) given by the formula (16) which represents Bayes estimator for the parameters vector \( \beta \) based on a normal conjugate prior function, that is, Bayes estimator is given by:

\[
\beta_{BE} = \left( X Q^T \right) \left( X Q^T \right)^{-1} \left( Y Q^T \beta_{P} \right) \quad \ldots (16)
\]

4. The Practical side and results:

In this section we present an actual experiment about one of electrical energy production stations in Iraq – Karbala station – here, we use eight variable one of them is the response (dependent) variable \( Y \) and the remaining variables represent the explanatory (independent) variables for (30) months in accordance with the following multiple linear regression model:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + u_i
\]

Where:

- \( Y_i \): Electrical energy production (mega watt)
- \( X_{i1} \): Heavy fuel (black oil) (litre)
- \( X_{i2} \): Light fuel (kerosene) (litre)
- \( X_{i3} \): Lubricating oil (litre)
- \( X_{i4} \): Chimirical materials (litre)
- \( X_{i5} \): Operating hours average (hour/month)
- \( X_{i6} \): Overtime wages (in Dinar)
- \( X_{i7} \): Monthly temperature average.

The statistical program (MATLAB 2105a) has been used to estimate parameters of the previous model.

Table (1) shows estimation of multiple model parameters based on Bayes approach

| Parameters | \( \beta_0 \) | \( \beta_1 \) | \( \beta_2 \) | \( \beta_3 \) | \( \beta_4 \) | \( \beta_5 \) | \( \beta_6 \) | \( \beta_7 \) | MSE | \( R^2 \) |
|------------|------|------|------|------|------|------|------|------|-----|------|
| Bases NC  | 0.976 | 0.001 | 0.225 | 0.098 | 0.182 | 0.174 | 0.311 | 0.47 | 0.003 | 0.001 | 0.976 |

From table (1), we find that mean square error (MSE) is decreasing value and very low, while value of the adjusted coefficient of determination \( (R^2) \) seems to be very high.

Table (2) represent ANOVA for regression model

| Source       | Sum of squares | d.f. | MS     | \( F_c \) | \( F_{table} \) |
|--------------|----------------|-----|--------|----------|-----------------|
| Regression   | 43143.0581     | 7   | 6163.2940 | 3.4923  | 2.4638          |
| Residuals    | 39485.2401     | 22  | 1764.7836 |          |                 |
| Total        | 82628.2982     | 29  |         |          |                 |

By comparing the computed \( F_c \) with value of the tabulated \( F(7,22,0.05) = 2.46 \), we see that \( F_c > F_{table} \), and that means there is at least one explanatory variable affects on the response variable \( Y \).
5. Conclusions:

1. The results have been showed that mean square error (MSE) was very low with high value of coefficient of determination $R^2$, and this is a good indicator.

2. The concluded results showed that the model is significance according to F-test, and that means there is a strong relationship between the dependent variable $Y$ and the other independent variables.

Reference.

[1] Baldi P. and Long A., "A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistic inferences of gene changes" Bioinformatics, 17(6): 509-519,2001.

[2] Bishop C., 2006, "Pattern recognition and machine learning" Springer.

[3] Christopher M. Strichland and Clair L.Alston (2013), "Bayesian analysis of the Normal linear Regression model", First Edition, university of Technology Brisbane, Australia, John Wiley & Sons, Ltd.

[4] Demichelis F., P. Magni, p. piergiorgi, M. Rubin, and R. Bellazi, 2016, "A hierarchical Naïve Bayes models for handling sample heterogeneity in classification problems: an application to tissue microarrays". BMS Bioinformatics, 7:514.

[5] Gordon, N., J., D. J. Salmond, and A.F.M. Smith. "Novel Approach to Nonlinear Non-Gaussian Bayesian state Estimation" IEEE Proceedings F on Radar and Signal Processing. Vol.140(2), 1993, pp. 107-113.

[6] Iswari A A I A C, IW Sumerjaya and IGAM Srisnadi (2014), Analisis Regresi Bayes Linear Sederhana dengan prior Noninformatif E- Journal Mathematika, 3(2) 38-44.

[7] Kevin P, Murphy, (2007),"Conjugate Bayesian analysis of the Gaussian distribution", murphy K.

[8] Park, T., and G.Casella. "The Bayesian Lasso". Journal of the American statistical Association. Vol.103, No, 482, 2008, pp.681-686.

[9] Raftery, E., Maigan, D. & Hoeting, J.A. (1997), " Bayesian model averaging for linear regression model", Jassa. Vol. 75, No, 327, p. 801-816.

[10] Michael J. zyphur and Frederick L. Oswald, " Bayesian Estimation and inference: A user's Guide", Journal of Management, 41(2): 390-420, 2013.

[11] Sinay M. S. and J. S. J. Hsu (2014), Baysian Inferensi of a multivariate regression model (Hindawi Publishing Corporation).

[12] Williams P. M. (1995), Bayesian regularisation and Pruning using a Laplace prior. Neural Computation, 7(1), 117-143.