HEBRON UNIVERSITY
FACULTY OF GRADUATE STUDIES AND SCIENTIFIC
RESEARCH PROGRAM OF MATHEMATICS
MASTER PROGRAMME IN MATH

BAYESIAN AND NON BAYESIAN ANALYSIS OF LINEAR REGRESSION MODEL USING INDUSTRIAL DATA SET

Researcher:
Hiba Talal Al-Mawi

Supervisor:
Dr. Inad Nawajah

2017

Faculty of Graduate Studies and Scientific Research
Program of Mathematics

Bayesian and Non Bayesian Analysis of Linear Regression Model Using Industrial Data Set

Prepared by

Hiba Talal Al-mawi

Supervisor

Dr. Inad Nawajah

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics, College of Graduate Studies and Academic Research, Hebron University, Palestine.

2017

Bayesian and Non Bayesian Analysis of Linear Regression Model Using Industrial Data Set

By
Hiba Talal Al-mawi

This thesis was defended successfully on 21/12/2017 and approved by:

Committee Members:		**Signature**
• Dr. Inad Nawajah	Supervisor	.....
• Dr. Bader Aljawadi	Internal Examiner	.....
• Dr. Mahmoud Almanasrah	External Examiner	.....

Dedications

To my parents, my husband Sa'id and my children Mohammed and Dana

Acknowledgements

First of all I am grateful to The Almighty Allah for helping me to complete this thesis, all the Praise and thanks to Allah.

I would like to express my sincere appreciation to my principal Supervisor, Dr. Inad Nawajah, for suggesting the problem and for his guidance and support and encouragement during the time working in this thesis, without which this work would not have been possible.
Also, I would like to thank the chief of department of biology Dr. Khaldon Najim for providing the data on shoe factory workers and for his notes. I'm also grateful to all lecturers in the department of mathematics, in particular Dr. Ali Tawyha and Dr. Bader Al-Jawadi for their support towards the successful completion of my Thesis.
Without the financial support and the encourage of my husband, Sa'id, this work would not have been possible. My special thanks to my parents, sisters, brothers, and my friends at Hebron university. and to my head teacher and teachers in Alrehia school for their support and encouragement.

declaration

Abstract

Regression analysis consists of techniques for modeling the relationship between dependent variable and one or more independent variables. In this thesis, simple linear regression will be reviewed theoretically in the first part in the thesis followed by a general overview for multiple linear regression where the least square estimation method is employed. In the third part, the Bayesian technique for regression will be handled with more details to illustrate the procedure. The theoretical investigation of the multiple linear regression and Bayesian approaches accompanied with real case study using real data set from shoes factories at Hebron city to make a real comparison between least square estimation method and Bayesian technique.

empty

Contents

List of Figures

List of Tables

CHAPTER 1

Simple Linear Regression Model

Linear regression is used to estimate the unknown effect of changing one variable called dependent or (response) variable over another variable/s called independent (or predictor) variable.
The purpose of regression analysis are three-folds as mentioned in (Xin and Xiaogang, 2009):

- Establish a casual relationship between response variable Y and regressors $X_1, X_2, ..., X_n$.
- Predict Y based on a set of values of $X_1, X_2, ..., X_n$.
- Screen variables $X_1, X_2, ..., X_n$ to identify which variables are more important than others to explain the response variable Y.

1.1 Correlation

In statistical terms we use a correlation to determine whether a relationship between two numerical or quantitative variables exists.
Statisticians use a numerical measure to determine whether two variables are related and to determine the strength of the relationship between or among the variables. This measure is called a correlation coefficient, which is denoted by r.
The range of the correlation coefficient is from -1 to $+1$. It gives us an indication of the strength and direction of the linear relationship. In general, $r > 0$ indicates positive linear relationship, and $r < 0$ indicates negative linear relationship, where $r = 0$ indicates no linear relationship.
There are many formulas to find the correlation coefficient. Here we have pearson's correlation formula for two observed data set x and y:

$$r = \frac{S_{xy}}{\sqrt{(S_{xx})(S_{yy})}}$$

$$r = \frac{\sum_{i=1}^{n} xy - \frac{(\sum_{i=1}^{n} x)(\sum_{i=1}^{n} y)}{n}}{\sqrt{[\sum_{i=1}^{n} x^2 - \frac{(\sum_{i=1}^{n} x)2}{n}][\sum_{i=1}^{n} y^2 - \frac{(\sum_{i=1}^{n} y)2}{n}]}}$$

1.2 Model description

Simple regression is a description of relationship between one independent variable X and one dependent variable Y as shown below:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i. \tag{1.1}$$

where
β_0 :is the y intercept.
β_1 : Slope of the equation.
ϵ_i :Random error in $Y$for observation i.
Y_i : Dependant variable (some times referred to as the response variable) for the observation i.
X_i : Independent variable (some times referred to as the explanatory variable) for the observation i, (Krehbiel, 2008).

1.3 Assumptions

The purpose of modal formulation in regression analysis is to allow the analyst to conceptualize how the observations are generated. This formulation of statistical theory will then allow for the study of properties of estimators of the parameters. The assumptions underlying the least squares procedure are important.
Let us assume that the X_i are nonrandom $\forall i = 1, 2, ..., n$, while the ϵ_i are random variables and assumed to follow the normal distribution with a mean of 0 and constant variance of σ^2. Since y is the sum of this random term and the mean value, $E(Y)$, which is constant, the variance of y at any given value of x is also σ^2. Therefore, at any given value of x, say x_i, the dependent variable follows a normal distribution with mean 0 and standard deviation σ. Also, we assume that the ϵ_i are uncorrelated from observation to observation.

1.4 Least Squares Formulation

Myers (2000) says that the method of least squares is used more extensively than any other estimation procedure for building regression models. Prior to the 1970s, it was employed almost exclusively. This method is designed to provide the sample y intercept b_0 and the sample slope b_1 as estimators of the respective population parameters β_0 and β_1.
Eq. (1.2) uses these estimators to form the simple linear regression equation. This straight line is often referred to as the prediction line.
Krehbiel (2008) define the predicted value of y as the y intercept plus the slope times the value of x .

$$\hat{y}_i = b_0 + b_1 x_i \tag{1.2}$$

where
$\hat{y}_i$: Predicted value of y for observation i.
x_i : Value of x for observation i.
b_0: y intercept.
b_1:Slope.

Now, we need to determine the two regression coefficients b_0 and b_1. This method minimizes the sum of squared differences between y_i and $\hat{y}_i$ using the prediction line as shown in eq. (1.2).
The sum of squared residual is equal to

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{1.3}$$

Since

$$\hat{y}_i = b_0 + b_1 x_i$$

then

$$\begin{aligned} SSE &= \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2 \\ &= \sum_{i=1}^{n}(y_i^2 - 2(b_0 + b_1 x_i)y_i + (b_0 + b_1 x_i)^2) \\ &= \sum_{i=1}^{n}(y_i^2 - 2b_0 y_i - 2b_1 x_i y_i + b_0^2 + 2b_1 b_0 x_i + b_1^2 x_i^2). \end{aligned}$$

We will differentiate SSE with respect to b_0 and make it equal zero in order to find b_0 formula

$$\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{n}(-2y_i + 2b_0 + 2b_1 x_i) = 0$$

$$\begin{aligned} -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} b_0 + b_1 \sum_{i=1}^{n} x_i &= 0 \\ -n\bar{y} + nb_0 + b_1 n\bar{x} &= 0 \\ -\bar{y} + b_0 + b_1\bar{x} &= 0 \end{aligned}$$

$$b_0 = \bar{y} - b_1\bar{x} \tag{1.4}$$

The same with respect to b_1 we have

$$\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{n}(-2x_i y_i + 2b_0 x_i + 2x_i^2 b_1) = 0$$

$$\sum_{i=1}^{n} -x_i y_i + b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = 0$$
$$-\sum_{i=1}^{n} x_i y_i + (\bar{y} - b_1 \bar{x}) \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = 0$$

substitute

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

and,

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

and multiply by n we have

$$-n \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - b_1 (\sum_{i=1}^{n} x_i)^2 + b_1 n \sum_{i=1}^{n} x_i^2 = 0$$

$$n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = b_1 (n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)$$

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}. \tag{1.5}$$

1.5 Estimation of error variance σ^2

In practical situations, an estimate of the error variance, σ^2, is required. The estimate is used in the calculation of estimated standard errors of coefficient for hypothesis testing, and in many instances, plays a major role in assessing quality of fit and prediction capabilities of the regression model

$\hat{y} = b_0 + b_1 x$. Sum of squares of residuals or error sum of squares (SSE) is defined as follows:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \tag{1.6}$$

Myers (2000) defined

$$s^2 = \frac{SSE}{n-2}$$

which is an estimator of σ^2.
The quantity s^2 is often called the mean squared error (MSE) i.e.

$$MSE = \frac{SSE}{n-2}.$$

Note that we divide by $n-2$ because there are two constraints on e_i, $i = 1, ..., n$, i.e. the normal equations.

It should be emphasized that s^2 is an unbiased estimator of a σ^2 under the important assumption that the model is correct, i.e. $E(s^2) = \sigma^2$

1.6 Inference about the slope b_1

Recall, we assumed that x_i are non random and $E(\epsilon_i) = 0$, so it is not difficult to show that the estimator is unbiased, using that the mean of the distribution of y_i is given by
$E(y_i) = \beta_0 + \beta_1 x_i$

The expectation of b_1 is given by

$$\begin{aligned} E(b_1) &= \frac{n\sum_{i=1}^n x_i E(y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n E(y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\ &= \frac{n\sum_{i=1}^n x_i(\beta_0+\beta_1 x_i) - \sum_{i=1}^n x_i \sum_{i=1}^n (\beta_0+\beta_1 x_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\ &= \frac{n\beta_0\sum_{i=1}^n x_i + n\beta_1\sum_{i=1}^n x_i^2 - n^2\bar{x}\beta_0 - n^2\bar{x}^2\beta_1}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} \\ &= \beta_1. \end{aligned}$$

For the variance properties, one should note that
$Var(y_i) = Var(\epsilon_i) = \sigma^2$.
The point estimate of b_1 is given by formula (1.5), this formula can also be expressed as:

$$b_1 = \frac{\sum_{i=1}^n (x_i-\bar{x})y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}.$$

So,

$$\begin{aligned} Var(b_1) &= \frac{\sum_{i=1}^n (x_i-\bar{x})^2 \text{Var}(y_i)}{(\sum_{i=1}^n (x_i-\bar{x})^2)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2} \\ &= \frac{\sigma^2}{S_{xx}}, \qquad \text{where} \quad S_{xx} = \sum_{i=1}^n (x_i-\bar{x})^2. \end{aligned}$$

Now, under the assumption of normal theory on the ϵ_i and where the slope is a linear function of the y_i, we can write $b_1 \sim N(\beta_1, \frac{\sigma^2}{S_{xx}})$, $s = \sqrt{\frac{\sigma^2}{S_{xx}}}$ then,

$$\frac{b_1-\beta_1}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N(0,1)$$

but,

$$s^2 = \frac{SSE}{n-2} \sim \frac{\sigma^2 \chi^2(n-2)}{n-2},$$

and s^2 is independent of b_1. So by some very straight forward applications of standard relationships between distribution will allow us to write

$$\frac{\frac{b_1-\beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}}}{\frac{s}{\sigma}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{(n-2)}}{n-2}}} \sim t(n-2)$$

then

$$\frac{(b_1-\beta_1)}{s}\sqrt{S_{xx}} \sim t_{n-2},$$

where t_{n-2} is the $t-$ distribution with $n-2$ degrees of freedom.

1.7 Inference about the intercept b_0

For the intercept, we have

$$\begin{aligned} E(b_0) &= E(\bar{y} - b_1\bar{x}) \\ &= \frac{1}{n}E(\sum_{i=1}^{n} y_i) - \beta_1\bar{x} \\ &= \frac{1}{n}(\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i)) - \beta_1\bar{x} \\ &= \beta_0. \end{aligned}$$

So, b_0 is unbiased estimators for β_0.

For the variance, since $b_0 = \bar{y} - b_1\bar{x}$ then we have

$$Var(b_0) = Var(\frac{\sum_{i=1}^{n} y_i}{n}) + Var(-\bar{x}b_1)$$
$$= \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})$$

We can use the fact that b_0 is a linear combination of a normal random variables and thus observe that

$$b_0 \sim N[\beta_0, \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})].$$

1.8 The confidence interval for β_0, β_1

A confidence interval (CI) is a type of interval estimate (of a population parameter) that is computed from the observed data. Regression coefficient confidence interval is a function to calculate the confidence interval, which represents a closed interval around the population regression coefficient of interest using the standard approach and the noncentral approach when the coefficients are consistent.
Now, using the notes in section (1.6) we can write a $(1 - \alpha)100\%$ confidence intervals on β_1 as

$$b_1 \pm t_{\frac{\alpha}{2},n-2}\sqrt{\frac{s^2}{S_{xx}}} \tag{1.7}$$

The same, a $(1-\alpha)100\%$ confidence intervals on β_0 is written as

$$b_0 \pm t_{\frac{\alpha}{2},n-2}s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}. \tag{1.8}$$

by using the notes in section (1.7).

1.9 Testing of hypothesis for b_1 and b_0

This section discusses hypothesis tests on the regression coefficients in simple linear regression. These tests can be carried out if it can be assumed that the random error term, ϵ_i, is normally and independently distributed with a mean of zero and variance of σ^2.
The t tests are used to conduct hypothesis tests on the regression coefficients obtained in simple linear regression. A statistic based on the distribution is used to test the two-sided hypothesis that the true slope, β_1, equals some constant value, $\beta_{1,0}$. The statements for the hypothesis test are expressed as:

$$H_0 : \beta_1 = \beta_{1,0}$$

The test statistic used for this test is:

$$t = \frac{(b_1 - \beta_{1,0})}{s}\sqrt{S_{xx}}.$$

For the intercept, if one interested in testing

$$H_0 : \beta_0 = \beta_{0,0}$$

the appropriate test statistic is given by

$$t = \frac{b_0 - \beta_{0,0}}{s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}.$$

1.10 Analysis of variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. In general, the

purpose of analysis of variance (ANOVA) is to test for significant differences between means. But in regression models it consists of calculations that provide information about levels of variability within a regression model and form a basis for tests of significance.
The total variation of the response variable y can be decomposed into two parts: the residual variation of y (error sum of squares (SSE)) and the explained variation of y (regression sum of squares (SSR)).

To calculate the test statistic we need to find:

- Sum squares total:$SST = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$.
- Sum squares regression :$SSR = b_1^2 S_{xx}$.
- Sum squares error: $SSE = SST - SSR$.

Each sum of squares can be divided by an appropriate constant (degrees of freedom, which indicates how many independent pieces of information involving the n independent numbers $y_1, y_2, ..., y_n$ are needed to compile the sum of squares (Draper and Smith, 1998). to get the mean sum of squares due to regression MSR, and the mean sum of squares due to error MSE, as shown in table (1.1)

Source of variation(Source)	Sum of squares (SS)	Degrees of freedom(df)	Mean sqares(MS)	F statistic
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$F_1 = \frac{MSR}{MSE}$
Error	SSE	n-2	$MSE = \frac{SSE}{n-2}$	
Total	SST	n-1		

Table 1.1: *ANOVA table for simple regression*

Some notes on ANOVA table:

- If the calculated value of the statistic falls in the critical region, we reject the null hypothesis and conclude that the regression coefficient is significant. In other words, we say that the explanatory variable has significant effect on the response variable. The critical region (or the rejection region) is determined by the value of $F-$tabulated,$F^1_{n-2}\alpha$. If the value of the statistic falls outside the critical region, we do not reject the null hypothesis and conclude that the regression coefficient is not significant, i.e., the explanatory variable has no significant effect on the response variable.

- The coefficient of determination r^2 is the amount of variance in y that explained by x.

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

  with ranges $0 \leq r^2 \leq 1$.

- The correlation coefficient r is equal to:

$$r = \pm\sqrt{r^2}$$

$$r = \pm\sqrt{\frac{b_1 S_{xy}}{S_{yy}}}$$

Illustrative Example

Let us take this example from (Kewan, 2015) of the relationship between income (X) and food expenditure (Y). Suppose we take a sample of seven households from a small city and collect information on their incomes and food expenditure (in

hundreds of dollars) in a certain month. The data obtained was as given in table (1.2).

Income	Food expenditure
55	14
83	24
38	13
61	16
33	9
49	15
67	17

Table 1.2: *Income and Food Expenditure of seven households*

In this example we want to find the best regression line for the data on income and food expenditure on the seven households given in table (1.2) by using the income as independent variable and the food expenditure as dependent variable.

But before any step, we must find the correlation coefficient r which can be computed from the formula :

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)2}{n}][\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)2}{n}]}}$$

Table (1.3) shows the calculations required for the computation of r.

X	Y	XY	X^2	Y^2
55	14	770	3025	196
83	24	1992	6889	576
38	13	494	1444	169
61	16	976	3721	256
33	9	297	1089	81
49	15	735	2401	225
67	17	1139	4489	289
$\sum X = 386$	$\sum Y = 108$	$\sum XY = 6403$	$\sum X^2 = 23058$	$\sum Y^2 = 1792$

Table 1.3: *Data on income and food expenditure*

After substitution we have:

$$r = \frac{6403 - \frac{(386)(108)}{7}}{\sqrt{[23058 - \frac{(386)^2}{7}][1792 - \frac{(108)2}{7}]}} = 0.96$$

We note that the correlation coefficient is very close to 1, which means that the relation between the income and the food expenditure is very strong.

Then, we want to find the values of b_0 and b_1 for the regression model

$$\hat{y}_i = b_0 + b_1 x_i$$

We denote the independent variable (income) by x and the dependent variable (food expenditure) by y, both in hundreds of dollars.

Using the calculations in table (1.3) and finding

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

$$\bar{x} = \frac{386}{7} = 55.14,$$

and

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\bar{y} = \frac{108}{7} = 15.43.$$

Substitute these calculations in the formula:

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}.$$

we find

$$b_1 = \frac{7(6403) - (386)(108)}{7(23058) - (386)^2} = 0.2525.$$

Also,

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_0 = (15.43) - (0.2525)(55.14) = 1.505$$

Thus, our estimated regression model is:

$$\hat{y}_i = 1.505 + 0.2525x_i$$

Some notes on this regression line:

- Using this estimated regression model, we can find the predicted value of y for any fixed value of x (during the month in which the data has been collected). For instance, suppose we randomly select a household whose monthly income is $6100\$$ so that $x = 33$. The expected value of food expenditure for this household is:

  $\hat{y} = 1.505 + 0.2525(33) = 9.8375$ hundred$\$ = 983.75\$.$

  In other words, based on our regression line, we predict that a household with a monthly income of $3300\$$ is expected to spend $983.75\$$ per month on food. In our data

on seven households, there is a one household whose income is $3300\$$. The actual food expenditure for that household is $900\$$ (see Table (1.3)). The difference between the actual and predicted values gives the error of prediction .

$$\epsilon = y - \hat{y} = 9 - 9.8375 = -0.8375 \text{ hundred\$} = 83.75\$$$

- $b_0 = 1.505$, is the expected value of y when $x = 0$. That is, a household with no income is expected to spend $150.5\$$ per month on food.

- The value of b_1 in the regression model gives the change in y due to increase of one unit in x. That is, for every one dollar increase in income, a household food expenditure is predicted to increase by $0.2525\$$.

- 95 % confidence interval of β_1 is calculated as follows: using eq. (1.7) we need to calculate :

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

$$= 23058 - \frac{(386)^2}{7} = 1772.85$$

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{12.7}{5} = 2.54.$$

$$\alpha = 0.05$$

from t-distribution we find $t_{0.025,5} = 2.571$
substitute these value in eq. (1.7) we have that the 95% confidence interval is $(0.1552, 0.3498)$.
Using the previous calculation and eq. (1.8) we can find the confidence interval of β_0 which is $(-4.066, 7.076)$.

- Set the null hupothesis H_0 and the alternative hypothesis H_1:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Compute the quantities SST, SSR, and SSE respectively as:

$$SST = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 1792 - \frac{(108)^2}{7} = 125.71$$

$$SSR = b_1^2 S_{xx} = (0.2525)^2(1772.85) = 113.03$$

$$SSE = SST - SSR = 12.7$$

The ANOVA table is shown below:

(Source)	(SS)	(df)	Mean (MS)	F statistic	p-value
Regression	113.03	1	113.03	$F = 44.49$	0.001
Error	12.7	5	2.54		
Total	125.71	6			

Table 1.4: *ANOVA table*

The F value from the table is $F_1^5 0.05 = 6.61$, we note that the tabulated F is smaller than the calculated F, i.e. the F-statistic falls in the rejection region, so we reject the null hypothesis. Also, the p-value less than 0.05. That is, the income is useful to explain the food expenditure in a satisfactory way.

- From table (1.4), we note that $r^2 = \frac{113.03}{125.71} = 0.899$. In general, the higher $r-$ squared, the better the model fits your data.

CHAPTER 2

Multiple Linear Regression Model

2.1 The model description

Some times we need to predict the value of one variable based on the value of two or more other variables, this model is called multiple linear regression model. Clearly multiple linear regression analysis is an extension of simple linear regression analysis (Montgomery et al., 2012).

For example, suppose that one want to study factors that might affect systolic blood pressures for women aged 45 to 65 years old. The response variable is systolic blood pressure (y). Suppose that the two predictor variables of interest are age (x_1) and body mass index (BMI) (x_2). The general structure of

a multiple linear regression model for this situation would be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

- The equation $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ describes the mean value of blood pressure for specific values of age and BMI.
- The error term (ϵ) describes the characteristics of the differences between individual values of blood pressure and their expected values of blood pressure.

In general, the multiple linear regression equation is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon \qquad (2.1)$$

where Y is the response variable, $X_1, X_2, X_3, ..., X_k$ are independent or predictor or explanatory variables. $\beta_0, \beta_1, ..., \beta_k$ are fixed (but unknown parameters). ϵ is a random variable representing the error or residuals that is normally distributed with mean 0 and variance σ_ϵ^2 .

We can generalize the eq. (2.1) as

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i, \qquad (i = 1, 2, \ldots, n; n \geq k+1) \qquad (2.2)$$

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i \qquad (2.3)$$

Myers (2000) defined a linear model as a model that is linear in the parameters.i.e., linear in the coefficients, the β's in eq. (2.2).
Examples of a linear model: (Myers, 2000):

- A model quadratic in x but of course linear in the β's is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

- A model contains interaction among a pair of regressor variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon.$$

- In some application there is a need to perform transformations on the regressor variables. For example, consider the case of three regressor variables x_1, x_2, and x_3. The following is a linear model in which the natural log transformation is made on each variable:

$$y = \beta_0 + \beta_1 lnx_1 + \beta_2 lnx_2 + \beta_3 lnx_3 + \epsilon.$$

  In each of the three illustrations provided here, transformations are made on the regressor variables but the model remains linear in the parameters.

- The analyst may be interested in using a log transformation on a y and, say, reciprocal transformations on x_1 and x_2. As a result, the linear model is written

$$lny = \beta_0 + \beta_1(\frac{1}{x_1}) + \beta_2(\frac{1}{x_2}) + \epsilon.$$

2.2 Assumptions

Four assumptions of multiple regression that researchers should always test

2.2.1 Linearity

There must be a linear relationship between the dependent variable and the independent variables. Scatterplots can show whether there is a linear or curvilinear relationship.

2.2.2 Normality

The linear regression analysis requires all variables to be normal. This assumption may be checked by looking at a histogram or a Q-Q-Plot.

2.2.3 Multicollinearity

Multiple linear regression assumes that there is no multicollinearity in the data. Multicollinearity arises when at least two highly correlated predictors are assessed simultaneously in a regression model. So, multicollinearity(also collinearity) is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy.

Multicollinearity causes the following two basic types of problems as found in (Frost, 2017):

- The coefficient estimates can swing wildly based on which other independent variables are in the model. The coefficients become very sensitive to small changes in the model.
- Multicollinearity reduces the precision of the estimate coefficients, which weakens the statistical power of your

regression model. You might not be able to trust the p-values to identify independent variables that are statistically significant.

Simon (2004) stated some of the common methods used of detecting multicollinearity include:

- The analysis exhibits the signs of multicollinearity such as, estimates of the coefficients vary from model to model.
- The t-tests for each of the individual slopes are non-significant ($P - value > 0.05$), but the overall F-test for testing all of the slopes are simultaneously 0 is significant ($P - value < 0.05$).
- The correlations among pairs of predictor variables are large.

It is possible that the pairwise correlations are small, and yet a linear dependence exists among three or even more variables. Many regression analysis often rely on what are called Variance Inflation Factor (VIF) to help detect multicollinearity. Variance Inflation Factor (VIF) quantifies how much the variance of the estimated coefficient is inflated.
The VIF for a coefficient β_j is:

$$\text{VIF} = \frac{1}{1 - R_j^2}.$$

where R_j^2 is the coefficient of multiple determination resulting from regressing the jth predictor variable x_j, on the remaining $n - 1$ predictor variables.
Statistical software calculates a VIF for each independent variable. VIFs start at 1 and have no upper limit. A value of 1 indicates that there is no correlation between this independent variable and any others. VIFs between 1 and 5 suggest that there

is a moderate correlation, but it is not severe enough to warrant corrective measures. VIFs greater than 5 represent critical levels of multicollinearity where the coefficients are poorly estimated, and the p-values are questionable.

Here we have an example from (Myers, 2000).

Site	x_1	x_2	x_3	x_4	x_5	Y
1	15.57	2463	472.92	18.0	40.45	566.52
2	44.02	2048	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	6505	568.33	36.7	3.90	1603.62
5	49.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.15	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20106	3655.08	180.5	6.15	3503.93
11	96.00	13313	2912.00	60.9	5.88	3571.89
12	131.42	10771	3921.00	103.7	4.88	3741.40
13	127.21	15543	3865.67	126.8	5.50	4026.52
14	252.90	36194	7684.10	157.7	7.00	10343.81
15	409.20	34703	12446.33	169.4	10.78	11732.17
16	463.70	39204	14098.40	331.4	7.05	15414.94
17	510.22	86533	15524.00	371.6	6.35	18854.45

Table 2.1: *Hospital manpower data*

Data are given in table (2.1) that reflect information taken from seventeen U.S Naval hospitals at various sites around the world. The regressors are workload variables, i.e., items that result in the need for manpower in a hospital installation. A brief description of the variables is as follows:
Y:Monthly man-hours
x_1:Average daily patient load

x_2:Monthly X-ray exposures
x_3:Monthly occupied bed days
x_4:Eligible population in the area $\div 1000$
x_5:Average length of patients stay in days.

The goal here is to produce an empirical equation that will estimate(or predict) manpower needs for Naval hospitals. The following are the least squares regression equation, the estimate of residual standard deviation, and the coefficient of the determination.

$$\hat{Y} = 1962.948 - 15.8517x_1 + 0.05593x_2 + 1.58962x_3$$

$$-4.21867x_4 - 394.314x_5$$

$s = 642.088$ man-hours/month , $R^2 = 0.99082$
These results seem to reflect a satisfactory fit between man-hours and the workload variables. However the signs of the coefficients could be classified as alarming if we interpret them literally. The coefficients of the variable x_1(average daily patient load), x_4 (eligible population), and x_5(average length of patients stay), are negative. This implies that, say, in the case of x_1, an increase in patient load, when other x's are held constant, is accompanied by a corresponding decrease in hospital manpower, a conclusion which, of course, is ludicrous. The correlation matrix, showing the empirical linear dependency among these regressor workload variables is as follows:

$$Correlation = \begin{bmatrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ x_1 & 1.00000 & 0.90738 & 0.99990 & 0.93569 & 0.67120 \\ x_2 & & 1.00000 & 0.90715 & 0.91047 & 0.44665 \\ x_3 & & & 1.00000 & 0.93317 & 0.67111 \\ x_4 & & & & 1.00000 & 0.46286 \\ x_5 & & & & & 1.00000 \end{bmatrix}$$

It would seem that, even though the linear regression model fits the data quite well, the rather curious signs on the regression coefficients may be a result of the effect of multicollinearity. The variance inflation factors, which are the diagonals of the inverse of the correlation matrix, are given by

$$\begin{aligned} x_1 &: VIF = 9597.57 \\ x_2 &: VIF = 7.94 \\ x_3 &: VIF = 8933.09 \\ x_4 &: VIF = 23.29 \\ x_5 &: VIF = 4.28 \end{aligned}$$

It is clear that at least two of the regression coefficients, b_1 and b_3, are estimated vary poorly in comparison to the ideal, i.e., the condition in which there is no multicollinearity.
In the case of the hospital data set, the correlation between x_1 and x_3 stands out as being noteworthy. If one were attempt to reduce the multicollinearity, but still confine the estimation procedure to ordinary least squares, the elimination of one of the regressors, either x_1 or x_3, would seem to be a promising or, perhaps necessary, approach. The implication is, perhaps, that a model containing x_1 does not need x_3 or, vice versa. The pair of variables together may prohibit quality estimation

of either coefficient. the variable x_1 was eliminated, and the following model and statistics were obtained:

$$\hat{Y} = 2032.188 + 0.05608x_2 + 1.0884x_3 - 5.0041x_4 - 410.083x_5$$

$$R^2 = 0.9908, s = 615.489$$

$$\begin{aligned} x_2 &: VIF = 7.92 \\ x_3 &: VIF = 23.93 \\ x_4 &: VIF = 12.70 \\ x_5 &: VIF = 3.36 \end{aligned}$$

The regression without x_1 has not only reduced the residual estimate of variance s^2, while not severely altering R^2, but has also substantially reduced the variance inflation factor on b_3 from 8933.09 to 23.93. Thus, from the results shown here, elimination of x_1 would seem to produce reduced multicollinearity and, perhaps an improved regression.

2.2.4 Homoscedasticity

Homoscedasticity means that the variance of errors is the same across all levels of the independent variables. When the variance of errors differs at different values of the independent variables, heteroscedasticity is indicated.
A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of the independent variables. Also, a scatterplot of residuals versus predicted values is good way to check for homoscedasticity. There should be no clear pattern in the distribution; if there is a cone-shaped pattern, the data is heteroscedastic.

2.3 The least squares procedure

The method of least squares can be used to estimate the regression coefficients in eq. (2.3).
So, the least squares function as shown in Montgomery et al. (2012) is:

$$SSE = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - b_0 - \sum_{j=1}^{k} b_j x_{ij})^2 \tag{2.4}$$

SSE must be minimized with respect to $b_0, b_1, \ldots, b_k$, so the least squares estimators of $b_0, b_1, \ldots, b_k$ must satisfy

$$\frac{\partial SSE}{\partial b_0} | b_0, b_1, ..., b_k = -2 \sum_{i=1}^{n} (y_i - b_0 - \sum_{j=1}^{k} (b_j x_{ij})) = 0$$

$$\sum_{i=1}^{n} (y_i - b_0 - \sum_{j=1}^{k} b_j x_{ij}) = 0$$

and hence

$$\sum_{i=1}^{n} y_i = n b_0 + \sum_{i=1}^{n} \sum_{j=1}^{k} b_j x_{ij}$$

$$n b_0 + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} + \ldots + b_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i \tag{2.5}$$

and

$$\frac{\partial S}{\partial b_j} | b_0, b_1, ..., b_k = -2 \sum_{i=1}^{n} (y_i - b_0 - \sum_{j=1}^{k} (b_j x_{ij})) x_{ij} = 0.$$

$$\sum_{i=1}^{n}(y_i - b_0 - \sum_{j=1}^{k}(b_j x_{ij}))x_{ij} = 0$$

$$\sum_{i=1}^{n} b_0 x_{ij} + \sum_{i=1}^{n}(\sum_{j=1}^{k} b_j x_{ij})x_{ij} = \sum_{i=1}^{n} y_i x_{ij}$$

$$b_0 \sum_{i=1}^{n} x_{ij} + b_1 \sum_{i=1}^{n} x_{i1}x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{ik}x_{ik} = \sum_{i=1}^{n} x_{ij}y_i \quad (2.6)$$

Now, we have the least− squares normal equations as follows:

$$\begin{aligned}
nb_0 + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} + \ldots + b_k \sum_{i=1}^{n} x_{ik} &= \sum_{i=1}^{n} y_i \\
b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{i1}x_{ik} &= \sum_{i=1}^{n} x_{i1}y_i \\
&\vdots \\
b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{ik}x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{ik}^2 &= \sum_{i=1}^{n} x_{ik}y_i.
\end{aligned}$$

The solution of these equations will be the least squares estimators $b_0, b_1, \ldots, b_k$.
the previous system can be represented in matrix notation as

$$\boldsymbol{X'Xb} = \boldsymbol{X'Y} \quad (2.7)$$

where

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

We want to find the vector of least $-$ squares estimators b that minimizes sum of squares residuals such that:

$$\begin{aligned} S = \sum_{i=1}^{n} \epsilon_i^2 = \acute{\epsilon}\epsilon = (y - xb)'(y - xb) \\ = \acute{y}y - b\acute{x}y - \acute{y}xb + b\acute{x}xb \end{aligned}$$

since $(\boldsymbol{b}\acute{\boldsymbol{X}}\boldsymbol{Y})' = \acute{\boldsymbol{Y}}\boldsymbol{X}\boldsymbol{b}$, then

$$S(b) = \acute{\boldsymbol{Y}}\boldsymbol{Y} - 2\boldsymbol{b}\acute{\boldsymbol{X}}\boldsymbol{Y} + \boldsymbol{b}\acute{\boldsymbol{X}}\boldsymbol{X}\boldsymbol{b}$$

The least squares estimator must satisfy

$$\begin{aligned}\frac{\partial(S)}{\partial(b)} &= -2\acute{\boldsymbol{X}}\boldsymbol{Y} + 2\acute{\boldsymbol{X}}\boldsymbol{X}\boldsymbol{b} = 0\\ \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} &= \boldsymbol{X}'\boldsymbol{Y}\\ \boldsymbol{b} &= (\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}\boldsymbol{Y}\end{aligned}$$

Note that $(\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}$ always exists if the regressors are linearly independent.

The fitted regression model corresponding to the levels of the regression variables $\boldsymbol{X} = [1, x_1, x_2, ..., x_n]$ is

$$\hat{Y} = \acute{X}b = b_0 + \sum_{j=1}^{k} b_j x_j. \tag{2.8}$$

2.4 Estimation of error variance σ^2

It is necessary to obtain a good estimate of σ^2 in multiple regression. We use the estimate in variable screening via hypothesis testing, or for assessing model quality.
Before we discuss the estimator, we should state the relationship between the total sum of squares(SST) and the regression sum of squares(SSR). The relationship shown in (Myers, 2000):

$$SST = SSR + SSE.$$

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

Where SSE is the familiar residual sum of squares.
SSR explains variation that accounts for k model terms. Thus

the total degrees of freedom partition as follows:

$$n - 1 = k + (n - k - 1)$$

The unbiased estimator, s^2, expresses variation in the residuals, i.e., variation about the regression $\hat{y} = xb$, with the denominator now becoming $n - p$, where p is the number of parameters estimated. In the notation of the model in (2.7), $p = k + 1$. As a result,

$$s^2 = \frac{(y - xb)'(y - xb)}{n - p} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n - p}$$

Where $\hat{y}_i$ is the predicted or fitted response at the ith data point. As in the simple linear regression case of chapter (1), this estimator, the residual mean square, expresses natural variation or experimental error variance and is an unbiased estimator, assuming that the model postulated, and thus fitted is correct.

2.5 Properties of the least squares estimators under ideal condition

One should recall that the ideal condition of the model in eq. (2.3) as mentioned in (Myers, 2000) are:

- the ϵ_i has a mean of zero.
- the ϵ_i are uncorrelated, and have common variance σ^2.

Under the condition that $E(\epsilon) = 0$, then we can easily show that b in eq. (2.8) is an unbiased estimator for β. since $E(y) = x\beta$, and x is not random.

$$E(\boldsymbol{b}) = E((\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}\boldsymbol{Y}) = (\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}E(\boldsymbol{Y}) = (\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

For the variance of $\boldsymbol{b}$, we have

$$\begin{aligned} Var(\boldsymbol{b}) &= (\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}Var(\boldsymbol{Y})\boldsymbol{X}(\acute{\boldsymbol{X}}\boldsymbol{X})^{-1} \\ &= (\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}\acute{\boldsymbol{X}}\sigma^2 I\boldsymbol{X}(\acute{\boldsymbol{X}}\boldsymbol{X})^{-1} \\ &= \sigma^2(\acute{\boldsymbol{X}}\boldsymbol{X})^{-1}. \end{aligned}$$

Note that $SE(b) = \sqrt{\sigma^2(\acute{x}x)^{-1}}$ where $SE(b)$ is the standard error of b and $b \sim N(\beta, \sigma^2(\acute{x}x)^{-1})$.

2.6 Hypothesis tests in multiple linear regression

This section from Myers (2000) discusses hypothesis tests on the regression coefficients in multiple linear regression. As in the case of simple linear regression, these tests can only be carried out if it can be assumed that the random error terms, ϵ_i, are normally and independently distributed with a mean of zero and variance of σ^2.
There are three types of hypothesis tests for multiple linear regression models:

- Test for significance of regression: This test checks the significance of the whole regression model.
  The test is used to check if a linear statistical relationship exists between the response variable and at least one of the predictor variables. The statements for the hypotheses are:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$

$$H_1 : \beta_j \neq 0, \text{for at least one } j$$

  The test for H_0 is carried out using the following statistic:

$$F_0 = \frac{MSR}{MSE}$$

Where MSR is the regression mean square and MSE is the error mean square. The null hypothesis, H_0, is rejected if the calculated statistic, F_0, is such that:

$$F_0 > f_{\alpha,k,n-(k+1)}$$

where k is a degree of freedom in the numerator and $n-(k+1)$ is a degree of freedom in the denominator.

- t test:
  The t test is used to check the significance of individual regression coefficients in the multiple linear regression model. The hypothesis statements to test the significance of a particular regression coefficient, β_j, are:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

The test statistic for this test is based on the t distribution

$$T_0 = \frac{b_j}{SE(b_j)}$$

where the standard error, $Se(b_j)$, is obtained. We would fail to reject the null hypothesis if the test statistic lies in the acceptance region:

$$-t_{\frac{\alpha}{2},n-2} < T_0 < t_{\frac{\alpha}{2},n-2}$$

- F test: This test can be used to simultaneously check the significance of a number of regression coefficients. It can also be used to test individual coefficients. This test can

be considered to be the general form of the t test mentioned in the second item. This is because the test simultaneously checks the significance of including many (or even one) regression coefficients in the multiple linear regression model. Adding a variable to a model increases the regression sum of squares, SSR. The test is based on this increase in the regression sum of squares. The increment in the regression sum of squares is called the extra sum of squares. Assume that the vector of the regression coefficients, $\boldsymbol{\beta}$, for the multiple linear regression model, $Y = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$, is partitioned into two vectors with the second vector, θ_2, containing the last r regression coefficients, and the first vector, θ_1, containing the first $(k + 1 - r)$ coefficients as follows:

$$\beta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

with:

$$\theta_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-r} \end{bmatrix}$$

and

$$\theta_2 = \begin{bmatrix} \beta_{k-r+1} \\ \beta_{k-r+2} \\ \vdots \\ \beta_k \end{bmatrix}$$

The hypothesis statements to test the significance of adding the regression coefficients in θ_2 to a model containing the

regression coefficients in θ_1 may be written as:

$$H_0 : \theta_2 = 0$$

$$H_1 : \theta_2 \neq 0$$

The test statistic for this test follows the F distribution and can be calculated as follows:

$$F_0 = \frac{\frac{SSR(\frac{\theta_1}{\theta_2})}{r}}{MSE}$$

where $SSR(\frac{\theta_1}{\theta_2})$ is the increase in the regression sum of squares when the variables corresponding to the coefficients in θ_2 are added to a model already containing θ_1, and $MSE = \frac{SSE}{n-2}$.
The null hypothesis, H_0, is rejected if

$$F_0 > f_{\alpha,r,n-(k+1)}.$$

Rejection of H_0 leads to the conclusion that at least one of the variables in $x_{k-r+1}, x_{k-r+2}, ..., x_k$ contributes significantly to the regression model.

2.7 The confidence interval

A confidence interval for β_j is as follow:

$$\frac{b_j - \beta_j}{Se(b)} = \frac{b_j - \beta_j}{\sqrt{s^2(x'x)^{-1}}} \sim t_{\frac{\alpha}{2p},n-p}.$$

where $j = 0, 1, 2, ..., k$, $p = k + 1$ and s^2 is the error mean square (an estimate of the variance σ^2).

So, a $(1-\alpha)100\%$ confidence interval on the regression coefficient β_j is obtained as follows

$$b_j \mp t_{\frac{\alpha}{2p},n-p}\sqrt{s^2(x'x)^{-1}}$$

2.8 Analysis of variance (ANOVA)

Draper and Smith (1998) stated that the ANOVA calculations for multiple regression are nearly similar to the calculations for simple linear regression, except that the degrees of freedom are adjusted to reflect the number of explanatory variables included in the model.
For k explanatory variables, the model degrees of freedom are equal to k, the error degrees of freedom are equal to $(n-k-1)$, and the total degrees of freedom are equal to $(n-1)$.

The corresponding ANOVA table is shown in table (2.2):

Source	SS	df	MS	F
Regression	SSR	k	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Error	SSE	n-k-1	$MSE = \frac{SSE}{n-k-1}$	
Total	SST	n-1		

Table 2.2: *ANOVA table for multiple regression*

The statements for the hypotheses are:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$

$$H_1 : \beta_j \neq 0, \text{for at least one } j$$

The test for H_0 is carried out using the following statistic:

$$F_0 = \frac{MSR}{MSE}$$

The null hypothesis, H_0, is rejected if the calculated statistic, F_0, is such that:

$$F_0 > f_{\alpha,k,n-(k+1)}$$

2.9 Applications of linear regression using Real Data Set

Regression analysis is one of the most commonly used statistical methods in practice. Applications of regression analysis can be found in many scientific fields including medicine, biology, agriculture, economics, engineering, sociology, geology, etc. In this study, we have used real biological data in order to explain the regression model in more details.
First of all, two factors that uses in biological analysis should be defined, exposure and outcome. S.Kramer (1988) defines the exposure as is the putative causal factor, or effector, that the investigator believes may be(at least partly) responsible for the outcome under study. And the outcome in an analytic study is the effect that the investigator believes may be caused by exposure. The main objective in most epidemiology studies is the measurement of the association between exposure and outcome.

Exposure to organic materials in shoe factory may cause lung function deterioration in workers.
Previous studies demonstrate an increased lung cancer risk for shoemakers and workers in shoe manufacturing and the risk seems to double after being 30 years in these occupations (Fu et al., 1996). Palestinian Statistics (2015) reported that Hebron city has the maximal number of shoe factories. This made us suspect that the employees in these factory encounter some lung function problems, we there fore conducted across-sectional survey in some employees in several factories at Hebron city in Palestine, aiming to characterize lung function and respiratory symptoms in factory employees and to estimate as-

sociations with exposures to organic compounds that found in the raw materials with existence of some confounders like height, weight, age, number of years of smoking and number of years of working in shoes factories.
This study included 113 employees chosen randomly from shoe factories in Hebron city. Lung function abnormality was measured by using a spirometer machine that measures the Forced Expiratory Volume in one second ($FEV1$) which is considered to be dependent variable. FEV1 means the volume exhaled during the first second of a forced expiratory maneuver started from the level of total lung capacity. (FEV1) was recorded by using a spirometer (Micro Direct Spiro USB) with spida 5 software program. All cases were at rest 15 minutes and were not allowed to smoke for 1 hour prior to the measurement. The procedure was carefully described to the employees with emphasis on the need to avoid leaks round the mouthpiece and to make a maximum inspiration without hesitation and without leaning forward. The procedure was demonstrated using a detached mouthpiece with nose clip in an upright posture and tight clothing was loosened . Employees were allowed to do 2 practice attempts before the actual measurement.

For exposure, four different tasks were performed in the factory mainly (painting, management, adhesive and molding), these tasks were considered as exposure independent variables. The data(shown in appendix A) were analysed using Statistical Package for Social Sciences $SPSS$ (version 20), guided by the article (Joaquim et al., 2007).
Dependent and independent variables were all included to formulate the following multiple linear regression:

$$\text{FEV1} = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{smoking years}) + \beta_3(\text{working years}) + \beta_4(\text{height}) + \beta_5(\text{weight}) + \beta_{61}(\text{management}) + \beta_{62}(\text{adhesive}) + \beta_{63}(\text{painting}) + \beta_{64}(\text{molding}) + \epsilon. \tag{2.9}$$

$FEV1$ expresses the Y variable such that

$$FEV1 = Y = \begin{bmatrix} 3.35 \\ 3.41 \\ \vdots \\ 3.48 \end{bmatrix}$$

with mean 3.5426 and standard deviation 0.7256.
Figure (2.1) shows the Boxplot of(FEV1), as you see the data are normal and have outliers points.

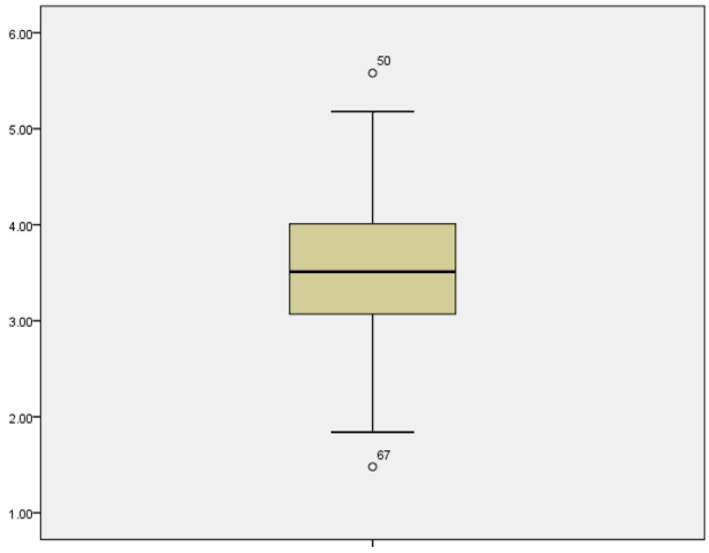

Figure 2.1: *The Boxplot of FEV1*

Also, figure (2.2) shows the normality of FEV1.

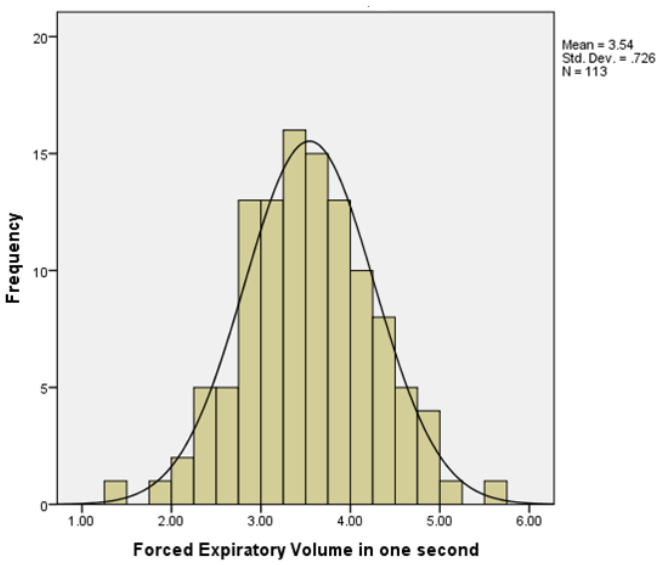

Figure 2.2: *The normality of FEV1*

The independent variables $x_1, ..., x_9$ represented the variables mentioned in eq. (2.9) respectively, such that

$$x = \begin{bmatrix} & x_1 & x_2 & x_3 & x_4 & x_5 & x_{61} & x_{62} & x_{63} & x_{64} \\ 1 & 33 & 17 & 8 & 165 & 73 & 0 & 0 & 0 & 1 \\ 1 & 44 & 10 & 14 & 170 & 78 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 23 & 0 & 7 & 185 & 85 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The table (2.3)shows the description for these variables.

	mean	standard deviation
Age	34.52	11.698
smoking years	6.889	9.2011
working years	11.020	8.9627
weight	77.10	13.567
height	169.97	7.644

Table 2.3: *The Descriptive Statistics for the independent variables*

Figure (2.3) shows the percentage of workers in each working task in the factory. Clearly the percentage of molding task is the highest one.

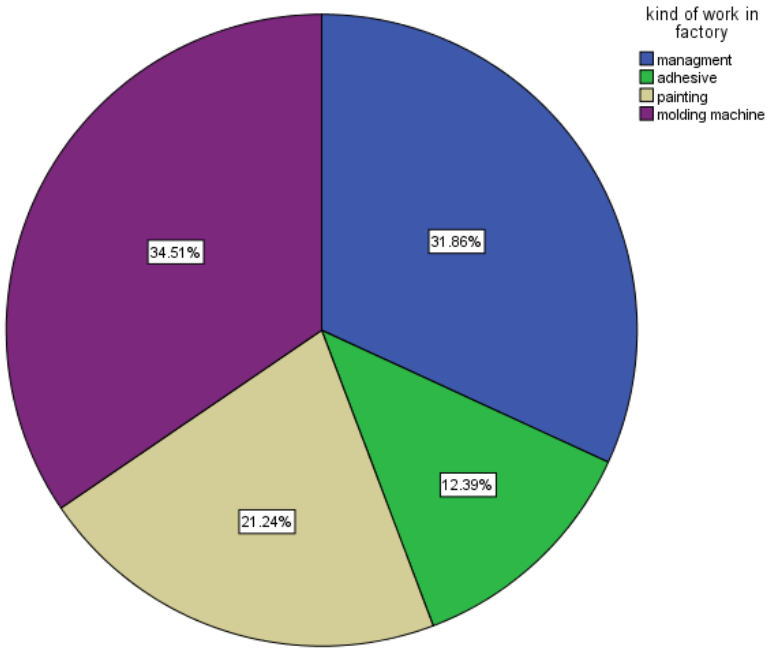

Figure 2.3: *Kind of work in the factory*

As we have four different working groups in the factory, then we have three dummy variables(adhesive, painting, and

molding) which contain all of the information needed to determine which observations are included in which group. Thus, each of the groups is defined by having one of the dummy variables equal to one except of one group which is all zeros. The group with all zeros is known as the reference group, in our example we have (management) as a reference group.

Before we make a regression, we must find the correlation between the variables, table (2.4) shows the correlation between our variables.

Table 2.4: *Correlations matrix for the variables*

	age	smoke	work	manag	adhesive	painting	mold	height	weight	FEV1
age	1	0.187	0.630	0.306	-0.082	0.1	-0.328	-0.047	0.214	-0.623
smoking	0.187	1	0.466	0.289	-0.154	-0.081	-0.107	-0.011	0.044	-0.144
working	0.630	0.466	1	0.401	-0.243	-0.068	-0.165	0.006	0.220	-0.411
management	0.306	0.289	0.401	1	-0.257	-0.355	-0.496	0.090	0.094	-0.160
adhesive	-0.082	-0.154	-0.243	-0.257	1	-0.159	-0.273	-0.006	-0.005	0.070
painting	0.1	-0.081	-0.068	-0.355	-0.195	1	-0.377	-0.157	0.003	-0.158
molding	-0.328	-0.107	-0.165	-0.495	-0.273	-0.377	1	0.051	-0.091	0.244
height	-0.047	-0.011	0.006	0.09	-0.006	-0.157	0.051	1	0.395	0.188
weight	0.210	0.044	0.220	0.094	-0.005	0.003	-0.091	0.395	1	-0.092
FEV1	-0.623	-0.144	-0.411	-0.160	0.070	-0.158	0.244	0.188	-0.092	1

We note from this table that the correlations between the predictors are very weak, which indicates less multicollinearity. We note a good correlation between FEV1 and other variables.

Table (2.5) provides the R and R^2 values. Where R value represents the simple correlation and equal to 0.649 which indicates a moderate degree of correlation. Also, we have R^2 value which indicates how much of the total variation in the dependent variable can be explained by the independent variable, in our case, 42.2% can be explained.

Table 2.5: *Model Summary*

Model	R	R Square	Adjusted R Square	Std.Error of the Estimate
1	0.649	0.422	0.377	0.57256

Now, the least square estimators b can be obtained by $b = (\acute{x}x)^{-1}\acute{x}y$ as shown in Table (2.6)

Table 2.6: *The coefficients and their 95% confidence interval*

		Lower Bound	Upper Bound	VIF	p-value
b_0	2.386	-0.163	4.934		0.066
b_1	-0.035	-0.048	-0.023	1.898	0.000
b_2	-0.002	-0.015	0.012	1.340	0.815
b_3	-0.003	-0.022	0.015	2.326	0.709
b_4	0.015	-0.001	0.031	1.250	0.060
b_5	-0.001	-0.01	0.008	1.292	0.778
b_{62}	-0.007	-0.396	0.383	1.442	0.974
b_{63}	-0.139	-0.458	0.179	1.485	0.387
b_{64}	0.010	-0.279	0.298	1.647	0.947

Looking at the p-value of the t-test for each predictor, we can see that only the age and the height contributes to the model. This is because all variables are introduced in one step using the enter method in SPSS.
So, the regression model is

$$\text{FEV1} = 2.386 - 0.035(\text{age}) - 0.002(\text{smoking years}) - 0.003(\text{working years}) + 0.015(\text{height}) - 0.001(\text{weight}) - 0.007(\text{adhesive}) - 0.139(\text{painting}) + 0.01(\text{molding}). \tag{2.10}$$

which has a normally distributed residuals with mean approximately to zero as in figure (2.4).

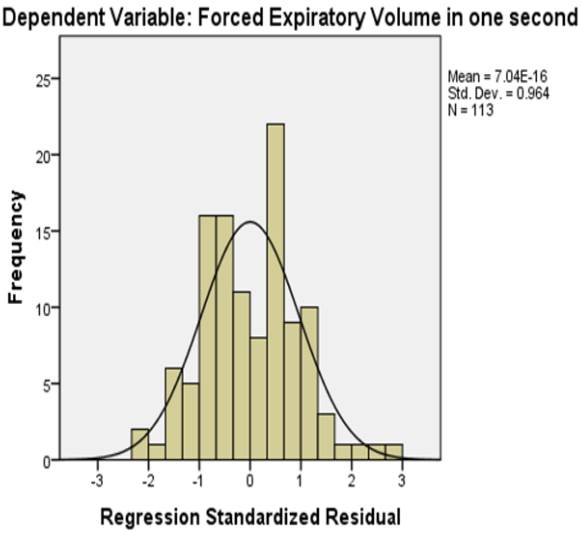

Figure 2.4: *The normality of residuals*

The ANOVA table (2.7) reports how well the regression equation fits the data, i.e. predicts the dependent variable. We note from this table that the sig. value (the statistical significant of the regression model that was run) is less than 0.05, and indicates that the regression model statistically significantly predicts the outcome variable(i.e.,it is good fit for the data).

Source	SS	df	MS	F	Sig
Regression	24.870	8	3.109	9.483	000
Error	34.094	104	0.328		
Total	58.964	112			

Table 2.7: *ANOVA table for multiple regression*

Based on this regression, we can predict the expected (FEV1) for any person as long as we know his age, smoking years, working years, height, weight, and the type of work he do in

the shoe factory.
We note from this regression that FEV1 decreases as age, smoking years, working years, weight, adhesive, and painting increase. But increases as height and molding increase.

CHAPTER 3

Bayesian Model

3.1 Introduction

Bayesian statistics, named for Thomas Bayes (1701 - 1761), as defined in (Edwards et al., 1963) is a theory in the field of statistics in which the evidence about the true state of the world is expressed in terms of degrees of belief known as Bayesian probabilities. Such an interpretation is only one of a number of interpretations of probability and there are other statistical techniques that are not based on 'degrees of belief'. One of the key ideas of Bayesian statistics is that probability is orderly opinion, and that inference from data is nothing other than the revision of such opinion in the light of relevant new information.

The purpose of bayesian analysis is to revise and update the

initial assessment of the event probabilities generated by the alternative solutions. This is achieved by the use of additional information. So the bayesian statistic consider θ to be a quantity whose variation can be described by a probability distribution called the prior distribution, which is formulated before the data are seen, and then this prior distribution is updated with the sample information θ. The updated prior is called the posterior distribution, which different from classical statistic (frequency statistic) that considers the parameter θ to be an unknown, but fixed quantity.
The prior distribution can be chosen to represent the beliefs of the researcher before observing the results of an experiment; this results in a proper subjective bayesian analysis. Often, however, it is difficult for a researcher to specify prior beliefs about model parameters, and to cast them into the form of a prior probability distribution.
Keying Ye and Wheeler (1999) defines a noninformative prior as is a function which is used in place of a subjective prior distribution when little or no prior information is available. The term noninformative is used to connote the lack of subjective beliefs used in formulating such a prior. However, one can think of a noninformative prior as simply being a function that is formally used in place of a subjective prior distribution, for the purpose of accomplishing some goal.

3.2 Advantages and disadvantages of Bayesian model

We want to state some advantages to use Bayesian analysis :

- It provides a natural and principal way of combining prior

information with data, within a solid decision theoretical framwork.

- Small sample inference proceeds in the same manner as if one had a large sample.
- It provides conditional and exact inference on the data without reliance on asymptotic approximation.
- It obeys the likelihood principle, while the classical inference does not in general obey it.
- It provides interpretable answers.
- It provides a convenient setting for a wide range of models, such as MCMC.
- Bolstad (2007) states that Bay's theorem gives the way to find the predictive distribution of future observations but this is not always easily done in a frequentist way.

as any models, there is some disadvantages to use Bayesian analysis:

- There is no correct way to choose a prior .
- It can produce posterior distributions that are heavily influenced by the priors.
- It often comes with a high computational cost, especially in models with a large number of parameters.

3.3 Some applications on Bayesian model

Explicitly (Spiegelhalter and Rice, 2009) stated that Bayesian statistical methods tend to be used in three main situations:

- The first situation is if one has no alternative but to include quantitative prior judgments, due to lack of data on some aspect of a model, or because the inadequacies of some evidence has to be acknowledged through making assumptions about the biases involved. These situations can occur when a policy decision must be made on the basis of a combination of imperfect evidence from multiple sources, an example being the encouragement of Bayesian methods by the Food and Drug Administration (FDA) division responsible for medical devices.

- The second situation is with moderate-size problems with multiple sources of evidence, where hierarchical models can be constructed on the assumption of shared prior distributions whose parameters can be estimated from the data. Common application areas include meta-analysis, disease mapping, multi-centre studies, and so on.

- The third area concerns where a huge joint probability model is constructed, relating possibly thousands of observations and parameters, and the only feasible way of making inferences on the unknown quantities is through taking a Bayesian approach: examples include image processing, spam filtering, signal analysis, and gene expression data.

3.4 The Model Description

Bayesian analysis uses the posterior distribution to form various summaries for the model parameters including point estimates such as posterior means, medians, percentiles, and interval estimates such as credible intervals. Moreover, all statistical tests about model parameters can be expressed as probability statements based on the estimated posterior distribution.

We can state bayes rule as

$$p(\text{hypothesis}|\text{data}) = \frac{p(\text{data}|\text{hypothesis})p(\text{hypothesis})}{p(\text{data})}$$

which tells us how to make inferences about hypotheses from data.

The posterior distribution is done with the use of this rule

$$P(A_i|B) = \frac{P(B|A_i)p(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)} \tag{3.1}$$

where A_1, A_2, ... be a partition of the sample space. B is any set. which is called also Bayes rule.
If we denote the prior distribution by $\Pi(\theta)$ and the sampling distribution by $f(x|\theta)$, then the posterior distribution, the conditional distribution of θ given the sample x as mentioned in (George, 2008) is

$$\Pi(\theta|x) = \frac{f(x|\theta)\Pi(\theta)}{g(x)}, \tag{3.2}$$

where $f(x|\theta)\Pi(\theta) = f(x, \theta)$, $g(x)$ is the marginal distribution of X, that is

$$g(x) = \int f(x|\theta)\Pi(\theta)d\theta.$$

and $f(x|\theta)$ is the maximum likelihood estimation which is the most commonly used method of estimating parameters and determining the extent of error in the estimation in social science statistics.

There are three types of noninformative priors Jeffreys' prior, the reference prior method and probability matching priors. Here we need to talk about jeffreys' prior, jeffreys' prior, named after Sir Harold Jeffreys, is one of the earliest methods of defining noninformative priors was based on the principle of insufficient reason. This method which found in 1961, sometimes referred to as Laplace's rule, prescribes a uniform prior on the parameter space. jeffreys proposed

$$\Pi(\theta) \propto |I(\theta)|^{\frac{1}{2}}$$

where $I(\theta)$ is the expected Fisher information matrix:

$$I(\theta) = -E_\theta[\frac{d^2}{d\theta^2} logp(y|\theta)]$$

Jeffreys' prior, like the uniform prior, may be improper. However, Jeffreys' prior is invariant, in the sense that if the Jeffreys' prior in one parameterization is transformed to a different parameterization, then the transformed prior will be the Jeffreys' prior in the new parameterization.

There are three general steps for bayesian modeling:

- Specify a probability model for unknown parameter values that includes some prior knowledge about the parameters if available.
- Update knowledge about the unknown parameters by conditioning this probability model on observed data.

- Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions.

Recall that the matrix notation $\boldsymbol{Y} = \boldsymbol{X\beta} + \epsilon$; $\epsilon \sim N(0, \sigma^2)$ where $\boldsymbol{X} = [x_1, x_2, ...x_p]$ and $\boldsymbol{Y} = (y_1, y_2, ...y_n)$ is the linear model of regression. Also, we proved that $\boldsymbol{b}$ is unbiased estimator in chapter (2) and $\boldsymbol{b} = (x'x)^{-1}x'y$.
Now, for the variance σ^2 (Banerjee, 2010) found that

$$s^2 = \frac{1}{n-k}(y - xb)'(y - xb).$$

where $k = p + 1$ is the number of columns of x.

The posterior distribution has two components: a likelihood, which includes information about model parameters based on the observed data, and a prior, which includes prior information (before observing the data) about model parameters. The likelihood and prior models are combined using the Bayes rule to produce the posterior distribution:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}.$$

Banerjee (2010) assumes the priors on β and $\log \sigma^2$, and then

$$p(\beta) \propto 1; p(\sigma^2) \propto \frac{1}{\sigma^2}$$

equivalently

$$p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$$

So,the joint posterior distribution for β and σ^2 as

$$p(\beta, \sigma^2|Y) = P(\beta|\sigma^2, y) \times P(\sigma^2|y)$$

but

$$p(\beta|\sigma^2, y) \propto p(y_i|\beta).p(\beta).$$

We assumed previously that ϵ_i are normally distributed with mean 0, variance σ^2 and ϵ_i and ϵ_j are independent $\forall\, i \neq j$. Also the x_i are fixed i.e., not random variables. Alston et al. (2013) found that $p(y_i|\beta, \sigma^2)$ is normal as it is a linear combinations of normals.
Then we have $E(y_i|\beta, \sigma^2) = \beta x_i$ and var$(y_i|\beta, \sigma^2) = \sigma^2$.

Now, the normal probability density function pdf with mean and variance as mentioned above is

$$p(y_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(y_i - \beta x_i)^2}{2\sigma^2}]$$

Now, since ϵ_i and ϵ_j are independent of one another for $i \neq j$, then y_i and y_j are also independent of one another.
The likelihood estimator of β is $b = (x'x)^{-1}x'y$ which is the same estimator as the one found in sec.(2.3)
So, $\beta|\sigma^2, y \sim N((x'x)^{-1}x'y, \sigma^2(x'x)^{-1})$. But we must find the marginal posterior distribution of σ^2.
Recall that

$$s^2 = \frac{1}{n-k}(y - xb)'(y - xb).$$

then (Banerjee, 2010) found the marginal probability density function (pdf) σ^2 belong to the inverse gamma (IG) distribution i.e.

$$\sigma^2|Y \sim IG(\frac{n-k}{2}, \frac{(n-k)s^2}{2})$$

which is the classical unbiased estimate of σ^2 in the linear regression model.
Now, we can find the marginal posterior distribution of β by

integrating out σ^2 as:

$$p(\beta, Y) = \int p(\beta|\sigma^2, Y)p(\sigma^2|Y)d\sigma^2.$$

this integration compute by the Marcov Chain Monte Carlo (MCMC) method which is defined by(Lynch, 2006) as is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest.
A Marcov Chain is a sequence of random variables, in which each random variable depends on the previous one. (i.e. generate a finite set of points in some parameters space that are drawn from a given distribution function).

Monte Carlo, as in Monte Carlo integration is used to approximate an expectation by using the Marcov Chain samples, and provide approximate solutions to a variety of mathematical problems by performing statistical sampling experiments on a computer, and it is a simple, fast, intriguingly educational method for solving stochastic (random) system problems.
This simulation used in any system or situation with some random variables, i.e production lines, tolerancing, instrumentation data handling, test schedules, bus schedules, trimming controls, waiting lines, etc.
We can summarize this method as $MCMC$ is an iterative procedure, such that given the current state of the chain, $\theta^{(i)}$,the algorithm makes a probabilistic update to $\theta^{(i+1)}$.
before any data is observed the distribution of unknown but observable y is

$$p(y) = \int \int p(\beta, \sigma^2)p(y|\beta, \sigma^2)d\beta d\sigma^2.$$

which is the marginal or prior predictive distribution of y.

Now, after y is observed the posterior predictive distribution of $\tilde{y}$ is:

$$p(\tilde{y}|y) = \int \int p(\tilde{y}|\beta, \sigma^2) p(\beta, \sigma^2|y) d\beta d\sigma^2.$$

since β and σ^2 were known, then

$$\tilde{y} \sim N(\tilde{X}\beta, \sigma^2 I)$$

where $\tilde{X}$ is the covariate matrix.

3.5 Applications on Bayesian model using Real Data Set

We illustrate a practical issue of simulation by fitting an example which described in section (2.9). After some background in section (2.9) we show in this section the results of fitting the model using the Bayesian inference package $winBugs$ (which based on bugs) operating from within the general statistical package R.

Gelman et al. (2003) defined $Winbugs$ as a software for Bayesian analysis using (MCMC) methods, it runs under microsoft windows.

After applying the Bayesian regression model to the industrial data as reported in sec (2.9). In order to compute the Bayesian estimates of the regression model, the model was implemented in Winbugs, with 4000 iterations as a burn-in, a thinning of 50 iterations, and a final sample size of 5000 iterations. The chains passed most of the standard convergence tests. As a first result, table (3.1) reports the posterior mean and 95% credible interval of the regression coefficients.

Table 3.1: *The posterior mean and their 95% credible interval*

	Posterior mean	Lower Bound	Upper Bound
b_0	2.391	-0.133	4.918
b_1	-0.036	-0.048	-0.023
b_2	-0.002	-0.015	0.011
b_3	-0.002	-0.021	0.017
b_4	0.015	-0.0004	0.031
b_5	-0.001	-0.01	0.008
b_{62}	-0.002	-0.383	0.393
b_{63}	-0.134	-0.457	0.186
b_{64}	0.009	-0.281	0.298

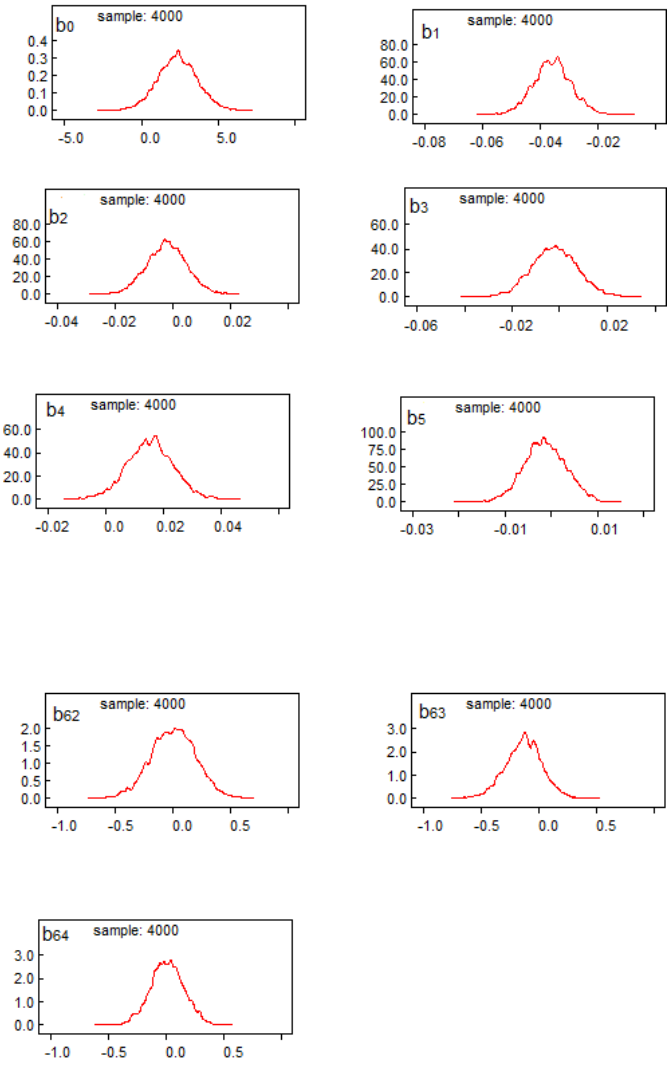

Figure 3.1: *The posterior distribution of the estimators*

It can be shown from fig (3.1) the density for the posterior coefficients β_j that is represents our prior beliefs about the explanatory variables.

3.6 Comparison Between Frequentist and Bayesian approaches

Perhaps a better question is when to use Bayesian analysis and when to use frequentist analysis. The answer to this question mainly lies in our research problem. i.e, the specific research questions determine which analysis is better. For example, if we are interested in estimating the probability that the parameter of interest belongs to some prespecified interval, we will need the Bayesian framework, because this probability cannot be estimated within the frequentist framework. If we are interested in a repeated-sampling inference about our parameter, the frequentist framework provides that. It is important to state some differences between frequency and Bayesian statistic:

- As we mentioned above the underlying parameters remain constant during the repeatable process in frequntist, while they are described probabilistically in Bayesian statistic.
- In frequentist the data are repeatable random sample i.e there is frequency, where in Bayesian the data are observed from the realized sample.
- We conclude from previous differences that the parameters are fixed in frequentist but the data are fixed in Bayesian.
- Prior information abound and it is important and helpful to use it in Bayesian, while no prior information to the model specification in frequentist.

  But when the sample size is large, the results of parametric models which provides by bayesian inference often be

vary similar to the results produced by frequentist methods.

- The interpretation of a 95% confidence interval in frequentist is that if we repeat the same experiment many times and compute confidence intervals for each experiment, then 95% of those intervals will contain the true value of the parameter. But 95% Bayesian credible interval provides a range for a parameter such that the probability that the parameter lies in that range is 95%.

Table (3.2) shows the output of our approaches,

Table 3.2: *The coefficients and their 95% confidence interval for two approaches*

	Frequency			Bayesian		
		Lower Bound	Upper Bound	Posterior mean	Lower Bound	Upper Bound
b_0	2.386	-0.163	4.934	2.391	-0.133	4.918
b_1	-0.035	-0.048	-0.023	-0.036	-0.048	-0.023
b_2	-0.002	-0.015	0.012	-0.002	-0.015	0.011
b_3	-0.003	-0.022	0.015	-0.002	-0.021	0.017
b_4	0.015	-0.001	0.031	0.015	-0.0004	0.031
b_5	-0.001	-0.01	0.008	-0.001	-0.01	0.008
b_{62}	-0.007	-0.396	0.383	-0.002	-0.383	0.393
b_{63}	-0.139	-0.458	0.179	-0.134	-0.457	0.186
b_{64}	0.010	-0.279	0.298	0.009	-0.281	0.298

We have now seen how Bayesian methods work in a simple case:the Normal linear regression model with multiple regression - many explanatory variables and natural conjugate prior. The credible interval for β_j smaller range than the confidence interval using frequentist approach.

Bibliography

Alston, C. L., Mengersen, K. L., and Pettitt, A. N. (2013). *Case Studies in Bayesian Statistical Modelling and Analysis*. Wiley and Sons.

Banerjee, S. (2010). *The Bayesian linear regression*. University of Minnesota.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. Wiley and Sons.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis, Third Edition (Wiley Series in Probability and Statistics)*. John Willy and Sons, third edition. Available from: `http://gen.lib.rus.ec/book/index.php?md5=903983CE6C8FBCBB1F81265002FE576B`.

Edwards, Ward, Lindman, Harold, and J, S. L. (1963). *Bayesian Statistical Inference for Psychological Research*. American Psychological Association.

Frost, J. (2017). Multicollinearity in regression analysis. *Statistics By Jim*.

Fu, H., Demers, P. A., Costantini, A. S., Winter, P., Colin, D., Kogevinas, M., and Boffetta, P. (1996). Cancer mortality among shoe manufacturing workers: an analysis of two cohorts. 53(6):394–398.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. CRC Press.

George, C. (2008). *Monte Carlo Statistical Methods*. Springer.

Joaquim, PMDS, and S, M. (2007). Applied statistics using spss, statistica, matlab and r. *Springer*.

Bibliography

Kewan, J. N. S. (2015). Estimation of multivariate multiple linear regression models and applications. Master's thesis, An Najah National University.

Keying Ye, a. C. J. C. and Wheeler, A. E. P. S. G. R. T. R. L. (1999). *Noninformative Prior Bayesian Analysis for Statistical Calibration Problems*. Daniel R. Eno.

Krehbiel, B. L. (2008). *Basic Business Statistics*. Prentice Hall.

Lynch, S. M. (2006). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer.

Montgomery, D., A.Aeck, E., and Vining, G. (2012). *Introduction to Linear Regression Analysis*. Wiley and Sons.

Myers, R. H. (2000). *Classical and Modern Regression with Applications*. Duxbury.

Simon, L. (2004). Detecting multicollinearity using variance inflation factors. *Penn State Department of Statistics, The Pennsylvania State University*.

S.Kramer, M. (1988). *Clinica Epidemiology and Biostatistics*. Springer Verlag.

Spiegelhalter, D. and Rice, K. (2009). Bayesian statistics. *Scholarpedia*, 4(8):5230. revision #91036.

Statistics, P. C. B. O. (2015). Finance and insurance survey. *Palestine Economy Portal*.

Xin, Y. and Xiaogang, S. (2009). *Linear regression analysis: theory and computing*. World Scientific.