Report on Linear Regression

Course: 837 Machine Learning

Group: VIII

**Submitted by**

Amran Hossain (BSSE 0917)

Niraj Chaudhary (BSSE 0942)

[Zahid Hasan Rifad](mailto:bsse0820@iit.du.ac.bd) (BSSE 0944)

**Submitted To**

Dr. B M Mainul Hossain

Associate Professor

Institute of Information Technology

University of Dhaka



Institute of Information Technology

University of Dhaka

Date: 9 Jan, 2021

Table of Contents

[Dataset Description: 3](#_Toc60937304)

[Model-1: 3](#_Toc60937305)

[Title: 3](#_Toc60937306)

[R^2 calculation: 3](#_Toc60937307)

[Model-2: 4](#_Toc60937308)

[Title: 4](#_Toc60937309)

[Difference between previous model: 4](#_Toc60937310)

[R^2 calculation: 4](#_Toc60937311)

[Model-3: 5](#_Toc60937312)

[Title: 5](#_Toc60937313)

[Difference between previous model: 5](#_Toc60937314)

[R^2 calculation: 5](#_Toc60937315)

[Model-4: 6](#_Toc60937316)

[Title: 6](#_Toc60937317)

[Difference between previous model: 6](#_Toc60937318)

[R^2 calculation: 6](#_Toc60937319)

Linear Regression

Linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables)

Dataset Description

We have chosen a dataset which contain diabetes type1 data and pick BMI related data for fitting the linear regression. There are 307 rows and 22 columns in the main dataset. For our purpose we have chosen 6 columns. These columns belongs to height, weight, BMI, gender, area and age. Also, there are 3 categorical variables such as gender, area, age. We have taken BMI as a y variable from our dataset.  
  
Model  
We have built many regression models from our dataset. There are four models in details.

# Model-1

## Title:

Here

Y=

X =

## R^2 calculation:

The R^2 value of this model is **12%**

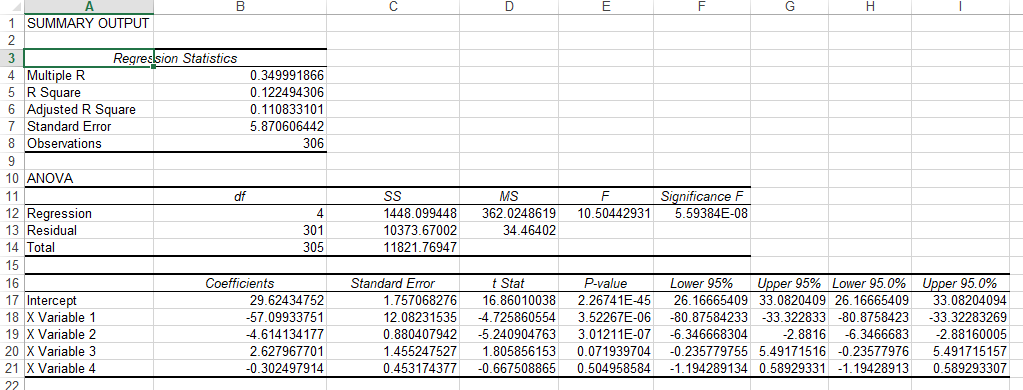


Figure 1: result of model 1

# Model-2:

## Title:

Here

Y=

X = ,

## Difference between previous model:

Previous model y variable was BMI but, in this model, we have used e base log of BMI. Independent variable we used (height)^2, weight, age and ln(age)^2 but previous model we used 1/weight, height^2, ln(age), area. For this reason, R^2 value has been increased.

## R^2 calculation:

The R^2 value of this model is **32%**

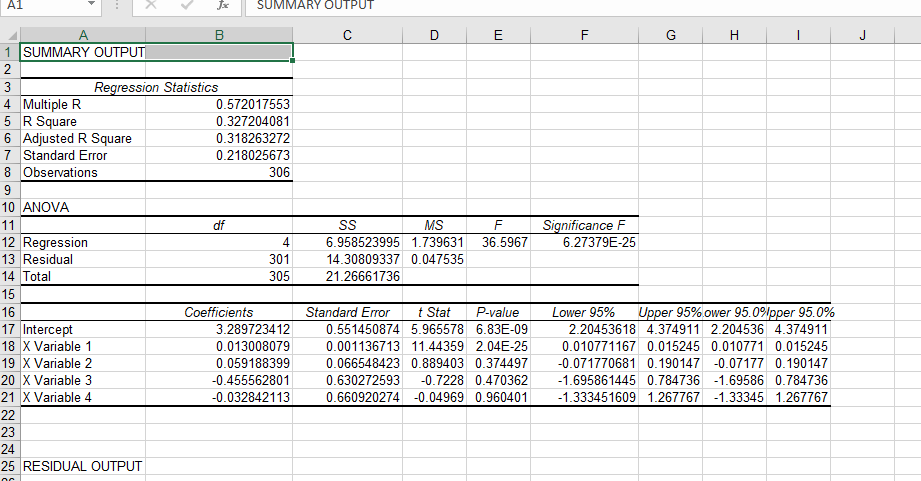


Figure-2: result of model 2

# Model-3:

## Title:

Here

Y=

X = ,

## Difference between previous model:

Previous model y variable was log of BMI but in this model, we have used normal BMI. Independent variable we used (height)^2 and weight but previous model we used age and log of age^2. For this reason, R^2 value has been increased.

## R^2 calculation:

The R^2 value of this model is **69%**

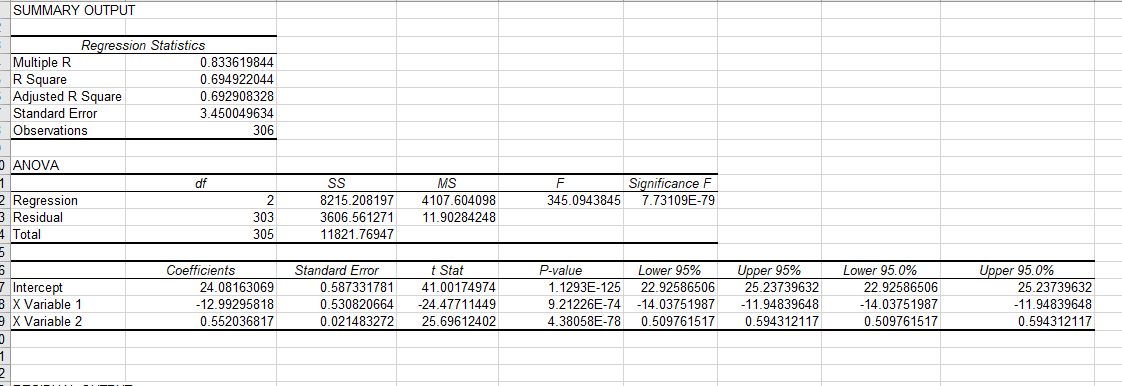
****

Figure-3: result of model 3

# Model-4:

## Title:

Here

Y=

X =

## Difference between previous model:

Previous model y variable was normal BMI but, in this model, we have used e base log of BMI. Independent variable we used height^2 and weight in previous model but in this model, we used e base log of all independent variables. E base log of (height, weight, gender, age, Residence). R^2 value has been increased.

## R^2 calculation:

The R^2 value of this model is **89.8%**

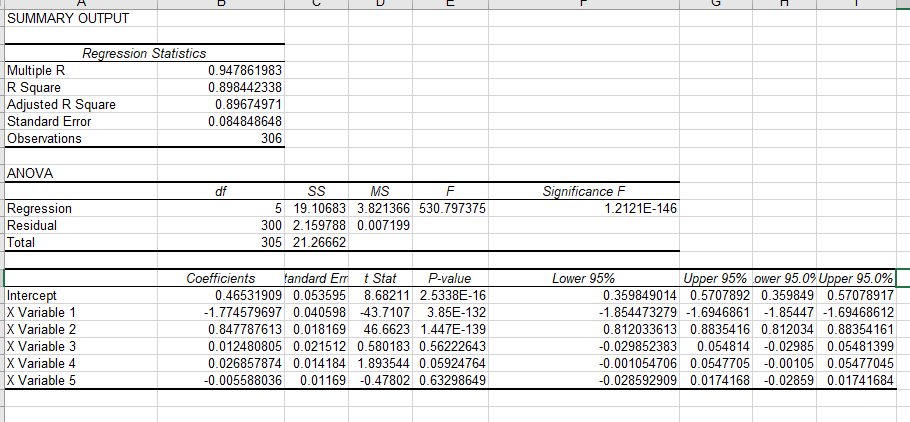
****

Figure-4: result of model 4