Project Unit 4:
Linear Regression:Prediction
Subject: Programming

Teacher: Didier Gamboa.

Autores:

Hernandez Escalante Hector

Catzin Cetz Jesus Alejandro

August 9 of 2019

1. Abstract

Linear Regression
One way to se the simple linear regression, is that is one way to predect one result using ecuations that are aplicated in the real life (this find the relation between two variables or more that can be dependent variables or independen) The variable we are predicting is called the criterion variable and is referred to as Y. The variable we are basing our predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called simple regression. In the end, linear regresion is one of the most techniques use to predict result in different enviroments as economy, industries, taking of decisitions etc.

2. Introduction

To begin with all, in this project we are going to program a code to solve and predict the results of data set with the Lineal Regression simple

The linear regression , as its name says, It is a straight line that across of a set of points in the cartesian plane that can adapt in the best way between them and how this can help us to make predictions or take decisions taking some intervals where the predictions or values can be and all of this is because the linear regression is a Statistics technique that uses the relation between the variables, in this case, the independent variable "X" and the dependent variable "Y" and in many cases more variables (multiple variable linear regression) that in this project is not going to be doing. However, in the problems of the investigation the simple linear regression is not enough to solve that, and for that reason is used the multiple variable linear regression that has more than one independent variable and this allows us to predict a better answer, but it is not going to be implemented

We are going to use the regression coefficients that allow us to make the best regression depending on our data and for this, we are going to use the Least Squares compute the error that both use the simple linear regression or multiple variable linear regression.

In this project, 54 data were taken about bears such as weight, size, age taken in months, sex, head weight, head measurement, chest measurement and neck measurement in inches. Based on these data, the neck and height measurement data were taken to study if there is a correlation and if it is possible to make a linear regression.

3. Motivation

The motivation to do this problem is that with the lineal regression we are using datasets as we said before and we can learn about the relation between the independent and dependent variable (including that this is a statistic technique) and how that relationship can help us to make predictions in a base of that.
Other motivations that why we do this, is that the linear regression can be used in different situations or fields as Social Investigation to compute the analyze of economic measure in the different aspects of the human behavior, also in marketplaces to find what is the best way to invest money or predict the amount of sell of a product, and that's not all, because also we can use in physics, mathematics, chemistry, biologic, etc. to find the relation between the variables and to compute or calibrate measures.
To conclude we can say that the best motivation that we get to do this, is that we can compute different things in a lot of fields, and that this, in the end, can help us to solve problems of the actuality, to avoid future problems or future mistakes, to help the nature or environment and even include to us with the predictions.

4. Objetive

Our objectives are:
Program a code to predict results using a model of simple lineal regression (this include the computing of different formulas as the Coefficient of Pearson, the prediction of the SCT, R square, Sxx and the Prediction Interval)

5. Problem Description

In a data collection in a bear research center, the problem arose that there is a relationship between the measurement of the neck and its length. Using the following 54 data samples, perform the following:

6. Table data

Sample	Neck	Lenght
1	16	53
2	28	67.5
3	31	72
4	31.5	72
5	22	66
6	21	70
7	26.5	73.5
8	27	68.5
9	20	64
10	18	58
11	29	73
12	13	37
13	10.5	63
14	21.5	67
15	17.5	52
16	21.5	59
17	24	64
18	12	36
19	19	59
20	30	72
21	19	57.5
22	20	61
23	17	54
24	13	40
25	24	63
26	13.5	43
27	22	66.5
28	17.5	60.5
29	21	60

Sample	Neck	Lenght
30	20	61
31	16	40
32	28	64
33	26	65
34	17	49
35	17	47
36	21	59
37	27	72
38	24	65
39	21.5	63
40	28	70.5
41	16.5	48
42	19	50
43	28	76.5
44	15	46
45	23	61.5
46	23	63.5
47	15.5	48
48	15	41
49	17	53
50	15	52.5
51	13	46
52	10	43.5
53	30.5	75
54	18	57.3

7. Proposed Solution

7.1. Formulas To compute the Lineal Regression

Lineal Regression Model

$$\hat{y} = b_0 + b_1 x,$$

Error

$$e_i = y_i - \hat{y}_i,$$

Sum of the Square of the Error

$$SCE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2.$$

Computing the coefficient b1 and b0

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

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}.$$

Computing the Sxx
(Variance of sample X)

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

S Square

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

Relation Coeficient:

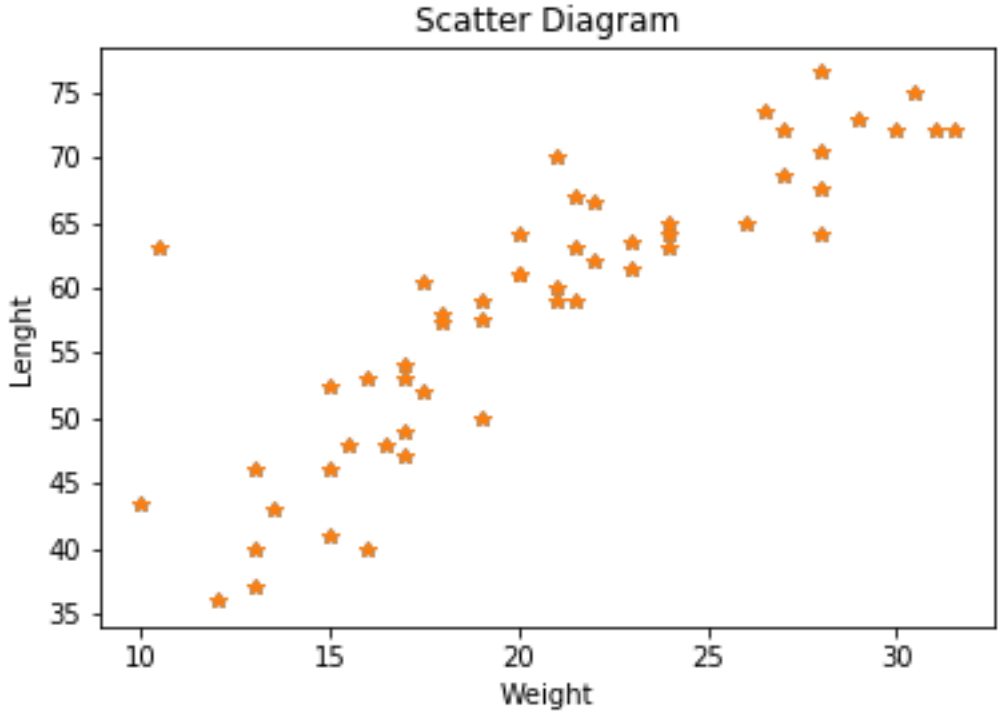
$$R^2 = 1 - \frac{SCE}{STCC}.$$

Prediction Interval

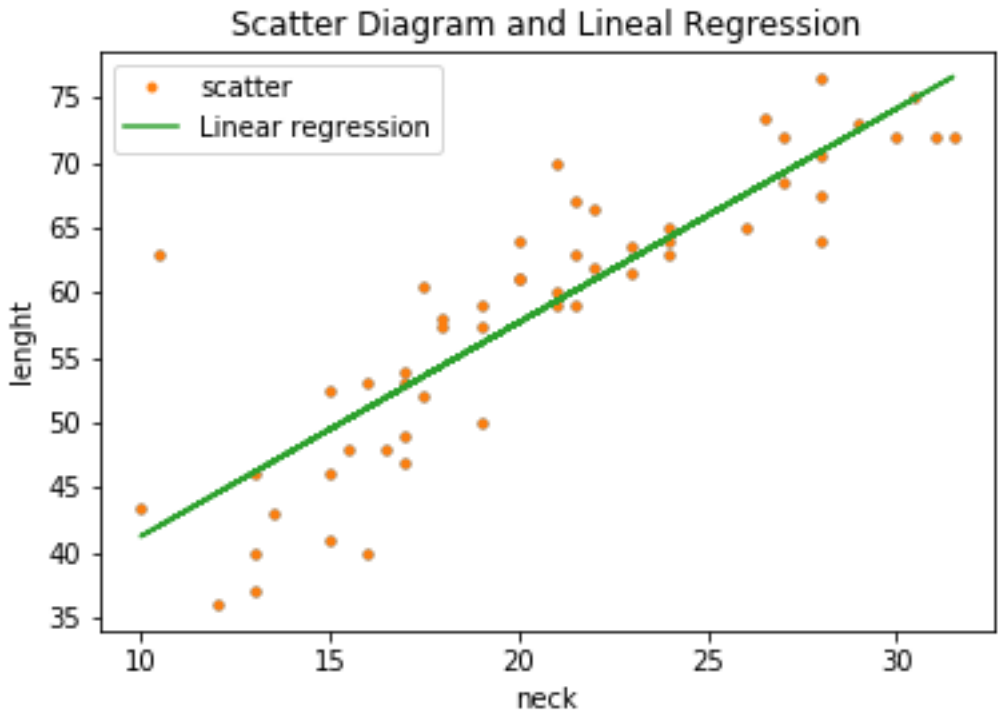
$$\hat{y}_0 - t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} < y_0 < \hat{y}_0 + t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}},$$

8. Results

Graph the data.



Fit a linear regression model and calculate R2.



R2 =75.23To make a prediction:
We used a confidance interval as alpha equal to 0.05 and we want to predict what is the lenght of the bear if the bear has a neck measure of 12 inches.
We got Interval Prediction:
34.2414 ¡y0 ¡54.8353

What can you say about the neck measure and the size of the bears?
R= We can see that the relation of the bears neck and lenght have a good correlation and also we have a positive linear regresion that we get from the pearson value :
R=0.8674
And the equation lineal regresion is:
lenght = 24.7920 + 1.6455(neck)
With this we can a better prediction with 75

9. References

Bibliografy

Brownlee, J. (30 de June de 2018). machinelearningmastery.com. Obtenido de https://machinelearningmastery.com/how-to-code-the-students-t-test-from-scratch-in-python/

Developers, S. (2006). Scipy. Obtenido de https://docs.scipy.org/doc/

Ronal E. Walpole Raymond H Miyers, S. L. (2012). Probabilidad y Estadistica para Ingeniera y Ciencias. Mexico: Pearson Education.