# Analisis y documentacion de practica 3A: Regresion Lineal

Angel Cruz

October 2022

**Abstract**

En este documento se desglozará los detalles importantes durante el desarrollo de la practica 3A. Tales como los antecedentes y las bases para el calculo de la regresion lineal y el coeficiente de correlacion.

Keywords: regresion lineal, coeficiente de correlacion, coeficiente R, formula

# 1 Introduction

## 1.1 Objectives

Calculo de la regresion lineal asi como de sus coeficientes de correlacion, aplicando las pruebas necesarias sugeridas dentro de los "issues" dentro del repositorio de GitHub

## 1.2 Problem Statement

Durante el calculo estadistico, la regresión lineal permite predecir el comportamiento de una variable (dependiente o predicha) a partir de otra (independiente o predictora). Esto resulta bastante util a la hora de necesitar alguna predicción para algun evento. Durante esta practica, la regresion lineal nos ayuda a predecir el como varian los tiempos de programacion de lineas de codigo en relacion a las horas de desarrollo. Aplicando las formulas necesarias se espera llegar a las pruebas deseadas

# 2 State of the art

## 2.1 Previous IoT architectures

La regresion lineal es un calculo es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, m variables independientes Xi y un término aleatorio. Este método es aplicable en muchas situaciones en las que se estudia la relación entre dos o más variables o predecir

un comportamiento La formula para aplicarla es: $y = \beta_0 + \beta_1 x$
Para calcular los valores de B0 y B1 se aplican las siguientes formulas:

$$\beta_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - \left(n x_{avg} y_{avg}\right)}{\left(\sum_{i=1}^{n} x_i^2\right) - \left(n x_{avg}^2\right)}$$

$$\beta_0 = y_{avg} - \beta_1 x_{avg}$$

El coeficiente de correlación es la medida específica que cuantifica la intensidad de la relación lineal entre dos variables en un análisis de correlación. En los informes de correlación, este coeficiente se simboliza con la r.
Las formulas para calcular los coeficientes Rxy y r2 son:

$$r_{x,y} = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$

$$r^2 = r * r$$

# 3 Methodology

## 3.1 General architecture

Para el desarrollo de la practica, se utilizaron las formulas anteriormente mencionadas. Aplicadas a la tabla de la relacion de horas y desarrollo siguiente:

| Program Number | Estimated Proxy Size | Plan Added and Modified size | Actual Added and Modified Size | Actual Development Hours |
|---|---|---|---|---|
| 1 | 130 | 163 | 186 | 15.0 |
| 2 | 650 | 765 | 699 | 69.9 |
| 3 | 99 | 141 | 132 | 6.5 |
| 4 | 150 | 166 | 272 | 22.4 |
| 5 | 128 | 137 | 291 | 28.4 |
| 6 | 302 | 355 | 331 | 65.9 |
| 7 | 95 | 136 | 199 | 19.4 |
| 8 | 945 | 1206 | 1890 | 198.7 |
| 9 | 368 | 433 | 788 | 38.8 |
| 10 | 961 | 1130 | 1601 | 138.2 |

Aplicamos los test correspondientes a la practica, los cuales son los siguentes:
Test 1:
Description: Calculate the regression parameters between estimated proxy size and actual added and modified size in data table
Calculate the linear regression given an estimated proxy of Xk = 386

-Should return B1= 1.7279, if X=[130, 650, 99, 150, 128, 302, 95, 945, 368, 961] and Y=[186, 699, 132, 272, 291, 331, 199, 1890, 788, 1601]
-Should return B0= -22.55, if B1=1.7279 and X=[130, 650, 99, 150, 128, 302, 95, 945, 368, 961] and Y=[186, 699, 132, 272, 291, 331, 199, 1890, 788, 1601]
- Should return yk = 644.429 if xk = 386 and B0=-22.55 and B1=1.7279

Test 2:
Description: Calculate the regression parameters between estimated proxy size and actual development time in data table
Calculate the linear regression given an estimated proxy of Xk = 386

-Should return B1 = 0.1681, if X=[130, 650, 99, 150, 128, 302, 95, 945, 368, 961] and Y=[15.0, 69.9, 6.5, 22.4, 28.4, 65.9, 19.4, 198.7, 38.8, 138.2]
-Should return B0 = -4.039, if B1 = 0.1681 and X=[130, 650, 99, 150, 128, 302, 95, 945, 368, 961] and Y=[15.0, 69.9, 6.5, 22.4, 28.4, 65.9, 19.4, 198.7, 38.8, 138.2]
-Should return yk = 60.858, if xk = 386 and B0=-4.039 and B1=1.7279

Test 3
Calculate the regression parameters between plan added and modified size and actual added and modified size in data table
Calculate the linear regresion given an estimated proxy of Xk = 386

-Should return B1 = 1.43097, if X=[163, 765, 141, 166, 137, 355, 136, 1206, 433, 1130] and Y=[186, 699, 132, 272, 291, 331, 199, 1890, 788, 1601]
-Should return B0 = -23.92, if B1 = 1.43097 and X=[163, 765, 141, 166, 137, 355, 136, 1206, 433, 1130] and Y=[186, 699, 132, 272, 291, 331, 199, 1890, 788, 1601]
-Should return yk =528.4294, if xk = 386 and B0=-23.92 and B1=1.43097

Test 4

Calculate the regression parameters between plan added and modified size and actual develpment time in data table
Calculate the linear regresion given an estimated proxy of Xk = 386

-Should return B1 = 0.140164, if X=[163, 765, 141, 166, 137, 355, 136, 1206, 433, 1130] and Y=[15.0, 69.9, 6.5, 22.4, 28.4, 65.9, 19.4, 198.7, 38.8, 138.2]
-Should return B0 = -4.604, if B1 = 0.140164 and X=[163, 765, 141, 166, 137, 355, 136, 1206, 433, 1130] and Y=[15.0, 69.9, 6.5, 22.4, 28.4, 65.9, 19.4, 198.7, 38.8, 138.2]
-Should return yk =49.4994, if xk = 386 and BO=-4.604 and ‘‘=0.140164

# 4 Results

Como resultado final deberiamos obtener los datos dados en la siguente tabla:

| **Test** | **Expected Values** | | | | |
|---|---|---|---|---|---|
| | $\beta_0$ | $\beta_1$ | $r_{x,y}$ | $r^2$ | $y_k$ |
| Test 1 | -22.55 | 1.7279 | 0.9545 | 0.9111 | 644.429 |
| Test 2 | -4.039 | 0.1681 | 0.9333 | .8711 | 60.858 |
| Test 3 | -23.92 | 1.43097 | .9631 | .9276 | 528.4294 |
| Test 4 | -4.604 | 0.140164 | .9480 | .8988 | 49.4994 |

# 5 Conclusion

Conclusion and future work

# 6 References