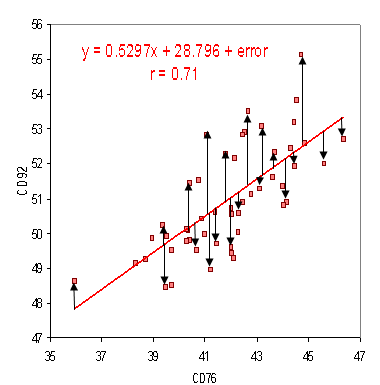
**Logistic Regression**

**What is Regression?**

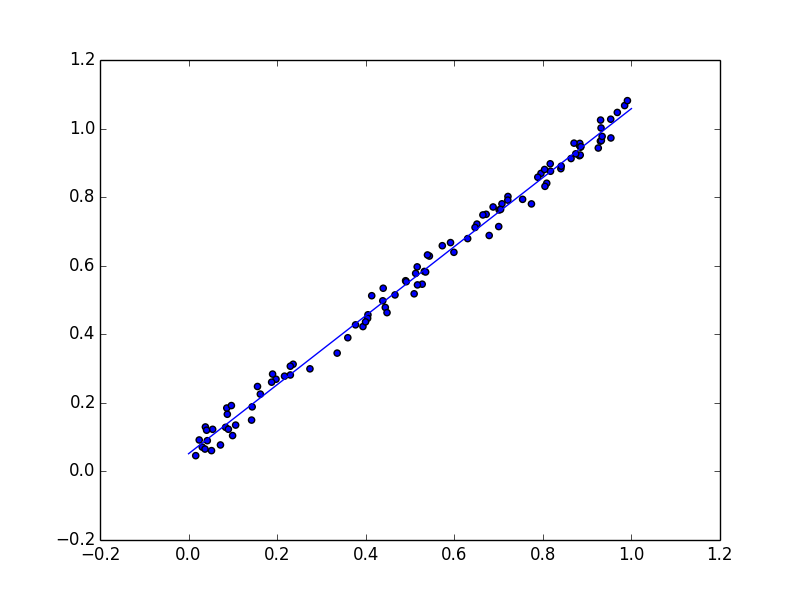
* **Regression analysis is a predictive modeling technique.**
* **It estimates the relationship between a dependent (target) and an independent variable(Predictor)**
* **Scatter plot with regression line.**

****

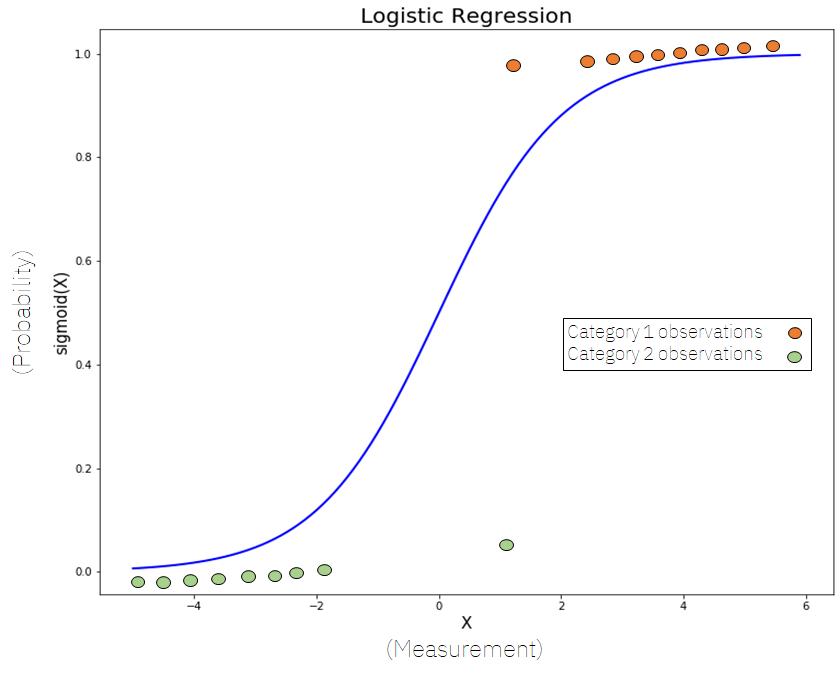
**Regression Equation: Y = 0.5297X+28.796, for any value of X , we can predict the value of Y.**

**Types of Regression**

1. **Linear Regression: When there is a linear relationship between independent and dependent variables.**

****

1. **Logistic Regression: When the dependent variable is categorical (0/1, True/False, Successful/Unsuccessful, A/B/C) in nature.**

****

**Sigmoide Curve(S-Curve)**

1. **Polynomial Regression: When the relationship between the independent and dependent variables is not linear.**

**Why Logistic Regression?**

**Whenever the outcome of the dependent variable (Y) is discrete like 0 or 1, Yes or No, A, B, C, we use logistic regression.**

**Why can’t we use linear regression?**

**Since our value of Y will be between 0 and 1 in logistic regression but in linear regression it may cross 0 or 1, so, the linear line has to be clipped at 0 and 1. With this our resulting curve cannot be formulated into a single formula. So we needed a new way to solve this kind of problem.. Hence logistic regression is required.**

**Equation for a straight line:**

**Y= β0+β1X1+β2X2+………..………. , Range of Y is from -**∞ to + ∞

Lts try to find the logistic regression from the above equation.

Y = **β0+β1X1+β2X2+………..**…………. In logistic equation Y can be only between 0 and 1.

Now, to get the range of Y between 0 to + ∞, lets transform Y

Y Y=0] 0

1-Y Y =1] ∞, Now, we have range between 0 to ∞

Let us transform it further, to get the range between - ∞ to ∞

Y

**Log = β0+β1X1+β2X2+………..……..**

**1-Y**

**What is logistic Regression?**

Logistic Regression or logit regression or logit model is a regression model where the dependent variable is categorical.

Categorical: Variables that can be only fixed values such as A,B or C , Yes or No.

Y= F(X), Y is dependent on X.

**How does logistic regression work?**

|  |
| --- |
| **IQ of Candidates**  Selected  147,120,121,128,110,119,133 |
| **110** |
| **147** |
| **120** |
| **107**    MODEL |
| **89** |
| **92** |
| **106** |
| **121** |
| **127** |
| **104**  Not Selected  107, 89, 92,106,104,114 |
| **137** |
| **133** |
| **114** |
| **126** |
| **121** |
| **119** |

Before creating the model, we divide our dataset into training data (estimation) and testing data (validation).

**Logistic Regression Equation:**

Y

**Log = β0+β1X1+β2X2+………..**

**1-Y**

**Logistic Regression Equation:**

Y e β0+β1X1+β2X2

**Logit(Y)=Log i.e. P(Y) =**

**1-Y 1+** e β0+β1X1+β2X2

**Example: Logistic Regression in R**

**Objective: To predict the patient is diabetic or not based on the following data.**

**Npreg= number of pregnancies**

**Glu= plasma glucose concentration**

**Bp=Blood Pressure**

**Skin: Triceps skin fold thickness**

**Bmi=body mass index**

**Ped =diabetes pedigree function**

**Age = Age in Years**

**Type: 1 for Yes and 0 for No diabetic**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sr No.** | **npreg** | **glu** | **bp** | **skin** | **bmi** | **ped** | **age** | **type** |
| 1 | 6 | 148 | 72 | 35 | 33.6 | 0.627 | 50 | 1 |
| 2 | 1 | 85 | 66 | 29 | 26.6 | 0.351 | 31 | 0 |
| 3 | 1 | 89 | 66 | 23 | 28.1 | 0.167 | 37 | 0 |
| 4 | 3 | 78 | 50 | 32 | 31.1 | 0.248 | 26 | 1 |
| 5 | 2 | 197 | 70 | 45 | 30.5 | 0.158 | 53 | 1 |
| 6 | 5 | 166 | 72 | 19 | 25.8 | 0.587 | 51 | 1 |
| 7 | 0 | 118 | 84 | 47 | 45.8 | 0.551 | 31 | 0 |
| 8 | 1 | 103 | 30 | 38 | 43.3 | 0.183 | 33 | 1 |
| 9 | 3 | 126 | 88 | 41 | 39.3 | 0.704 | 27 | 0 |
| 10 | 9 | 119 | 80 | 35 | 29 | 0.263 | 29 | 1 |
| 10 | 6 | 148 | 72 | 35 | 33.6 | 0.345 | 39 | 1 |
| 10 | 1 | 47 | 66 | 29 | 26.6 | 0.351 | 31 | 1 |
| 10 | 1 | 89 | 72 | 23 | 28.1 | 0.167 | 21 | 0 |
| 10 | 3 | 78 | 50 | 32 | 31.1 | 0.248 | 26 | 1 |
| 10 | 2 | 197 | 70 | 45 | 30.5 | 0.158 | 53 | 0 |
| 10 | 5 | 166 | 67 | 19 | 25.8 | 0.587 | 51 | 1 |
| 10 | 0 | 148 | 69 | 49 | 45.8 | 0.341 | 31 | 1 |
| 10 | 1 | 103 | 30 | 38 | 43.3 | 0.245 | 33 | 0 |
| 10 | 3 | 126 | 88 | 41 | 39.3 | 0.704 | 27 | 0 |
| 10 | 9 | 119 | 80 | 35 | 29 | 0.263 | 29 | 1 |
| 10 | 6 | 148 | 72 | 35 | 33.6 | 0.627 | 50 | 0 |
| 10 | 1 | 85 | 66 | 29 | 26.6 | 0.456 | 31 | 1 |
| 10 | 1 | 89 | 66 | 23 | 28.1 | 0.167 | 21 | 0 |
| 10 | 3 | 78 | 50 | 32 | 31.1 | 0.248 | 26 | 1 |
| 10 | 2 | 197 | 82 | 45 | 30.5 | 0.158 | 53 | 1 |
| 10 | 5 | 160 | 72 | 19 | 25.8 | 0.587 | 54 | 0 |
| 10 | 0 | 139 | 67 | 47 | 45.8 | 0.551 | 31 | 1 |
| 10 | 1 | 103 | 30 | 34 | 43.3 | 0.183 | 39 | 0 |
| 10 | 3 | 126 | 88 | 41 | 39.3 | 0.704 | 27 | 1 |
| 10 | 9 | 125 | 80 | 35 | 29 | 0.263 | 27 | 1 |