

## Logistic Regression: Example 1

**Question:** To Predict whether a person has disease yes/no based on his age, gender and smoking status; for the given model parameters b1, b2, b3 and a using Logistic Regression;

Given Dataset:

Sr No	Age	Gender	Smoking	Disease
1	25	0	1	1
2	37	1	0	0
3	40	0	0	0
4	49	0	1	1
5	55	1	1	1

Given model parameters

b0 = 0.523  
b1 = -0.1125  
b2 = -2.05  
b3 = 7.521

**Answer:** IV is Independent variable and DV is Dependent variable

	x1	x2	x3	y	e = 2.71828	
Sr No	Age (x1)	Gender (x2)	Smoking (x3)	Disease (y)	Calculated Y (Fitted Value)	Prediction
1	25	0	1	1	0.9947	1
2	37	1	0	0	0.0034	0
3	40	0	0	0	0.0184	0
4	49	0	1	1	0.9263	1
5	55	1	1	1	0.4518	0

Determine Predicted value of y:

Disease = Yes = 1 i.e P(y=1) and Disease = No = 0 i.e P(y=0)

$$f(z) = P(\text{Disease}) = 1 / (1 + e^{-(b_0 + b_1 \cdot \text{Age} + b_2 \cdot \text{Gender} + b_3 \cdot \text{Smoking})})$$

C' Matrix	Predicted 1	Predicted 0
Actual 1	TP	FN
Actual 0	FP	TN

C' Matrix	Predicted 1	Predicted 0	Total (N)
Actual 1	2	1	3
Actual 0	0	2	2
<b>Total (N)</b>	<b>2</b>	<b>3</b>	<b>5</b>

P, Precision or PPV =  $TP / (TP + FP) = 1.000$   
NPV =  $TN / (TN + FN) = 0.667$   
False Omission Rate (FOR) =  $1 - NPV = 0.333$   
R, Recall (Sensitivity) or TPR =  $TP / (TP + FN) = 0.667$   
Specificity or NPV or TNR =  $TN / (TN + FP) = 1.000$   
False Positive Rate (FPR) =  $FP / (FP + TN) = 0.000$   
False Negative Rate (FNR) =  $FN / (FN + TP) = 0.333$   
Accuracy =  $(TP + TN) / (TP + TN + FP + FN) = 0.800$   
F1 score =  $2 * P * R / (P + R) = 0.8$