Linear Regression Probability Model

We know by Multiple Linear Regression that:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \; - (\text{eqn(i)})$$

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} - (\text{eqn(ii)})$$

E is unconditional Expectation.

By Gaussian White Noise we know:

$\epsilon_i \approx iid\ N(0, \sigma^2)$ or $\epsilon_i \approx GW\ N(0, \sigma^2)$ $where$,

$E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$.

$Here, Var$ is $Variance$ and GWN is $Gaussian\ White\ Noise$.

From (eqn(i)) we get:

$$P(y_i) = \begin{cases} 0, & Failure \\ 1, & Success \end{cases}$$

i.e.

when :

$$P(y_i) = 1\ , when\ event\ is\ success.$$

$$P(y_i) = 0, when\ event\ is\ a\ failure.$$

Where P $stands\ for$ $probability$.

Therefore for a variable $'y'$ **:**

$$E(y_i) = 1 \times P(y_i = 1) + 0 \times P(y_i = 0)$$

i.e.

$$E(y_i) = P(y_i = 1)$$

Therefore,

$$E(y_i|x_i) = 0 \times P(y_i = 0|x_i) + 1 \times P(y_i = 1|x_i)$$

i.e.

$$E(y_i|x_i) = P(y_i = 1|x_i)$$

We know ,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

and we have:

$$E(y_i|x_i) = P(y_i = 1|x_i)$$

Therefore,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} = P(y_i = 1|x_i)$$

For 2-D model, we represent it as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

Then,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1}$$

and,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} = P(y_i = 1|x_i)$$

So we can say that,

when x = 0 :

$$\beta_0 = P(y_i = 1|x_i = 0)$$

when x increased 0 to 1 ,

$$\beta_1 x_{i1} = \frac{\partial P(y_i = 1|x_i)}{\partial x}$$

Now,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

$$=> E(y|x) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

Let,

$y\hat{}_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$ are estimates of $P\left(y_i = 1\middle|x_{ij}\right)$

Therefore,

$$\beta_j = \frac{\partial P(y_i = 1|x_{ij})}{\partial x_j}$$

where j = 1,2,3.....k times

Now,

let,

$$E(Y_i|X_i) = \beta_0 + \beta_1 x_{i1} = P(Y_i = 1|X_i)$$

The model assumes : $E(\varepsilon_i) = 0$

If $P_i = probability\ then\ Y_i = 1, that\ is\ event\ occurs\ and$

$(1 - P_i) = probability\ then\ Y_i = 0(that\ is\ event\ doesnot$

$occur)$.

Y_i	Probability
0	$(1 - P_i)$
1	P_i

Here

$\boldsymbol{Y_i, follows\ the\ Bernoulli\ or\ Binomial\ Probability\ Distribution.}$

$\boldsymbol{E(Y_i) = 0 \times (1 - P_i) + 1 \times (P_i) = (P_i)}$

Now,

$$\boldsymbol{E(Y_i|X_i) = \beta_0 + \beta_1 x_{i1} = P_i}$$

and $\boldsymbol{E(Y_i|X_i)\ can\ be\ further\ represented\ as\ Y\hat{}_i\ i.e.}$

$$\boldsymbol{Y\hat{}_i = E(Y_i|X_i) = \beta_0 + \beta_1 x_{i1} = P_i}$$

If there are $'\boldsymbol{n}'$independent trials, each with a probability $\boldsymbol{P}$ of success and probability $\boldsymbol{(1 - P)}$ of failure, and $\boldsymbol{X}$ of these trials represent the number of successes, then $\boldsymbol{X}$ is said to follow the **binomial distribution.**

Binomial Distribution:

$$\boldsymbol{P(x) = (n_{C_x})} \times P^x \times q^{n-x}$$

i.e.

$$P(x) = \left(\frac{n!}{(n - x!) \times x!}\right) \times P^x \times q^{n-x}$$

where,

$n = the\ number\ of\ trials(or\ the\ number\ being\ sampled)$

$x = number\ of\ success\ desired.$

$p = probability\ of\ getting\ success\ in\ trial.$

$q = (1 - P) = probability\ of\ getting\ a\ failure\ in\ one\ trial.$

The mean of the binomial distribution is $\boldsymbol{np}$ and its variance is $\boldsymbol{np(1 - P)}$.The term success is defined in the context of the problem.

Since the probability P_i must lie between 0 and 1 , we have the restriction where,

$$0 \le E(Y_i|X_i) \le 1$$

Note ,

$$E(\varepsilon_i) = 0 \; and \; cov(u_i, u_j) = 0 \; for \; i = j \; i.e. no \; serial \; correlation$$

and

$$var(\varepsilon_i) = P_i(1 - P)$$

To solve the uneven scattered problem of the model we have to take out the weight:

$$\sqrt{E(Y_i|X_i) \times (1 - E(Y_i|X_i))}$$
$$= \sqrt{P_i \times (1 - P_i))}$$

Say ,

$$\sqrt{P_i \times (1 - P_i))} = \sqrt{w_i}$$

Then,

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_0}{\sqrt{w_i}} + \beta_1 \times \frac{x_1}{\sqrt{w_i}} + \frac{\varepsilon_i}{\sqrt{w_i}}$$

where w_i *is serving as weights*.

Therefore how to get the model for Simple Linear Regression Probability Model:

Using Ordinary Least Square(OLS)

Step 1: **Find,** $Y_i = \beta_0 x_i^0 + \beta_1 x_i^1 + \epsilon_i$

Step 2: **Find,** $Y\hat{}_i = E(Y_i|X_i) = \beta_0 x_i^0 + \beta_1 x_i^1 = P_i$

Where, P_i *is true*.

Using Weighted Least Square(WLS)

Step 3: **Find,** $E(w_i) = w\hat{}_i = Y\hat{}_i(1 - Y\hat{}_i)$

Step 4: **Find,**

$$\frac{Y_i}{\sqrt{w\hat{}_i}} = \frac{\beta_0}{\sqrt{w\hat{}_i}} + \beta_1 \times \frac{x_{i1}}{\sqrt{w\hat{}_i}} + \frac{\varepsilon_i}{\sqrt{w\hat{}_i}}$$

To remove the uneven scattering problem i.e. heteroscedasticity problem.

Multiple Linear Regression Probability Model:

Using Ordinary Least Square(OLS)

Step 1: Find, $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$

Step 2: Find,$Y\hat{}_i = E(y_i|x_i) = \beta\hat{}_0 + \beta\hat{}_1 x_i^1 + \beta\hat{}_2 x_i^2 + .. + \beta\hat{}_p x_i^p \; = P_i$

Where, $P_i\ is\ true.$

Using Weighted Least Square(WLS)

Step 3: Find, $E(w_i) = w\hat{}_i = Y\hat{}_i(1 - Y\hat{}_i)$

Step 4: **Find,**

$$\frac{Y_i}{\sqrt{w\hat{}_i}} = \frac{\beta_0}{\sqrt{w\hat{}_i}} + \beta_1 \times \frac{x_{i1}}{\sqrt{w\hat{}_i}} + \cdots + \beta_p \times \frac{x_{ip}}{\sqrt{w\hat{}_i}} + \frac{\varepsilon_i}{\sqrt{w\hat{}_i}}$$

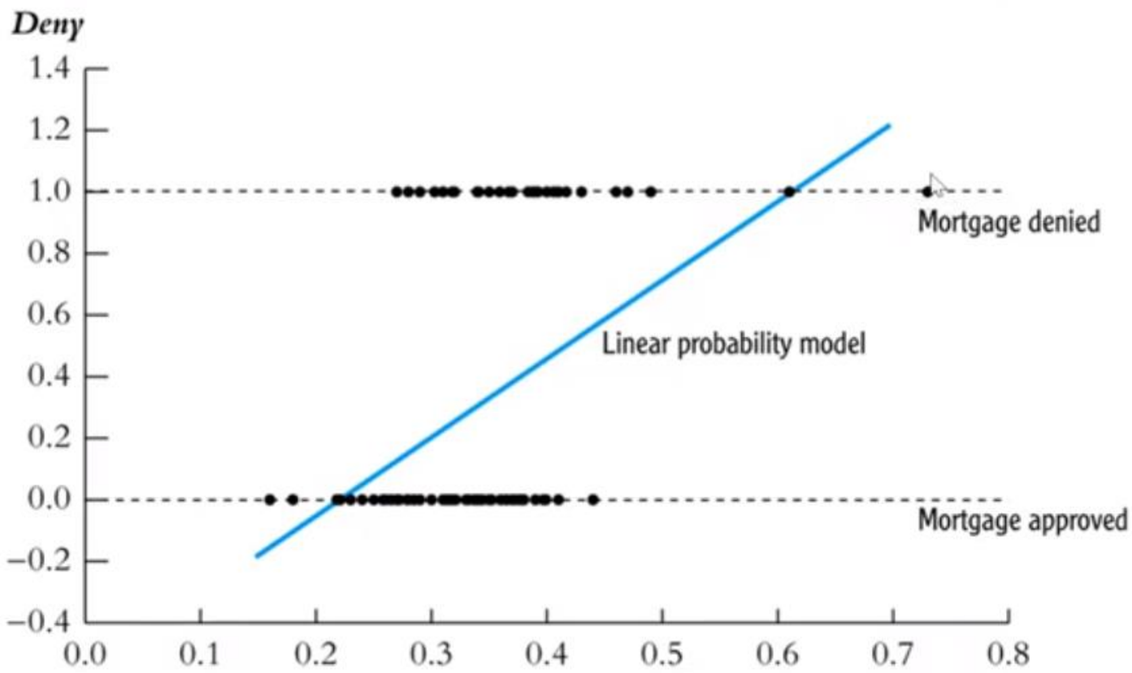
To remove the uneven scattering problem i.e. heteroscedasticity problem where the errors are large in the regression model.

Problems with LPM :

1. P_i **Are not restricted to 0 and 1.**

2. $E(Y_i|X_i)$ **Also are not restricted to 0 and 1.**

3. $\frac{\partial P(y_i=1|x_{ij})}{\partial x_j}$ **Are constant (they do not depend on any x-variable). This is typically unrealistic,** $\frac{\partial P(y_i=1|x_{ij})}{\partial x_j}$ **should eventually decrease with** x_j **as** x_j $\boldsymbol{become\ large}$ **. This causes** Heteroscedasticity problem where the errors are large in the regression model.



4. **These disadvantages can be solved by nonlinear probability model: probit and logit regression.**