So the:

$$Q = e'e = y'y - 2\beta'x'y + \beta'x'x\beta \tag{2.13}$$

To find the $\beta$ that makes $\varepsilon'\varepsilon$ as low as possible we take the partial derivative for each $\beta_i$ and then make it equal to zero:

$$\frac{\partial Q}{\partial \beta} = \begin{bmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \\ \vdots \\ \frac{\partial Q}{\partial \beta_p} \end{bmatrix} = -2x'y + 2x'x\beta = 0 \tag{2.14}$$

After simplification we find that:

$$X'X\beta = X'Y \tag{2.15}$$

Thus the estimations of the least squares are:

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \tag{2.16}$$

(Yuan, 2007).

**2.3. Logistic Regression Model (LR)**

The MLR model is applicable when the dependent variable is continuous, and not categorical, while LR is different from the MLR and used when the dependent variable is binary and the independent variables are quantity, categorical variables, or both (Marill, 2004).

Simple and multiple linear regression assumes linear relationship between the independent variables and the dependent variable, while logistic regression does not assume that. It is a useful tool for modeling and forecasting data that consists of binary dependent variable. Nowadays, the researchers state the fact that it is not appropriate to propose multiple linear regression for a binary dependent variable and the better choice institute of multiple linear regression is LR for accurate modeling and forecasting. Categorical events will be coded as binary variables with a value of one represents the positive outcome, or success as target outcome, and a value of zero represents the negative outcome, or failure (Hosmer *et al.*, 1997; Midi *et al.*, 2010; Pohlman & Leitner, 2003).

Figure2.1 explains the simple linear regression equation when the dependent variable is continuous, while Figure2.2 explains the simple linear regression equation when the dependent variable is binary.

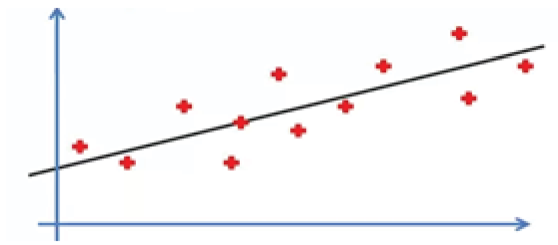

Figure 2.1 Simple linear regression equation with continuous response variable.

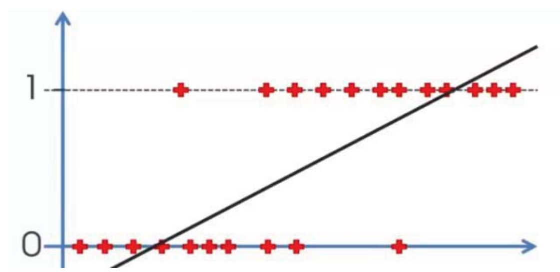


Figure 2.2 Simple linear regression equation with binary response variable.

From Figure2.2, it is difficult to represent the categorical or binary data using the line of simple regression. Therefore, the line of regression should be changed to another shape to be more fit with the binary data as shown in Figure2.3.

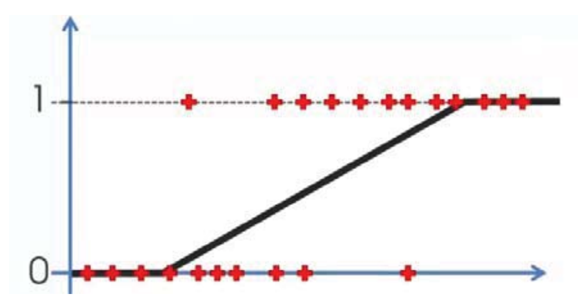


Figure 2.3 Supposed Simple regression line for binary response variable.

The regression line in figure 2.3 represents the data more appropriately, but this line was not used scientifically. Therefore, the most scientific style of regression line that can represent most of data points is like that in figure2.4,which is called logistic regression line. Therefore, the logistic regression can be proposed with categorical dependent variable for minimum random error.

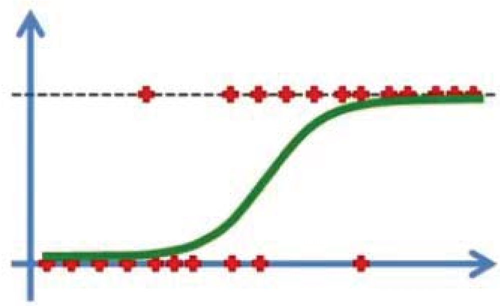

Figure 2.4 Logistic regression line for binary response.

Therefore, the logistic regression can be proposed with categorical dependent variable for minimum random error.

LR is based on probabilities associated with the values of $y$ . For simplicity, and since it is case most commonly encountered in practice, we assume that $y$ is dichotomous, taking on values of 1 and 0. In theory, the hypothetical, population proportion of cases for which $y = 1$ is defined as $p = p(y = 1)$ . Then, the theoretical proportion of cases for which $y = 0$ is $1 - p = p(y = 0)$ . In the absence of other information, we would estimate p by the sample proportion of cases for which $y = 1$ . However, in the regression context, it is assumed that there is a set of predictor variables, $x_1, \ldots, x_p$ , that are related to $y$ and, therefore, provide additional information for predicting $y$. For theoretical, mathematical reasons, LR is based on a linear model for the natural logarithm of the odds (i.e., the log-odds) in favor of $y = 1$ : (Dayton, 1992)

$$Ln\left[\frac{P\left(y=1\mid x_1,x_2,\ldots,x_p\right)}{1-P\left(y=1\mid x_1,x_2,\ldots,x_p\right)}\right]=Ln\left[\frac{\pi}{1-\pi}\right] \quad (2.17)$$

$$=\beta_0+\beta_1 x_1+\ldots+\beta_p x_p=\beta_0+\sum_{j=1}^{p}\beta_j x_j \quad (2.18)$$

Where $\beta_0$ is the constant parameter, $\beta_j$ are the regression coefficients , $x_j$ are independent variables and $\pi$ represents the conditional probability of success which takes the form $P\left(y=1\mid x_1,\ldots,x_p\right)$ and depends on combinations of independent variables. $Ln\left[\frac{\pi}{1-\pi}\right]$ term is the log-odds which is known as the logit transformation of p or the natural logarithmic of odd. There are two advantages of using the logistic model as development of the model. First is using multiplicative, rather than additive concepts for the odds and probabilities. Since the natural logarithm convert the multiplication into addition, taking the natural logarithm of the probabilities and odds allows for more simplicity using additive concepts. Relatively, the second one is using a simple exponential transformation that transfers the log-odds to probabilities such as in the following form.

$$\frac{P\left(y=1\mid x_1,x_2,\ldots,x_p\right)}{1-P\left(y=1\mid x_1,x_2,\ldots,x_p\right)}=\frac{\pi}{1-\pi}=e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}=e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j} \quad (2.19)$$

where the probability of success can be as follows.

$$\pi=P\left(y=1\mid x_1,x_2,\ldots,x_p\right)=\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}=\frac{e^z}{1+e^z}=\frac{1}{1+e^{-z}} \quad (2.20)$$

and the probability of failure can be as follows.

$$1-P\left(y=1\mid x_1,x_2,\ldots,x_p\right)=1-\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}=1-\frac{e^z}{1+e^z}=\frac{1}{1+e^z} \quad (2.21)$$

The logit model will classify the corresponding observation as 1 if the value of the probability $p\left(y_i = 1 | x_i\right) = \pi$ is greater than 0.5,while if this probability value is less than 0.5 then the corresponding observation will be classified as 0 (King & Zeng,2001).

### 2.3.1. Advantages and Disadvantages of Logistics Regression

**First: Advantages of Logistic Regression**

1-It deals with qualitative or quantitative variables and contains limits of interactions and a test of the significance of coefficients.

2-It gives the researcher an idea of the effect of each independent variable on the binary response variable.

3-Logistic regression computes the effect of variables, leading the researcher to conclude that a variable is stronger than the other variable.

4-The trend of deviations from the distribution modality of the study variables is less sensitive.

5-Parameter Estimation according to the logistic model is acceptable in the absence of some constraints on linear and logarithmic regression.

**Second: Disadvantages of Logistics Regression**

This method faces the problem of being unable to deal with data that depends on the change in the time point prior to the occurrence of the event, and contains data on the disappearance of Censored Data, Cross Sectional CCs, so there was the most manipulated alternative to this case, a Cox Regression (Leffondré *et al.*, 2003).

### 2.3.2. Maximum Likelihood Estimation

In order to estimate logistic regression coefficients, a higher probability method is used. Maximum Likelihood Method is one of the most appropriate methods for all linear and nonlinear models, and the most likely method is a repetitive method based on repetition calculations multiple times, until the best estimation of the transactions is achieved view data can be interpreted.

Also, since logistic regression predicts probabilities, rather than just classes, we can fit it using likelihood. For each training data-point, we have a vector of features, $x_i$ ,and an observed class, $y_i$ , the probability of that class was either $\pi(x)$ , if $y_i = 1$ or $1-\pi(x)$ , if $y_i = 0$ .(Santner & Duffy, 1986). The likelihood is then;

$$L(\beta_0, \beta) = \prod_{i=1}^{p} \pi(\mathrm{x}_i)^{y_i} (1-\pi(\mathrm{x}_i))^{1-y_i} \tag{2.22}$$

The log-likelihood turns products into sums:

$$L(\beta_0, \beta\ ) = \sum_{i=1}^{p} (\mathrm{y}_i \log \pi(\mathrm{x}_i) + (1-\mathrm{y}_i) \log(1-\pi(\mathrm{x}_i)) \tag{2.23}$$

$$= \sum_{i=1}^{p} (\log(1-\pi(\mathrm{x}_i)) + \mathrm{y}_i \log \pi(\mathrm{x}_i) - y_i \log(1-\pi(\mathrm{x}_i)))$$

$$= \sum_{i=1}^{p} (\log(1-\pi(\mathrm{x}_i)) + \mathrm{y}_i \log \frac{\pi(\mathrm{x}_i)}{(1-\pi(\mathrm{x}_i))}) \tag{2.24}$$

$$= \sum_{i=1}^{p} \log(1-\pi(\mathrm{x}_i) + \sum_{i=1}^{p} y_i (\beta_0 + x_i . \beta_i) \tag{2.25}$$

$$= \sum_{i=1}^{p} -\log 1 + (1 + e^{(\beta_0 + x_i . \beta_i)}) + \sum_{i=1}^{p} y_i (\beta_0 + x_i . \beta_i) \tag{2.26}$$

Where in the next - to-last step, we finally use equation (2.26)

Typically, to find the maximum likelihood estimates wed differentiate the log likelihood with respect to the parameters, set the derivatives equal to zero, and solve. To start that, take the derivative with respect to one component of $\beta$,say $\beta_j$ .

With reference to Equation ( 2.2) , we get the following:-

$$\frac{\partial L}{\partial \beta_j} = -\sum_{i=1}^{p} \frac{1}{1 + e^{(\beta_0 + x_i . \beta)}} e^{(\beta_0 + x_i . \beta)} x_i + \sum_{i=1}^{p} y_i x_i \tag{2.27}$$

$$= \sum_{i=1}^{p} (y_i - \pi(x_i; \beta_0, \beta)) x_i \tag{2.28}$$

The procedure is equivalent to iteratively re-weighted least squares (IRLS) (Nelder & Wedderburn, 1972).

We are not going to be able to set this to zero and solve exactly. (That is a transcendental equation, and there is no closed-form solution.) We can however approximately solve it numerically.