Contents

1 Meaning of Response Function when Outcome Variable is Binary or Dichotomous

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad Y_i = 0, 1$$

where the outcome Y_i is binary, taking on the value of either 0 or 1 . The expected response $E\{Y_i\}$ at each level of X_i has a special meaning in this case. Since $E\{\varepsilon_i\} = 0$ we have:

$$E\{Y_i\} = \beta_0 + \beta_1 X_i \tag{1}$$

Consider Y_i to be a Bernoulli random variable for which we can state the probability distribution as follows:

Y_i	Probability
1	$P(Y_i = 1) = \pi_i$
0	$P(Y_i = 0) = 1 - \pi_i$

Thus, π_i is the probability that $Y_i = 1$, and $1 - \pi_i$ is the probability that $Y_i = 0$. By the definition of expected value of a random variable, we obtain:

$$\begin{aligned} E\{Y_i\} &= \sum_{i=1}^{2} Y_i P(Y = Y_i) & (2) \\ &= 1P(Y_i = 1) + 0P(Y_i = 0) & (3) \\ &= 1(\pi_i) + 0(1 - \pi_i) & (4) \\ &= \pi_i & (5) \\ &= P(Y_i = 1) & (6) \end{aligned}$$

Equating equation 1 and equation 5, we thus find:

$$E\{Y_i\} = \beta_0 + \beta_1 X_i = \pi_i = P(Y_i = 1) \tag{7}$$

The mean response $E\{Y_i\} = \beta_0 + \beta_1 X_i$ as given by the response function is $\therefore$ **simply the probability that $Y_i = 1$ when the level of the predictor variable is X_i.** This interpretation of the mean response applies whether the response function is a simple linear one, as here, or a complex multiple regression one. **The mean response, when the outcome variable is a 0,1 indicator variable, always represents the probability that $Y = 1$ for the given levels of the predictor variables.**

2 Special Problems When Response Variable is Binary

2.1 Non-normal Error Terms

For a binary 0,1 response variable, each error term $\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$ can take on only two values:

$$\text{When } Y_i = 1: \quad \varepsilon_i = 1 - \beta_0 - \beta_1 X_i$$
$$\text{When } Y_i = 0: \quad \varepsilon_i = -\beta_0 - \beta_1 X_i$$

The classical normal error regression model, which assumes that the ε_i are normally distributed, is not appropriate.

2.2 Non-constant Error Variance

Another problem with the error terms is that they do not have equal variances when the response variable is an indicator variable. To see this, we shall obtain $\sigma^2\{Y_i\}$ for the simple linear regression model, utilizing (A.15):

$$\begin{aligned}
\sigma^2\{Y\} &= E\{(Y - E\{Y\})^2\} \\
&= E\{(Y - \mu_Y)^2\} \\
&= E\{(Y^2 - 2\mu_Y Y + \mu_Y^2)\} \\
&= E\{(Y^2\} - E\{2\mu_Y Y\} + E\{\mu_Y^2)\} \\
&= E\{(Y^2\} - 2\mu_Y E\{Y\} + \mu_Y^2) \\
&= E\{(Y^2\} - 2\mu_Y \mu_Y + \mu_Y^2) \\
&= E\{(Y^2\} - 2\mu_Y^2 + \mu_Y^2) \\
&= E\{(Y^2\} - \mu_Y^2) \\
&= E\{Y^2\} - (E\{Y\})^2
\end{aligned}$$

Therefore:

$$\sigma^2\{Y_i\} = E\{(Y_i - E\{Y_i\})^2\}$$

Since $Y_i = 1$ or 0 and $E\{Y_i\} = \pi_i$:

$$\begin{aligned}
\sigma^2\{Y_i\} &= E\{(Y_i - E\{Y_i\})^2\} \\
&= (Y_i - E\{Y_i\})^2 P(Y_i = 1) + (Y_i - E\{Y_i\})^2 P(Y_i = 0) \\
&= (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i) \\
&= (1 - 2\pi_i + \pi^2)\pi_i + \pi_i^2 (1 - \pi_i)
\end{aligned}$$

$$
\begin{aligned}
&= \pi_i - 2\pi_i^2 + \pi_i^3 + \pi_i^2 - \pi_i^3 \\
&= \pi_i \cancel{-2\pi_i^2} \cancel{+\pi_i^3} \cancel{+\pi_i^2} \cancel{-\pi_i^3} \\
&= \pi_i - \pi_i^2 \\
&= \pi_i(1 - \pi_i)
\end{aligned}
$$

We know that $E\{Y_i\} = \pi_i$:

$$\sigma^2\{Y_i\} = \pi_i(1 - \pi_i) = (E\{Y_i\})(1 - E\{Y_i\})$$

The variance of ε_i is the same as that of Y_i because $\varepsilon_i = Y_i - \pi_i$ and π_i is just a constant:

$$\sigma^2 \{\varepsilon_i\} = \pi_i (1 - \pi_i) = (E \{Y_i\}) (1 - E \{Y_i\})$$

Or equivalently, since $E\{Y_i\} = \beta_0 + \beta_1 X_i$:

$$
\begin{aligned}
\sigma^2 \{\varepsilon_i\} &= (\beta_0 + \beta_1 X_i) (1 - (\beta_0 + \beta_1 X_i)) \\
&= (\beta_0 + \beta_1 X_i) (1 - \beta_0 - \beta_1 X_i)
\end{aligned}
$$

From the equation above, it can be seen that $\sigma^2 \{\varepsilon_i\}$ depends on X_i. Hence, the error variances will differ at different levels of X, and ordinary least squares will no longer be optimal.

2.3 Constraints on Response Function

Since the response function represents probabilities when the outcome variable is a 0,1 indicator variable, the mean responses should be constrained as follows:

$$0 \leq E\{Y\} = \pi \leq 1$$

Many response functions do not automatically possess this constraint. A linear response function, for instance, may fall outside the constraint limits within the range of the predictor variable in the scope of the model.

3 Sigmoidal Response Functions For Binary Response

3.1 Probit Mean Response Function

Consider a health researcher studying the effect of a mother's use of alcohol ($X-$ an index of degree of alcohol use during pregnancy) on the duration of her pregnancy (Y^c). Here we use the superscript c to emphasize that the response variable, pregnancy duration, is a continuous response. This can be represented by a simple linear regression model:

$$Y_i^c = \beta_0^c + \beta_1^c X_i + \varepsilon_i^c \tag{8}$$

and we will assume that ε_i^c is normally distributed with mean zero and variance σ_c^2. If the continuous response variable, pregnancy duration, were available, we might proceed with the usual simple linear regression analysis. However, in this instnce, researchers coded each pregnancy duration as preterm or full term using the following rule:

$$Y_i = \begin{cases} 1 & \text{if } Y_i^c \leq 38 \text{ weeks (preterm)} \\ 0 & \text{if } Y_i^c > 38 \text{ weeks (full term)} \end{cases}$$

It follows from equations 7 and 8 that:

$$\begin{aligned} P\left(Y_i = 1\right) = \pi_i &= P\left(Y_i^c \leq 38\right) \\ &= P\left(\beta_0^c + \beta_1^c X_i + \varepsilon_i^c \leq 38\right) \\ &= P\left(\varepsilon_i^c \leq 38 - \beta_0^c - \beta_1^c X_i\right) \\ &= P\left(\frac{\varepsilon_i^c}{\sigma_c} \leq \frac{38 - \beta_0^c}{\sigma_c} - \frac{\beta_1^c}{\sigma_c} X_i\right) \\ &= P\left(Z \leq \beta_0^* + \beta_1^* X_i\right) \end{aligned}$$

Where

- $\beta_0^* = \frac{38-\beta_0^c}{\sigma_c}$
- $\beta_1^* = \frac{\beta_1^c}{\sigma_c}$
- $Z = \frac{\varepsilon_i^c}{\sigma_c}$

And $Z = \frac{\varepsilon_i^c}{\sigma_c}$ follows a standard normal distribution. So if we let $P(Z \leq z) = \phi(z)$, then:

$$P\left(Y_i = 1\right) = \Phi\left(\beta_0^* + \beta_1^* X_i\right) \tag{9}$$

Equations 7 and 9 together yield the nonlinear regression function known as the **probit mean response function**:

$$E\left\{Y_i\right\} = \pi_i = \Phi\left(\beta_0^* + \beta_1^* X_i\right) \tag{10}$$

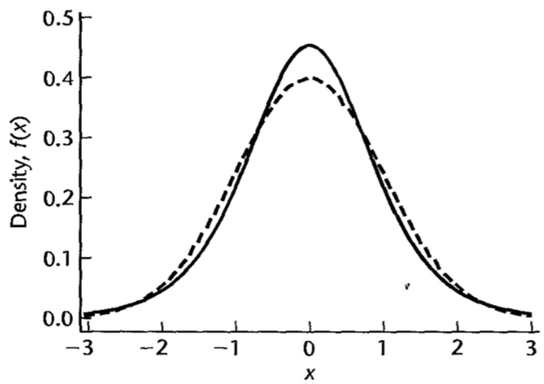
The inverse function, Φ^{-1} of the standard normal cumulative distribution function Φ, is sometimes called the probit transformation. We solve for the linear predictor, $\beta_0^* + \beta_1^* X_i$ in equation 10 by applying the probit transformation to both sides of the expression, obtaining:

$$\Phi^{-1}\left(\pi_i\right) = \pi_i' = \beta_0^* + \beta_1^* X_i \tag{11}$$

The resulting expression, $\pi_i' = \beta_0^* + \beta_1^* X_i$, is called the **probit response function**, or more generally, the **linear predictor**.

3.2 Logistic Mean Response Function

We have seen that the assumption of normally distributed errors for the underlying continuous response variable in 8 led to the use of the standard normal cumulative distribution function, Φ, to model the probability of mean response π_i. An alternative error distribution that is very similar to the normal distribution is the **logistic distribution**. The figure below presents plots of the standard normal density function (dashed line) and the logistic density function (solid line), each with mean zero and variance one. The plots are nearly indistinguishable, although the logistic distribution has slightly heavier tails.



The logistic distribution has the following statistics:

Table 1

Notation	logistic(μ,s)
Parameter	$0 \leq \mu \leq \infty$ & $s > 0$
Distribution	$0 \leq x \leq \infty$
Pdf	$\frac{\exp\left(-\frac{x-\mu}{s}\right)}{\mathrm{s}\left(1+\exp\left(-\frac{x-\mu}{s}\right)\right)^2}$
Cdf	$\frac{1}{1+\exp\left(-\frac{x-\mu}{s}\right)}$
Mean	μ
Variance	$\frac{1}{3}s^2\pi^2$
Skewness	0
Kurtosis	$\frac{6}{5}$

Using theses statistics, we find that the density of a logistic random variable ε_L having mean zero ($\mu = 0$) and standard deviation $\sigma = \pi/\sqrt{3}$ (which implies that $s = 1$) has a simple form:

$$f_L\left(\varepsilon_L\right) = \frac{\exp\left(\varepsilon_L\right)}{\left[1+\exp\left(\varepsilon_L\right)\right]^2}$$

Its cumulative distribution function is:

$$F_L(\varepsilon_L) = \frac{\exp(\varepsilon_L)}{1 + \exp(\varepsilon_L)}$$

Suppose now that ε_i^c in equation 8 has a logistic distribution with mean zero and standard deviation σ_c. Then, we have:

$$\begin{aligned}
P(Y_i = 1) = \pi_i &= P(Y_i^c \leq 38) \\
&= P(\beta_0^c + \beta_1^c X_i + \varepsilon_i^c \leq 38) \\
&= P(\varepsilon_i^c \leq 38 - \beta_0^c - \beta_1^c X_i) \\
&= P\left(\frac{\varepsilon_i^c}{\sigma_c} \leq \frac{38 - \beta_0^c}{\sigma_c} - \frac{\beta_1^c}{\sigma_c} X_i\right) \\
&= P\left(\frac{\varepsilon_i^c}{\sigma_c} \leq \beta_0^* + \beta_1^* X_i\right)
\end{aligned}$$

where $\frac{\varepsilon_i^c}{\sigma_c}$ follows a logistic distribution with mean zero and standard deviation one instead of a standard normal distribution.

$$\frac{\varepsilon_i^c}{\sigma_c} \sim \text{Logistic}(0,\, 1) \text{ rather than } \frac{\varepsilon_i^c}{\sigma_c} = Z \sim \text{N}(0,\, 1)$$

Multiplying both sides of the inequality inside the probability statement $P()$ by $\frac{\pi}{\sqrt{3}}$ does not change the probability; therefore:

$$\begin{aligned}
P(Y_i = 1) = \pi_i &= P\left(\frac{\pi}{\sqrt{3}} \frac{\varepsilon_i^c}{\sigma_c} \leq \frac{\pi}{\sqrt{3}} \beta_0^* + \frac{\pi}{\sqrt{3}} \beta_1^* X_i\right) \\
&= P(\varepsilon_L \leq \beta_0 + \beta_1 X_i) \\
&= \text{Using the CDF } F_L \text{ of logistic distribution instead of } \Phi \\
&= F_L(\beta_0 + \beta_1 X_i) \\
&= \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}
\end{aligned}$$

where:

- $\beta_0 = (\pi/\sqrt{3})\beta_0^*$
- $\beta_1 = (\pi/\sqrt{3})\beta_1^*$

denote the logistic regression parameters. **Notice that the logistic regression parameters are essentially scaled probit response function parameters.** To summarize, the **logistic mean response function** is:

$$E\{Y_i\} = \pi_i = F_L(\beta_0 + \beta_1 X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \tag{12}$$

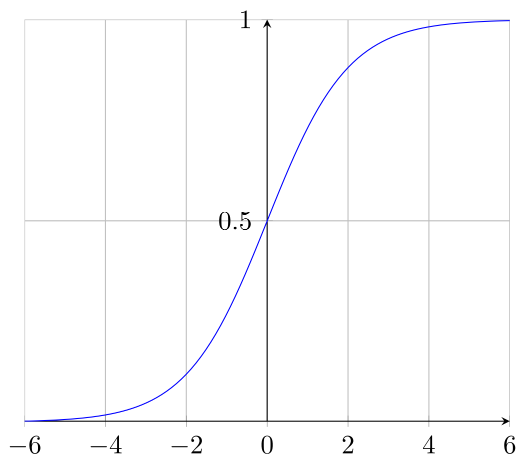
An equivalent form of equation 12 can be obtained as follows. For simplicity, we substitute $x = \beta_0 + \beta_1 X_i$:

$$
\begin{aligned}
E\{Y_i\} = \pi_i &= \frac{\exp(x)}{1+\exp(x)} \\
&= \frac{\exp(x)}{1+\exp(x)} \cdot \frac{\exp(x)}{\exp(x)} \\
&= \frac{\exp(x)}{1+\exp(x)} \cdot \frac{1}{\exp(-x)} \cdot \frac{1}{\exp(x)} \qquad \text{Because} \Longrightarrow \frac{\exp(x)}{1} = \frac{1}{\exp(-x)} \\
&= \frac{\exp(x)}{(1+\exp(x))\exp(-x)} \cdot \frac{1}{\exp(x)} \\
&= \frac{\cancel{\exp(x)}}{\exp(-x)+1} \cdot \frac{1}{\cancel{\exp(x)}} \qquad \text{Because } e^x e^{-x} = e^{x+(-x)} = e^0 = 1 \\
&= \frac{1}{\exp(-x)+1}
\end{aligned}
$$

Substituting $E\{Y_i\}$, we get:

$$E\{Y_i\} = \pi_i = \Bigg[\exp(-(\beta_0 + \beta_1 X_i)) + 1\Bigg]^{-1} \tag{13}$$

From the equation above, we can see why the logistic mean response function ranges between 0 and 1.



- As $x \to \infty$, then e^{-x} or $\exp(-x)$ will approach 0, and so the output of the logistic function will approach $\frac{1}{1} = 1$.
- As $x \to -\infty$, e^{-x} or $\exp(-x)$ will become very large, and so the output of the logistic function will approach $\frac{1}{\text{large number}} \approx 0$.

Applying the inverse of the cumulative distribution function F_L to equation 12 yields the **logit response function**:

$$F_L^{-1}(\pi_i) = \pi_i' = \beta_0 + \beta_1 X_i \tag{14}$$

The transformation $F_L^{-1}(\pi_i)$ is called the logit transformation of the probability π_i. The name comes from the fact that the logit function, defined as $\log_e\left(\frac{\pi_i}{1-\pi_i}\right)$, is the inverse of the logistic function:

$$F_L^{-1}(\pi_i) = \log_e\left(\frac{\pi_i}{1-\pi_i}\right) = \pi_i' = \beta_0 + \beta_1 X_i \tag{15}$$

where the ratio $\frac{\pi_i}{1-\pi_i}$ is called the odds, which is the log of the ratio between the estimated probability for the positive class and the estimated probability for the negative class. The linear predictor $\beta_0 + \beta_1 X_i$ is referred to as the **logit response function**. Notice again that the logit parameters are scaled probit parameters.

3.3 Properties of Probit and Logit Mean Response Functions

- Both are bounded between 0 and 1, and both approach these limits asymptotically.
- For $\beta_1 > 0$ and $\beta_1^* > 0$ the mean response functions are monotone increasing; they are monotone decreasing when $\beta_1 < 0$ and $\beta_1^* < 0$.
- Increasing or decreasing the intercepts β_0 or β_0^* shifts the mean response functions horizontally. The direction depends on the signs of both the intercept and the slope.
- When $\beta_1 > 0$ and $\beta_1^* > 0$, increasing β_1 and β_1^* makes the mean response functions more S-shaped, changing more rapidly in the center.
- The probit and logit response functions are symmetric. In other words, if the coefficients are recoded by changing 1s to 0s and vice versa— the signs of all the coefficients are reversed.

4 Simple Logistic Regression

First, we require a formal statement of the simple logistic regression model. Recall that when the response variable is binary, taking on the values 1 and 0 with probabilities π and $1-\pi$, respectively, Y is a Bernoulli random variable with parameter $E\{Y\} = \pi$. We could state the simple logistic regression model in the usual form:

$$Y_i = E\left\{Y_i\right\} + \varepsilon_i$$

Since the distribution of the error term ε_i depends on the Bernoulli distribution of the response Y_i, it is preferable to state the simple logistic regression model in the following fashion:

> Y_i are independent Bernoulli random variables (instead of independent normal random variables) with expected values $E\left\{Y_i\right\} = \pi_i$, where:

$$E\left\{Y_i\right\} = \pi_i = \frac{\exp\left(\beta_0 + \beta_1 X_i\right)}{1 + \exp\left(\beta_0 + \beta_1 X_i\right)} = \left[\exp(-\beta_0 - \beta_1 X_i) + 1\right]^{-1} \tag{16}$$

The X observations are assumed to be known constants. Alternatively, if the X observations are random, $E\left\{Y_i\right\}$ is viewed as a conditional mean, given the value of X_i.

4.0.1 Likelihood Function

Since each Y_i observation is an ordinary Bernoulli random variable, where:

$$P\left(Y_i=1\right)=\pi_i$$
$$P\left(Y_i=0\right)=1-\pi_i$$

we can represent its probability distribution as follows:

$$f_i\left(Y_i\right)=\pi_i^{Y_i}\left(1-\pi_i\right)^{1-Y_i} \quad Y_i=0,1; \quad i=1,\ldots,n$$

Note that $f_i(1)=\pi_i$ and $f_i(0)=1-\pi_i$. Hence, $f_i\left(Y_i\right)$ simply represents the probability that $Y_i=1$ or 0 Since the Y_i observations are independent, their joint probability function is:

$$g\left(Y_1,\ldots,Y_n\right)=\prod_{i=1}^{n} f_i\left(Y_i\right)=\prod_{i=1}^{n} \pi_i^{Y_i}\left(1-\pi_i\right)^{1-Y_i}$$

Again, it will be easier to find the maximum likelihood estimates by working with the logarithm of the joint probability function:

$$\begin{aligned}
\log_e g\left(Y_1,\ldots,Y_n\right) &= \log_e \prod_{i=1}^{n} \pi_i^{Y_i}\left(1-\pi_i\right)^{1-Y_i} \\
&= \text{Apply log rule } \ln(xy)=\ln(x)+\ln(y) \quad \& \quad \ln(x^k)=k\ln(x) \\
&= \sum_{i=1}^{n}\left[Y_i \log_e \pi_i+\left(1-Y_i\right)\log_e\left(1-\pi_i\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \log_e \pi_i+\left[\log_e\left(1-\pi_i\right)-Y_i\log_e(1-\pi_i)\right]\right] \\
&= \sum_{i=1}^{n}\left[Y_i \log_e \pi_i+\log_e\left(1-\pi_i\right)-Y_i\log_e(1-\pi_i)\right] \\
&= \sum_{i=1}^{n}\left[Y_i[\log_e \pi_i-\log_e(1-\pi_i)]+\log_e\left(1-\pi_i\right)\right] \\
&= \text{Apply log rule } \ln(x)-\ln(y)=\ln(\frac{x}{y}) \\
&= \sum_{i=1}^{n}\left[Y_i \log_e\left(\frac{\pi_i}{1-\pi_i}\right)+\log_e\left(1-\pi_i\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \log_e\left(\frac{\pi_i}{1-\pi_i}\right)\right]+\sum_{i=1}^{n}\log_e\left(1-\pi_i\right)
\end{aligned}$$

Since $E\{Y_i\} = \pi_i$ for a binary variable, we can solve for $1-\pi_i$ using equation 12. If we again let $x = \beta_0+\beta_1 X_i$ then it follows that:

$$
\begin{aligned}
1-\pi_i &= 1-\frac{\exp(x)}{1+\exp(x)} \\
&= \frac{1+\exp(x)}{1+\exp(x)} - \frac{\exp(x)}{1+\exp(x)} \\
&= \frac{(1+\exp(x))-\exp(x)}{1+\exp(x)} \\
&= \frac{1}{1+\exp(x)} \\
&= [1+\exp(x)]^{-1} \\
&= \text{Substituting for } x = \beta_0+\beta_1 X_i \\
&= [1+\exp(\beta_0+\beta_1 X_i)]^{-1}
\end{aligned}
$$

Furthermore, from equation 15, we obtain:

$$
\log_e\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0+\beta_1 X_i
$$

Hence, the log likelihood function can be expressed as follows:

$$
\begin{aligned}
\log_e L(\beta_0,\beta_1) &= \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) + \sum_{i=1}^{n} \log_e\left(\frac{1}{1+\exp(\beta_0+\beta_1 X_i)}\right) \\
&= \text{Apply log rule } \ln(x)-\ln(y) = \ln(\frac{x}{y}) \\
&= \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) + \sum_{i=1}^{n} \log_e(1) - \log_e[1+\exp(\beta_0+\beta_1 X_i)] \\
&= \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) + \sum_{i=1}^{n} 0 - \log_e[1+\exp(\beta_0+\beta_1 X_i)] \\
&= \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) + \sum_{i=1}^{n} -\log_e[1+\exp(\beta_0+\beta_1 X_i)] \\
&= \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) - \sum_{i=1}^{n} \log_e[1+\exp(\beta_0+\beta_1 X_i)]
\end{aligned}
$$

where $L(\beta_0,\beta_1)$ replaces $g(Y_1,\ldots,Y_n)$ to show explicitly that we now view this function as the likelihood function of the parameters to be estimated, given the sample observations.

$$
\log_e L(\beta_0,\beta_1) = \sum_{i=1}^{n} Y_i(\beta_0+\beta_1 X_i) - \sum_{i=1}^{n} \log_e[1+\exp(\beta_0+\beta_1 X_i)] \tag{17}
$$

4.0.2 Maximum Likelihood Estimation

The maximum likelihood estimates of β_0 and β_1 in the simple logistic regression model are those values of β_0 and β_1 that maximize the log-likelihood function in equation 17. No closed-form solution exists for the values of β_0 and β_1 in equation 17 that maximize the log-likelihood function. So we would need to use numerical search procedures to find the maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. Once the maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are found, we substitute these values into the response function in equation 16 to obtain the fitted response function. We shall use $\hat{\pi}_i$ to denote the fitted value for the i th case:

$$\hat{\pi}_i = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)}{1 + \exp\left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)}$$

The fitted value, $\hat{\pi}_i$, for the i th case is the estimated probability that $Y_i = 1$ when the level of the predictor variable is X_i. The **fitted simple logistic response function** is as follows:

$$\hat{\pi} = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 X\right)}{1 + \exp\left(\hat{\beta}_0 + \hat{\beta}_1 X\right)} \tag{18}$$

If we utilize the logit transformation in equation 15, we can express the fitted response function in 18 in another way as follows:

$$\hat{\pi}' = \hat{\beta}_0 + \hat{\beta}_1 X \tag{19}$$

where:

$$\hat{\pi}' = \log_e\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right)$$

Equation 19 is another way to express the **fitted logit response function**.

4.1 Interpretation of coefficient

The interpretation of the estimated regression coefficient $\hat{\beta}_1$ in the fitted logistic response function is not the straightforward interpretation of the slope in a linear regression model. The reason is that the effect of a unit increase in X varies for the logistic regression model according to the location of the starting point on the X scale. An interpretation of $\hat{\beta}_1$ is found in the property of the fitted logistic function that the estimated odds $\hat{\pi}/(1-\hat{\pi})$ are multiplied by $\exp\left(\hat{\beta}_1\right)$ for any unit increase in X. To see this, we consider the value of the fitted logit response function 19 at $X = X_j$:

$$\hat{\pi}'\left(X_j\right) = \hat{\beta}_0 + \hat{\beta}_1 X_j$$

The notation $\hat{\pi}'\left(X_j\right)$ indicates specifically the X level associated with the fitted value. We also consider the value of the fitted logit response function at $X = X_j + 1$(One unit increase):

$$\hat{\pi}'\left(X_j + 1\right) = \hat{\beta}_0 + \hat{\beta}_1\left(X_j + 1\right)$$

The difference between the two fitted values is simply:

$$
\begin{aligned}
\hat{\pi}'(X_j+1) - \hat{\pi}'(X_j) &= [\hat{\beta}_0 + \hat{\beta}_1 (X_j+1)] - [\hat{\beta}_0 + \hat{\beta}_1 X_j] \\
&= \cancel{\hat{\beta}_0} + \hat{\beta}_1 (X_j+1) \cancel{-\hat{\beta}_0} - \hat{\beta}_1 X_j \\
&= \hat{\beta}_1 (X_j+1) - \hat{\beta}_1 X_j \\
&= \hat{\beta}_1 [(\cancel{X_j} + 1) \cancel{-X_j}] \\
&= \hat{\beta}_1 (1) = \hat{\beta}_1
\end{aligned}
$$

Now according to equation 19, $\hat{\pi}'(X_j)$ is the logarithm of the estimated odds when $X = X_j$:

$$\hat{\pi}'(X_j) = \log_e \left(\frac{\hat{\pi}(X_j)}{1 - \hat{\pi}(X_j)} \right)$$

we shall denote this by $\log_e(odds_1)$. Similarly, $\hat{\pi}'(X_j+1)$ is the logarithm of the estimated odds when $X = X_j + 1$:

$$\hat{\pi}'(X_j+1) = \log_e \left(\frac{\hat{\pi}(X_j+1)}{1 - \hat{\pi}(X_j+1)} \right)$$

so we shall denote it by $\log_e(odds_2)$. Hence, the difference between the two fitted logit response values can be expressed as follows:

$$\log_e(\text{odds}_2) - \log_e(\text{odds}_1) = \log_e \left(\frac{\text{odds}_2}{\text{odds}_1} \right) = \hat{\beta}_1$$

Taking antilogs of each side, we see that the estimated ratio of the odds, called the **odds ratio** and denoted by $\widehat{OR}$, equals $\exp\left(\hat{\beta}_1\right)$:

$$\widehat{OR} = \frac{\text{odds}_2}{\text{odds}_1} = \exp\left(\hat{\beta}_1\right)$$

The odds ratio is interpreted as saying that the odds of $Y = 1$ increases by $(\exp\left(\hat{\beta}_1\right) - 1) \cdot 100$ percent per unit increase in X. In general, the estimated odds ratio when there is a difference of c units of X is:

$$\exp(c\hat{\beta}_1)$$

In words, the coefficient is interpreted as saying that the odds of $Y = 1$ increases over $\exp(c\hat{\beta}_1)$ fold.

5 Multiple Logistic Regression

The simple logistic regression model 16 is easily extended to more than one predictor variable. In fact, several predictor variables are usually required with logistic regression to obtain adequate description and useful predictions. In extending the simple logistic regression model, we simply replace $\beta_0 + \beta_1 X$ in equation 12 by $\beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$. To simplify the formulas, we shall use matrix notation and the following three vectors:

$$\underset{p\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad \underset{p\times 1}{\mathbf{X}} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_{p-1} \end{bmatrix} \quad \underset{p\times 1}{\mathbf{X}_i} = \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{i,p-1} \end{bmatrix}$$

We then have:

$$\mathbf{X}'\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$$
$$\mathbf{X}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1}$$

With this notation, the simple logistic response function 16 extends to the multiple logistic response function as follows:

$$E\{Y\} = \frac{\exp(\mathbf{X}'\boldsymbol{\beta})}{1+\exp(\mathbf{X}'\boldsymbol{\beta})}$$

and the equivalent simple logistic response form extends to:

$$E\{Y\} = [1+\exp(-\mathbf{X}'\beta)]^{-1}$$

Similarly, the logit transformation:

$$\pi' = \log_e\left(\frac{\pi}{1-\pi}\right)$$

now leads to the logit response function, or linear predictor:

$$\pi' = \mathbf{X}'\boldsymbol{\beta}$$

The multiple logistic regression model can therefore be stated as follows:

Y_i are independent Bernoulli random variables with expected values $E\{Y_i\} = \pi_i$, where

$$E\{Y_i\} = \pi_i = \frac{\exp(\mathbf{X}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i'\beta)} \tag{20}$$

This is called the *logistic mean response function*. Again, the X observations are considered to be known constants. Alternatively, if the X variables are random, $E\{Y_i\}$ is viewed as a conditional mean, given the values of $X_{i1}, \ldots, X_{i,p-1}$. Like the simple logistic response function (14.16), the multiple logistic response function (14.41) is monotonic and sigmoidal in shape with respect to $X'\beta$ and is almost linear when π is between .2 and .8. The X variables may be different predictor variables, or some may represent curvature or interaction effects. Also, the predictor variables may be quantitative, or they may be qualitative and represented by indicator variables. This flexibility makes the multiple logistic regression model very attractive.

5.1 Maximum Likelihood Estimation

Again, we shall utilize the method of maximum likelihood to estimate the parameters of the multiple logistic response function denoted by equation 20. The log-likelihood function for simple logistic regression in 17 extends directly for multiple logistic regression:

$$\log_e L(\boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i \left(\mathbf{X}_i'\boldsymbol{\beta}\right) - \sum_{i=1}^{n} \log_e \left[1 + \exp\left(\mathbf{X}_i'\boldsymbol{\beta}\right)\right]$$

Numerical search procedures are used to find the values of $\beta_0, \beta_1, \ldots \beta_{p-1}$ that maximize $\log_e L(\beta)$. These maximum likelihood estimates will be denoted by $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{p-1}$. Let $\hat{\boldsymbol{\beta}}$ denote the vector of the maximum likelihood estimates:

$$\underset{p\times 1}{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{bmatrix}$$

The fitted logistic response function and fitted values can then be expressed as follows:

$$\hat{\pi} = \frac{\exp\left(\mathbf{X}'\hat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{X}'\hat{\boldsymbol{\beta}}\right)} = \left[1 + \exp\left(-\mathbf{X}'\hat{\boldsymbol{\beta}}\right)\right]^{-1}$$

$$\hat{\pi}_i = \frac{\exp\left(\mathbf{X}_i'\hat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{X}_i'\hat{\boldsymbol{\beta}}\right)} = \left[1 + \exp\left(-\mathbf{X}_i'\hat{\boldsymbol{\beta}}\right)\right]^{-1}$$

where:

$$\mathbf{X}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_{p-1} X_{p-1}$$

$$\mathbf{X}_i'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_{p-1} X_{i,p-1}$$