Interpretation of the binary logistic regression

Example:

The explanatory variable is gender, which we have coded using an indicator variable with values $x=1$ for men and $x=0$ for women. The response variable is also an indicator variable. that is "Do you consume Alcohol". Thus, either responded consume alcohol or not.

The model says that the probability ($\pi(x)$) that the respondent consume alcohol depends upon the respondent gender ($x=1$ or $x=0$). So there are two possible value for $\pi(x)$, say $\pi(\text{men})$ and $\pi(\text{Women})$.

The logistic regression model specifies the relationship between $\pi(x)$ and x.

Since there are only two values for x, we write both equations. For men

$$\underbrace{\log\left(\frac{\pi(\text{men})}{1-\pi(\text{men})}\right)}_{\log(\text{odds men})} = \beta_0 + \beta_1 \qquad \text{—— (i)}$$

and for women

$$\underbrace{\log\left(\frac{\pi(\text{women})}{1-\pi(\text{women})}\right)}_{\log(\text{odds women})} = \beta_0 \qquad \text{—— (ii)}$$

Note that there is a β_1 term in the equation for men because $x=1$ but it is missing in the equation for women because $x=0$.

⇒ The fitted logistic regression model is (by software):

$$\underbrace{\log\left(\frac{\pi(x)}{1-\pi(x)}\right)}_{\log(\text{odds})} = -1.59 + 0.36x$$

where $b_0 = -1.59$ is the estimate of the intercept (simply the log(odds) for the women (ii).

& $b_1 = 0.36$ is the slop of the logistic regression model which shows the difference between the log(odds) for the men and the log(odds) for Women

Most people are not comfortable thinking in the log (odds) scale, so interpretation of the results in terms of the regression slope is difficult. Usually, we apply a transformation to help us. with a little algebra we can transforms the logistic regression slope into an odds ratio.

By (i) & (ii)

$odd_{men} = e^{\beta_0+\beta_1}$

$odd_{women} = e^{\beta_0}$

$$\text{odd Ratio} = \frac{odd_{men}}{odd_{women}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = \frac{e^{\beta_0} \cdot e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

So $e^{0.36} = 1.43$

$$\frac{odd_{men}}{odd_{women}} = 1.43$$

$$odd_{men} = 1.43\ odd_{women}$$

the odd for men are 1.43 times the odd for women.

<u>Interpretation for previous model</u>: ($\log(odds) = -1.59 + 0.36x$)

If gender change to men from women, the odd of the consuming alcohol is <u>increasing</u> by about 1.43 times
sign of the slop.

⇒ Had we coded women as 1 and men as 0, the signs of the parameters would be reversed. the fitted equation would be $\log(odds) = 1.59 - 0.36x$ and the odd ratio would be $e^{-0.36} = 0.70$.

<u>Interpretation:</u>

If gender change to women from men, the odd of the consuming alcohol is decreasing by about 0.70 times.

⇒ Now we consider the explanatory variable is quantitative.

we use the example of video in which here in multiple logistic regression (more than one explanatory variable are used) we can also called that binary logistic regression because dependent/response variable has two possible outcome.

dependent variable → Do you consume Alcohol? ⇒ categorical variable

Independent variables → Gender (x_1) ⇒ categorical variable

Age (x_2)
Self concept score (x_3)
Anxiety score (x_4)
] ⇒ Quantitative. variable

then he use forward selection technique to making model.

Step 1: self-concept score(x_3) in model
step 2: Self concept score(x_3) and anxiety score(x_4) in model
step 3: Age (x_2), self concept score (x_3) and anxiety score (x_4) are in model.

Here the gender was not included in the model because gender has an adequate impact on response variable.

Hosmer lemeshow goodness of fit test shows that the model of step 3 adequatly fits the data.

So the logistic regression model will be:

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = 6.228 - 0.075x_2 - 0.261x_3 + 0.245x_4$$

also the exp(β) are given in SPSS software results.

exp(b) ⇒ (0.928) (0.770) (1.277)

<u>Interpretation:</u>

If we increase the age (x_2) by one unite, the odds of the consuming alcohol is decreasing by about 0.9 times. If we increase the self concept score (x_3) by one unit the odds of the consuming →

alcohol is decreasing by about 0.8 times. And if we increase the anxity score(x_4) by one unit the odds of the consuming alcohol is increasing by about 1.3 times.

4