Solutions of Exercises for Logistic Regression (Chapter 15)

1. **a)** $\hat{\beta} = 0.02$. Since it is positive, we conclude that the curve of $\hat{P}(y = 1)$ is increasing with x; i.e the estimated probability of voting for the Republican candidate increases with the voter's total family income.
b) We have

$$\hat{P}(y = 1) = \frac{exp(-1.00 + 0.02x)}{1 + exp(-1.00 + 0.02x)} \tag{1}$$

(i) So, when income is 10 thousand dollars, estimated probability of voting for the Republican candidate would be

$$\frac{exp(-1.00 + 0.02(10))}{1 + exp(-1.00 + 0.02(10))} = 0.31 \tag{2}$$

(ii) Similarly, when income is 100 thousand dollars, estimated probability of voting for the Republican candidate is 0.73.
c) (i) $\hat{P}(y = 1) = 0.50$ when

$$x = -\frac{\hat{\alpha}}{\hat{\beta}} = -\frac{-1}{0.02} = 50. \tag{3}$$

i.e when the voter's income is 50 thousand Dollars.
(ii) Since the curve of $\hat{P}(y = 1)$ increases with x, the estimated probability of voting for the Republican candidate is more than 0.50 for incomes greater than 50 thousand Dollars.
d) For the region of x values for which $P(y = 1)$ is near 0.50, an approximate rate at which the above probability changes (increases) for 1 thousand dollar increase in income is

$$\frac{\hat{\beta}}{4} = 0.02/4 = 0.005. \tag{4}$$

i.e the probability of voting for the Republican candidate increases by 0.005 for every one thousand dollar increase in voter's income.
e) It can be shown that every unit increase in x has a multiplicative effect of $exp(\hat{\beta})$ on the odds. In our case, $exp(\hat{\beta}) = \exp(0.02) = 1.02$. So, we conclude that, for one thousand Dollar increase in income, the estimated odds of voting for the Republican increases by 2%.

3. **a)** The prediction equation is given by

$$logit[\hat{P}(y = 1)] = log[\frac{\hat{P}(y = 1)}{1 - \hat{P}(y = 1)}] = 2.0429 - 0.2821x. \tag{5}$$

where $\hat{P}(y = 1)$ is the estimated probability that symptoms of senility are present in an elderly person. Since $\hat{\beta}$ is negative, the curve of $\hat{P}(y = 1)$ decreases with x. So, the estimated probability of senility symptoms decreases at higher levels of WAIS.

b) $\hat{P}(y = 1) = 0.50$ at $x = -\frac{\hat{\alpha}}{\hat{\beta}} = -\frac{2.0429}{-0.2821} = 7.24$. Since $\hat{\beta}$ is negative, $\hat{P}(y = 1)$ is less than 0.50 for x greater than 7.24.

c) At $x = 20$, probability of senility would be

$$\frac{exp(2.0429 - 0.2821(20))}{1 + exp(2.0429 - 0.2821(20))} = 0.027. \tag{6}$$

d)For the linear probability model, estimated probability of senility at $x = 20$ is 0.847 - 0.051(20) = -.173 which is absurd (and hence doesn't make any sense) because a probaility can't be negative.

e) The P-value for this test is 0.0051 (from the given table). Since the P-value is quite small, there is strong evidence that the WAIS score has a negative association with symptoms of senility.

4.**a)**The prediction equation (obtained using SPSS) is given by

$$logit[\hat{P}(y=1)] = log[\frac{\hat{P}(y=1)}{1-\hat{P}(y=1)}] = 5.427 - 0.772x \tag{7}$$

b)The estimated probabilities are respectively:
(i)

$$\frac{exp(5.427 - 0.772(0))}{1 + exp(5.427 - 0.772(0))} = 0.9956. \tag{8}$$

and
(ii)

$$\frac{exp(5.427 - 0.772(20))}{1 + exp(5.427 - 0.772(20))} = 0.000045. \tag{9}$$

c) $\hat{P}(y=1) = 0.50$ when

$$x = -\frac{\hat{\alpha}}{\hat{\beta}} = -\frac{5.427}{-0.772} = 7.029. \tag{10}$$

Since $\hat{\beta}$ is negative, the estimated probability of senility is greater than 0.50 for x less than 7.029.
d) Since $exp(\hat{\beta}) = .46$, for a 1-unit increase in x, the estimated odds of senility multiply by 0.46.

7. **a)** Since the coefficient of m is 0.09, $exp(0.09)$ $(= 1.094)$, is the estimated odds ratio between mother's educational level and whether a child obtains a high school degree, controlling for the other predictors. Thus, we conclude that as mother's educational level increases by 1 year, the estimated odds that her child will get a high school degree increases by 9.4%
. **b)** Since $exp(-0.92) = 0.4$, we conclude that, controlling for the other predictors, the estimated odds that a child of a working mother gets a high school degree is 0.4 times the estimated odds that a child of a non-working mother gets a degree.
c) The coefficient of mother's occupational level is 0.21 and $exp(0.21) = 1.24$. So, controlling for the other predictors, the estimated odds that a child will get a high school degree increases by 24% for every unit increase in his/her mother's occupational level. So, the author was correct in making the interpretation.

9. **a)** Let A = 1 for AZT use and 0 otherwise,
R = 1 for blacks and 0 for whites and
Y = 1 for AIDS symptoms and 0 otherwise
The prediction equation is given by:

$$logit[\hat{P}(y=1)] = -1.0736 - 0.7195A + 0.0555R. \tag{11}$$

b) Since the coefficient of A (AZT) is negative, the estimated probability of developing AIDS symptoms decreases with AZT use, controlling for race, and since the coefficient of R (Race) is positive, the estimated probability of developing AIDS symptoms is slightly higher for Blacks compared to Whites, controlling for AZT use.
c) The estimated probability is given by :

$$\hat{P}(y=1) = \frac{exp(-1.0736 + 0.0555)}{1 + exp(-1.0736 + 0.0555)} = 0.265. \tag{12}$$

d) The estimated conditional odds ratio between AZT use and developing symptoms of AIDS will be $exp(-0.7195) = 0.486$. Thus, we conclude that, the estimated odds of developing AIDS symptoms for those using AZT is 0.486 times the estimated odds for those not using AZT.
e) Since the P-value corresponding to AZT for the Wald-Chi square test statistic is 0.0099, there is very strong evidence that, controlling for Race, AIDS symptoms are less likely to occur for those using AZT.

30. **a)** Taking logarithms of the respective Odds ratios, we get the prediction equation:

$$logit(\hat{P}(y = 1)) = C + 1.40G + 0.322S + 1.76SES + 1.17P \qquad (13)$$

Here Y is a dummy variable denoting whether or not the educational program makes adolescents more likely to use condoms, C is the intercept term (some constant), G, S and SES respectively denote the dummy variables for Group, Sex and SES, and P denotes lifetime number of partners.
b) The logarithms of the end points of the C.I. corresponding to Sex are (0.21, 2.56). Now, 1.38 lies in the middle of this C.I. Thus, the given C.I is not that of the odd's ratio but that of the log-odds ratio. So, if the given C.I. is true, then 1.38 should be the log-odds ratio and the true estimated Odds ratio should be $exp(1.38) = 3.97$.