STAT 659 Spring 2016
Homework 6 Solution

4.1

(a) From the output of the software, we know $\text{logit}\hat{\pi} = -3.7771 + 0.1449\text{LI}$. So for $\text{LI} = 8$, $\hat{\pi} = \frac{\exp(-3.7771+0.1449*8)}{1+\exp(-3.7771+0.1449*8)} = 0.068$.

(b) Similarly to (a), you can plug in $\text{LI} = 26$ and obtain the corresponding probability is 0.5.

(c) The rate of the change of the probability is $\pi(1-\pi)\beta$. When $\text{LI} = 8$, we have $\hat{\pi} = 0.068$. Then the rate of the change of the probability is $(0.068)(1-0.068)0.1449 = 0.00918$; for $\text{LI} = 26$, $\hat{\pi} = 0.4976$, so the rate of the change of the probability is 0.03622.

(d) Since $\hat{\pi}_{14} = 0.14824, \hat{\pi}_{28} = 0.56957$, then $\hat{\pi}_{28} - \hat{\pi}_{14} = 0.4213$.

(e) The multiplier for the estimated odds is $\exp(\beta) = 1.16$.

4.2

(a) The test statistic is $\frac{0.1449-0}{0.0593} = 2.4435$ which has a standard normal distribution. The P-value is $0.0145 < 0.05$, so we reject the null hypothesis and conclude that there exists LI effect.

(b) When LI increases by one unit, the log odds ratio increases by β. The 95 percent CI for β is $(0.02867, 0.26113)$, so the corresponding CI for the odds ratio is $(\exp(0.02867), \exp(0.26113)) = (1.029, 1.298)$.

(c) The likelihood ratio test statistic is $2\log l(\hat{\alpha}, \hat{\beta}) - 2\log l(\hat{\alpha}, 0) = 8.3$ with P-value 0.004, thus we reject the null hypothesis and conclude that there exists LI effect.

(d) When LI increases by one, the 95 percent CI likelihood ratio CI for odds ratio is $(\exp(0.0425), \exp(0.2846)) = (1.043416, 1.32923)$. We can see the CI excludes 1, so the LI effect exists.

4.5

(a) $\text{logit}\hat{\pi} = 15.0429 - 0.2322T$, where T is the temperature at the time of the flight and $\hat{\pi}$ is the estimated probability of thermal distress. We can see when the temperature increases by one, the log odds of the probability of thermal distress will decrease by -0.2322.

(b) $\hat{\pi} = \frac{\exp(15.0429-0.2322*31)}{1+\exp(15.0429-0.2322*31)} = 0.99961$.

(c) The temperature is $-\frac{\hat{\alpha}}{\hat{\beta}} = 64.78$ and the ratio of change in estimated probability is $\hat{\pi}(1-\hat{\pi})\hat{\beta} = 0.2322/4 = -0.05805$.

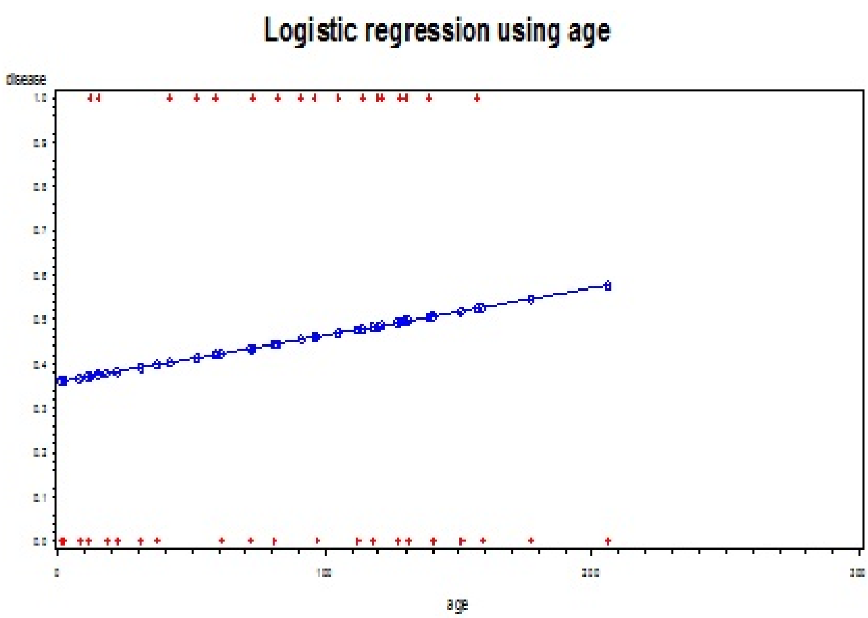
(d) When the temperature increases by one degree, the odds of the probability decreases by a multiplier $\exp(-0.2322) = 0.793$.

(e) The test statistic for Wald test is $(\frac{-0.2322}{0.1082})^2 = 4.60$ with P-value $0.032 < 0.05$; the likelihood ratio test statistic is 7.952 with P-value 0.0048. So we reject the null hypothesis and conclude that there is significant of temperature effect.

4.7

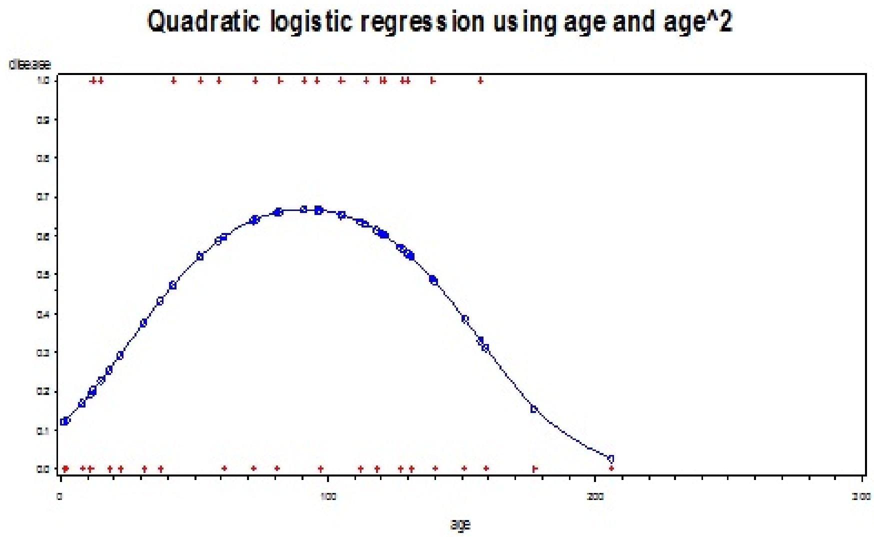
(a) The fitted logistic regression model is $\text{logit} = -0.5727 + 0.0043x$, where x is the age in months. The Wald test statistic is $\left(\frac{0.0043-0}{0.00585}\right)^2 = 0.539$ with P-value 0.463; the likelihood ratio test statistic is 0.5469 with P-value 0.4596. So we can not reject the null hypothesis that there is no effect of age.

(b) The plot of the data is as follows:



There exists more variability of age without kyphosis.

(c) The fitted model is $\text{logit}\pi(x) = -2.046 + 0.06x - 0.00033x^2$. The LR test statistic for the squared age term is 6.276, df=1, P-value 0.012. So the squared age term is significant.



4.8

(a) The prediction equation is $\text{logit}\pi(x) = -3.6947 + 1.8151x$, where x is the weight of the crab(unit is kg).

(b) $\hat{\pi}_{1.2} = \frac{\exp(-3.6947+1.8151*1.2)}{1+\exp(-3.6947+1.8151*1.2)} = 0.17997$; similarly, $\hat{\pi}_{2.44} = 0.67571, \hat{\pi}_{5.2} = 0.99681$.

(c) The weight is $-\frac{\hat{\alpha}}{\hat{\beta}} = 2.0355$.

(d) (i). The linear approximation of estimated effect is $\hat{\beta}/4 = 0.4538$;(ii). The estimated effect is $\hat{\beta}/4 * 0.1 = 0.0454$;(iii). The estimated effect is $\hat{\beta}/4 * 0.58 = 0.2633$.

(e) The 95 percent CI for β is $(1.0768, 2.5534)$, so the corresponding CI for the odds when weight increases by one unit is $(\exp(1.0768), \exp(2.5534)) = (2.9352, 12.8511)$.

(f) The Wald test statistic is 23.22 with P-value 1.443×10^{-6}; the likelihood ratio test statistic is 30.021 with P-value 4.274×10^{-8}. So we reject the null hypothesis and conclude that there is a weight effect.

4.9

(a) The fitted equation is $\text{logit}\pi = 1.0986 - 0.1226\text{col2} - 0.7309\text{col3} - 1.8608\text{col4}$. So the estimated odds ratio of probability for crabs having satellite with color two is $\exp(-0.1226) = 0.8846$ times as that for the crab with lightest color.

(b) The likelihood ratio test statistic is 13.698 which has a chi-squared distribution with degree of freedom 3. The P-value is 0.0033, so there is significant evidence of color effect.

(c) The fitted model is logit$\pi = 2.3635 - 0.7147$col. So when color increased by one unit, the odds will decreases by a multiplier $\exp(-0.7147) = 0.4893$.

(d) The test statistic is 12.461 with a chi-squared one distribution. The P-value is 0.00042, so there is significant evidence of color effect.

(e) We can see the test has more power if we see the color variable as a quantitative variable. But the disadvantage of treating the color as a quantitative variable is that the log odds may not have a linear trend with the color.

4.15

(a) The test statistic is 7.815, which has a chi-squared distribution with $df = 1$. The P-value is 0.0052, so we reject the null hypothesis that the merit pay decision is independent of the race, conditional on the district.

(b) Let $x = \begin{cases} 1 & \text{Black} \\ 0 & \text{white} \end{cases}$ and $z_i = \begin{cases} 1 & \text{when data belongs to district i} \\ & 0 \;\; \text{otherwise} \end{cases}$. The Wald test for β is $(\frac{\hat{\beta - 0}}{se_{\hat{\beta}}})^2 = 7.6914$ with P-value $0.0055 < 0.05$. So we reject the null hypothesis and conclude that there is an effect for the race.

(c) A model-based analysis in (b) gives information about size of effect. For example: the estimated odds ratio between pay decision and race controlling for district is $\exp(0.791) = 2.206$.

(d) The likelihood ratio test for equality of odds ratios between race and merit pay for the five districts is equivalent to the test of the hypothesis that all interaction terms between race and the five district are zero. The test statistic is 2.07, which has a chi-squared distribution with $df = 4$. The P-value is 0.7227, so we fail to reject the null hypothesis that odds ratios between race and merit pay for the five districts are equal. The result of the likelihood ratio test is consistent with the one of the Breslow-Day test. The test statistic of the Breslow-Day test is 2.15 with a chi-squared distribution 4 df, and the P-value is 0.7227.

4.16

(a) $\text{logit}(\hat{\pi}(x)) = -2.2086 + 0.555x_1 + 0.4292x_2 - 0.6873x_3 - 0.2022x_4$, where x_1, x_2, x_3 and x_4 are 1 for E,N,F,J respectively.

(b) $\text{logit}(\hat{\pi}(x_{ESTJ})) = -2.2086 + 0.555 - 0.2022 = -1.8558$.
$\hat{\pi}(x_{ESTJ}) = (1 + e^{-(-1.8558)})^{-1} = 0.135$

(c) since $\hat{\pi} = odds/(1+odds) = 1 - 1/(1+odds)$, the estimated probability attains its maximum when the corresponding odds attains the maximum. From the fitted equation, we can see the personality type ENTP keeps all the positive coefficients and let all negative coefficients be zero, so it has the maximum odds and also the maximum estimated probability.

4.17

(a) $\pi = \frac{\exp(-2.8291)}{1+\exp(-2.8291)} = 0.056$

(b) The odds ratio is $\exp(0.5805) = 1.79$

(c) $(\exp(0.1589), \exp(1.008)) = (1.172, 2.740)$

(d) The odds ratio is $\exp(-0.5805) = 0.560$ and the CI is $(\exp(-1.0080), \exp(-0.1589)) = (0.3649, 0.8531)$.

(e) The likelihood ratio test statistic is 7.28 with P-value $0.007 < 0.05$. So we reject the null hypothesis that there is no E/T effect on the response.

4.19

(a) Since $\exp(0.16) = 1.1735$, so the odds ratio of supporting legalized abortion between females/males is 1.1735.

(b) (i). $\text{logit}\hat{\pi}_1 = -0.11 - 0.66 - 1.67 = -2.44$, so $\hat{\pi}_1 = 0.0802$;(ii)$\text{logit}\hat{\pi}_2 = -0.11 + 0.16 + 0.84 = 0.89$, so $\hat{\pi}_2 = 0.7089$.

(c) In this case, $\hat{\beta}_2^G = -0.16$, then $\frac{\exp(0)}{\exp(-0.16)} = 1.1735$.

(d) $\hat{\beta}_1^G = 0.08, \hat{\beta}_2^G = -0.08$. So the odds ratio for gender is $\frac{\exp(0.08)}{\exp(-0.08)} = 1.1735$.

4.22

(a) The fitted model equation is $\text{logit}\pi = -4.5266 + 1.2694\text{col1} + 1.4143\text{col2} + 1.0833\text{col3} + 1.6928\text{weight}$. So the crabs with median color have the highest probability to have the satellites. When weight increases by one kg, the odds of the probability of satellites will increases by $\exp(1.6928) = 5.435$.

(b) The likelihood ratio test statistic is 7.1949 which has a chi-squared distribution with $df = 3$. The P-value is $0.0659 > 0.05$, so the color is not significant when $\alpha = 0.05$.

(c) (i). The fitted model is $\text{logit}\hat{\pi} = -2.0316 - 0.5142\text{col} + 1.6531\text{weight}$, so the light color crabs have the highest probability to have satellites; (ii). The fitted model is $\text{logit}\pi = -3.6947 + 1.8151\text{weight}$. The likelihood ratio test statistic is 5.4684 with P-value $0.0194 < 0.05$. So color is significant if we assign quantitative scores to it.

4.24

(a) The fitted model is $\text{logit}\hat{\pi} = -1.41734 + 0.06868D - 1.65895T$. So the odds ratio between the tracheal tube and laryngeal mask airway is $\exp(-1.65895) = 0.1903$; when the duration of the surgery increases by one minute, the odds will increases by a multiplier $\exp(0.06868) = 1.07$.

(b) Since $\hat{\beta}_D = 0.06868$ with estimated standard error 0.02641, the Wald test statistic is $2.6^2 = 6.76$ with P-value 0.00932. So we reject H_0 that there is no effect for the duration of the surgery.

(c) The fitted model is $\text{logit}\hat{\pi} = 0.04979 + 0.02848D - 4.47224T + 0.0746DT$. So when $T = 1$, $\text{logit}\hat{\pi} = -4.42245 + 0.10308D$; when $T = 0$, $\text{logit}\hat{\pi} = 0.04979 + 0.02848D$.

(d) The likelihood ratio test statistic for the interaction term is 1.8169 with P-value $0.17768 > 0.05$. Thus the interaction term is not statistically significant.

4.30

For no interaction model:
$\text{logit}\pi = -0.8537 + 1.0151R + 0.3524G$, where R : race G : gender. Using a 0.05 significance level, Race given gender was significant and Gender given race was significant. For given gender, we are 95 % confident that the odds of graduating for a White athlete are between $(2.326088, 3.273998)$ times higher than that for a black athlete. For given Race, we are 95 % confident that the odds of graduating for a female are between $(1.22, 1.67)$ times higher than that for a male.
For interaction model:
To test whether there is interaction effect, $H_0 : \beta_{interaction} = 0 \quad vs. \quad H_a : \beta_{interaction} \neq 0$. Wald Chi-square Statistics = 0.0002, $df = 1$, p-value = 0.9889. We fail to reject H_0. There is no significant evidence of interaction.

Additional Problem

(a) Let g be the indicator variable for the gender, $g = \begin{cases} 1 & \text{Male} \\ 0 & \text{otherwise} \end{cases}$ For department one, the fitted model is $\text{logit}\pi = 1.5442 - 1.0521g$ and the standard error for the coefficient of the gender is 0.2627. So the 95 percent CI for conditional odds ratio between male and female is $(\exp(-1.0521 - 1.96 * 0.2627), \exp(-1.0521 + 1.96 * 0.2627)) = (0.2086, 0.5844)$. For department two, the fitted model is $\text{logit}\pi = 0.7538 - 0.22g$ and the CI for conditional odds ratio is $(0.3404, 1.8921)$. For department 3, the fitted model is $\text{logit}\pi = -0.66044 + 0.12492g$ and the CI for conditional odds ratio is $(0.8545, 1.5024)$. For department 4, the fitted model is $\text{logit}\pi = -0.62197 - 0.08199g$ and the CI for conditional odds ratio is $(0.6863, 1.2367)$. For department 5, the fitted model is $\text{logit}\pi = -1.1571 + 0.2002g$ and the CI for conditional odds ratio is $(0.8251, 1.8087)$. For department 6, the fitted model is $\text{logit}\pi = -2.5808 - 0.1889g$ and the CI for conditional odds ratio is $(0.4552, 1.5058)$.

(b) The CMH test is to test the conditional independence of the gender and the admission decision, conditional on the department. To test whether the gender is independent with the admission decision, we fit a logistic model to the data with gender and the indicator variables of each department as covarites. The fitted model is $\text{logit}\pi = 0.68192 - 0.09987g - 0.04340\text{dep2} - 1.26260\text{dep3} - 1.29461\text{dep4} - 1.73931\text{dep5} - 3.30648\text{dep6}$. The null hypothesis is that there is no effect for the gender. The Wald test statistic is 1.526 with P-value 0.2167, so the gender effect is not significant. The Breslow-Day test is to test whether $\beta_i^{depart*gender} = 0$ for $i = 1, ..., 5$. The test statistic is 17.9011, df=5 with P-value 0.0031. We reject the null hypothesis and conclude that the odds ratios among departments are not homogeneous.

(c) We fit a logistic model for the data with out the department one with gender,the indicator variables of the department and the interaction terms as covariates. To check if common odds ratio is reasonable, we have to test $\beta_i^{depart*gender} = 0$. The Wald Chi-square statistics is 2.5532, df=4, p-value=0.6351, so the common odds ratio is possible. The estimated common odds ratio is $\exp(0.03069) = 1.031166$, and the 90 percent CI for it is $(0.89403, 1.18934)$.

Only for students having taken STAT 414, 610, 630, or another mathematical statistics course

4.34

When $x = 0$, $\text{logit}\pi = \alpha$. So $\frac{\pi}{1-\pi} = e^{\alpha}$. For $x = 1, 2, 3$, $\frac{\pi}{1-\pi} = e^{\alpha+\beta}, e^{\alpha+2\beta}, e^{\alpha+3\beta}$ respectively. We can see when x increases by one, the odds ratio will increases by a multiplier e^{β}. If we generalize it to the multiple logistic regression model, then β_i can be interpreted as the increase of the log odds ratio when x_i increases by one unit while keeping other co-variates fixed.

4.35

(a) Let $y = 0.4\beta \exp(-(\alpha + \beta x)^2/2)$. Then $\frac{dy}{dx} = 0.4\beta^2(\alpha + \beta x)\exp(-(\alpha + \beta x)^2/2)$. Letting it equal to 0 yields $x = -\frac{\alpha}{\beta}$. Then $y = 0.4\beta$.

(b) $x = -\frac{\alpha}{\beta} = \frac{7.502}{0.302} = 24.8$

(c) $0.4\beta exp(-(\alpha + \beta x)^2/2) = 0.4 \times 0.302 = 0.121$