STAT 659 Spring 2016
Homework 9 Solution

5.21

For the conditional exact test, the score test statistic of alcohol effect is 6.5699 with P-value 0.0172, so the alcohol effect is significant which is contrary to the result in 5.20 part (a).

5.22

(a) For the perfect fit, when $x \leq 40$, then $\hat{\pi} = 0$; when $x > 60$, then $\hat{\pi} = 1$. So the logitπ increases from $-\infty$ to $+\infty$ when x increases from 40 to 60 to produce a perfect fit, then $\hat{\beta} = \infty$ for this model.

(b) $\hat{\beta} = -0.633, \text{se}(\hat{\beta}) = 1.6721$ in SAS proc logistic.

(c) After adding these two data, $\hat{\beta} = -0.7328, \text{se}(\hat{\beta}) = 2.7575$. It is not correct, because the gap at $x = 50$ is even more shaper than part (a) and the algorithm for obtaining MLE does not converge.

(d) $\hat{\beta} = 0.5277$ and $\text{se}(\hat{\beta}) = 0.9816$. In this case, the collection of covariates for two outcome groups overlaps and there is no complete separation. So the algorithm converges and the results are correct.

5.23

(a) If we let $Y = 1$ when patients got cured and $Y = 0$ otherwise, then the fitted model is $\text{logit}\pi = -2.5496\text{delay} - 20.7932\beta_1^Z - 0.2339\beta_2^Z + 2.2592\beta_3^Z + 4.2626\beta_4^Z + 23.1827\beta_5^Z$. For Penicillin level 1/8, $Y = 0$ for all responses, so $\text{logit}\pi = -\infty$ for the perfect fit. Thus $\hat{\beta}_1^Z = -\infty$. Similarly, for Penicillin level 4, $Y = 1$ for all responses. Then $\text{logit}\pi = +\infty$ for the perfect fit. Thus $\hat{\beta}_5^Z = +\infty$. The software gives a huge value of standard deviations for these two variables.

(b) The P-value for the Wald test on the coefficient of delay variable is $0.03 < 0.05$, so delay effect is significant and XY is not conditional independent. Also the deviance for the model with delay variable and the the model without delay variable is $6.7994 \sim \chi_1^2$. The P-value is $0.009 < 0.05$ which implies Y depends on X, conditional on Z.

(c) The conditional odds ratio is $\exp(-2.5496) = 0.0781$. So conditional on the penicillin level, the odds of cured proportion with delay is 0.0781 times of that without delay.

5.24

For the conditional exact test, the P-value of the score test for delay effect is $0.0399 < 0.05$, so the delay effect is significant. The estimate for the delay effect is -2.3381. So the conditional odds ratio is 0.0965 which is similar to the results in the previous question.

6.1

(a) $\log\left(\frac{\hat{\pi}_{\mathrm{R}}}{\hat{\pi}_{\mathrm{D}}}\right) = \log\left(\frac{\hat{\pi}_{\mathrm{R}}}{\hat{\pi}_{\mathrm{I}}}\right) - \log\left(\frac{\hat{\pi}_{\mathrm{D}}}{\hat{\pi}_{\mathrm{I}}}\right) = -2.3 + 0.5x$. Since $\exp(0.5) = 1.649$, then the estimated odds of preferring Republicans over Democrats increase by 64.9% for every $10,000$ increase.

(b) $\frac{\hat{\pi}_{\mathrm{R}}}{\hat{\pi}_{\mathrm{D}}} = \exp(-2.3 + 0.5x) > 1 \Rightarrow -2.3 + 0.5x > 0 \Rightarrow x > 4.6$.

(c) Since $\hat{\pi}_D = \exp(3.3 - 0.2x)\hat{\pi}_I$ and $\hat{\pi}_R = \exp(1.0 + 0.3x)\hat{\pi}_I$, then by plugging in $\hat{\pi}_D, \hat{\pi}_R$ in the equation $\hat{\pi}_R + \hat{\pi}_D + \hat{\pi}_I = 1$, we can obtain $\hat{\pi}_{\mathrm{I}} = \frac{1}{1+e^{3.3-0.2x}+e^{1+0.3x}}$.

6.2

(a) The odds ratio is $\exp(-2.4654) = 0.08497$.

(b) From the output, we know $\log\frac{\hat{\pi}_F}{\hat{\pi}_o} = 1.6177 - 0.1101x, \log\frac{\hat{\pi}_I}{\hat{\pi}_o} = 5.6974 - 2.4654x$, then $\log\frac{\hat{\pi}_F}{\hat{\pi}_I} = -4.0797 + 2.3553$. By plugging in $\hat{\pi}_F, \hat{\pi}_o$ into the equation $\hat{\pi}_F + \hat{\pi}_o + \hat{\pi}_I = 1$, we can obtain $\hat{\pi}_I = \frac{1}{1+\exp(-4.0797+2.3553x)+\exp(-5.6974+2.4654x)}$. So for $x = 3.9$, the probability is 0.00462.

(c) Let $\log\frac{\hat{\pi}_I}{\hat{\pi}_o} = 5.6974 - 2.4654x = 0$ and we can obtain $x = 2.3109$.

6.3

When using <u>Fish</u> as the baseline, $x_h = \begin{cases} 1 & \text{Hancock} \\ 0 & \text{o.w.} \end{cases}$; $x_o = \begin{cases} 1 & \text{Oklawaha} \\ 0 & \text{o.w.} \end{cases}$; $x_t = \begin{cases} 1 & \text{Trafford} \\ 0 & \text{o.w.} \end{cases}$; $x_s = \begin{cases} 1 & \text{Size} \leq 2.3 \\ 0 & \text{o.w.} \end{cases}$.

(a)
$$\log\left(\frac{\hat{\pi}_{\mathrm{I}}}{\hat{\pi}_{\mathrm{F}}}\right) = -1.549 - 1.6583x_h + 0.9372x_o + 1.122x_t + 1.4582x_s.$$
$$\log\left(\frac{\hat{\pi}_{\mathrm{R}}}{\hat{\pi}_{\mathrm{F}}}\right) = -3.3139 + 1.2422x_h + 2.4583x_o + 2.9347x_t - 0.3513x_s.$$
$$\log\left(\frac{\hat{\pi}_{\mathrm{B}}}{\hat{\pi}_{\mathrm{F}}}\right) = -2.0931 + 0.6951x_h - 0.6532x_o + 1.0878x_t - 0.6307x_s.$$

$$\log\left(\frac{\hat{\pi}_{\mathrm{O}}}{\hat{\pi}_{\mathrm{F}}}\right) = -1.9043 + 0.8262x_h + 0.00565x_o + 1.5164x_t + 0.3316x_s.$$

(b)

	$\hat{\pi}_{\text{Fish}}$
Size ≤ 2.3	0.25819
Size > 2.3	0.45844

6.4

(a) The probability is $\frac{\exp(0.883+0.419)}{1+\exp(0.883+0.419)+\exp(-0.758+0.105)} = 0.7074$.

(b) The estimated odds ratio for (i) is $\exp(0.105) = 1.1107$; the estimated odds ratio for (ii) is $\exp(0.419 - 0.105) = 1.3689$.

6.5

(a) The subjects will be more satisfied if $\text{logit}[\hat{P}(Y \leq j)]$ is lower, so when x_1 increases and x_2, x_3 decrease, the subjects tend to be more satisfied.

(b) $x_1 = 4, x_2 = x_3 = 1$.

6.6

(a) Based on the output, the prediction equations are:
$\log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right) = -2.5551 - 0.2275x$.
$\log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) = -0.3513 - 0.0962x$.

(b) When income level increases by one, the estimated odds of being unhappy instead of very happy is $\exp(-0.2275) = 0.7965$ for the below average income family.

(c) From the outputs, the Wald test statistic is 0.9432 with P-value $0.624 > 0.05$. So the marital happiness is independent of family income.

(d) The Pearson's goodness of fit test statistic is 3.1510 with P-value $0.2069 > 0.05$, so the model fit is adequate.

(e) The estimated probability is $\frac{1}{1+\exp(-2.5551-0.2275(2))+\exp(-0.3513-0.0962(2))} = 0.6135$.

6.7

(a) Because the model assumes that the predictor effect x is identical for all two cumulative logits, then there is only one income effect; there are two cumulative probabilities, so there are two cumulative logit model of two different intercepts.

(b) When income increases by one, then the estimated odds of marital happiness below a certain level will multiply by $\exp(-0.1117) = 0.8943$.

(c) From the outputs, the likelihood ratio test statistic for the income is 0.8876 with P-value $0.3461 > 0.05$, which indicates that the martial happiness is independent of family income.

(d) The Pearson's goodness of fit test statistic is 3.2292 with P-value $0.3576 > 0.05$, so the model is adequate.

(e) It asks for the probability $P(Y = 3) = 1 - P(Y \leq 2) = 1 - \frac{\exp(-0.2378-2*0.1117)}{1+\exp(-0.2378-2*0.1117)} = 1 - 0.3867 = 0.6133$.

6.8

Let ordinal response, from Progressive Disease to Complete Remission: $j = 1 \sim 4$;
$x_1 = \begin{cases} 1, & \text{Therapy is alternating} \\ 0, & \text{o.w.} \end{cases}; x_2 = \begin{cases} 1, & \text{Male} \\ 0, & \text{Female} \end{cases}.$

(a) $\text{logit}(\hat{P}(Y \leq j)) = \hat{\alpha}_j + 0.5807x_1 - 0.5414x_2$, where $\begin{cases} \hat{\alpha}_1 = -0.7767 \\ \hat{\alpha}_2 = 0.7906 \\ \hat{\alpha}_3 = 1.8414 \end{cases}.$
Controlling for gender, at alternating therapy, the estimated odds of response to chemotherapy below at any fixed level is $e^{0.5807} \approx 1.79$ times the estimated odds of that at sequential therapy. It also suggests that the sequential therapy has more remission.

(b) $\text{logit}(\hat{P}(Y \leq j)) = \hat{\alpha}_j + 1.0786x_1 - 0.274x_2 - 0.5906x_1x_2$, where $\begin{cases} \hat{\alpha}_1 = -1.0016 \\ \hat{\alpha}_2 = 0.5697 \\ \hat{\alpha}_3 = 1.6192 \end{cases}.$
For females, the estimated effect of therapy is 1.0786. For males, the estimated effect of therapy is 0.488. Therefore, the impact of therapy seems to be more severe for females.

(c) $H_0 : \beta_{\text{interaction}} = 0$ $v.s.$ $H_a : \beta_{\text{interaction}} \neq 0$
(i) $\chi^2 = 0.9901$, $df = 1$, p-value $= 0.3197$.
(ii)$G^2 = 0.953$, $df = 1$, p-value $= 0.329$.
At significant level $\alpha = 0.05$, we fail to reject H_0. There is no evidence that the model with interaction gives a significantly better fit.

6.9

(a) (i).There are four $\{\hat{\alpha_j}\}$ because there are four cumulative probabilities to model. (ii). Because the model is $\text{logit}P(Y \leq j) = \alpha_j + \beta_1\text{Protestant} + \beta_2\text{Catholic} + \beta_3\text{Jewish}, j \leq 4$, where Protestant, Catholic and Jewish are indicator variables. When all indicator variables are zero, since the cumulative probability is increasing with j and so does the $\text{logit}P(Y \leq j)$, $\hat{\alpha_1} < \hat{\alpha_2} < \hat{\alpha_3} < \hat{\alpha_4}$.

(b) (i). Since logit$P(Y \leq 1) = \alpha_1 + \beta_1$Protestant $+ \beta_2$Catholic $+ \beta_3$Jewish, then to make this value to be the largest one, we need to set all indicator variables to be zero because they all have negative coefficients. Therefore, the None religious preference group is most liberal. (ii). For the most conservative group, the logit$P(Y \leq 4)$ should have the smallest value. Since the Protestant variable has the smallest coefficient, thus the Protestant group is most conservative.

(c) For Protestant group, the probability is $\frac{\exp(-1.03-1.27)}{1+\exp(-1.03-1.27)} = 0.0911$ while the probability of None group is $\frac{\exp(-1.03)}{1+\exp(-1.03)} = 0.2631$.

(d) For the Protestant and None group, the estimated odds that Protestant falls in relatively more liberal categories (rather than more conservative categories) is $\exp(-1.27) = 0.2808$ times estimated odds for someone with no religious preference. For the Protestant and Catholic group, the estimated odds that Protestant falls in relatively more liberal categories (rather than more conservative categories) is $\exp(-1.27 + 1.22) = 0.9512$ times estimated odds for someone with Catholic religious preference.

Only for students having taken STAT 414, 610 or STAT 630

Fill in the mathematical details for the derivation of the probabilities for the three categories from the logits on slide 6, Chapter 6

Since $\log\frac{\pi_{i,1}}{\pi_{i,3}} = \alpha_1 + \beta_1 x_i, \log\frac{\pi_{i,2}}{\pi_{i,3}} = \alpha_2 + \beta_2 x_i$, then $\pi_{i,1} = e^{\alpha_1+\beta_1 x_i}\pi_{i,3}, \pi_{i,2} = e^{\alpha_2+\beta_2 x_i}\pi_{i,3}$. Then plug in $\pi_{i,1}, \pi_{i,2}$ into the equation $\pi_{i,1} + \pi_{i,2} + \pi_{i,3} = 1$, we can obtain that $\pi_{i,3} = \frac{1}{1+e^{\alpha_1+\beta_1 x_i}+e^{\alpha_2+\beta_2 x_i}}$. Thus $\pi_{i,1} = \frac{e^{\alpha_1+\beta_1 x_i}}{1+e^{\alpha_1+\beta_1 x_i}+e^{\alpha_2+\beta_2 x_i}}$ and $\pi_{i,2} = \frac{e^{\alpha_2+\beta_2 x_i}}{1+e^{\alpha_1+\beta_1 x_i}+e^{\alpha_2+\beta_2 x_i}}$.

Fill in the mathematical details on slide 30, Chapter 6

$$
\begin{aligned}
\text{logit}P(Y \leq j | Y > j-1) &= \text{logit}\frac{P(Y \leq j \cap Y > j-1)}{P(Y > j-1)} \\
&= \text{logit}\frac{P(Y = j)}{P(Y > j-1)} \\
&= \log\frac{\frac{P(Y=j)}{P(Y>j-1)}}{1 - \frac{P(Y=j)}{P(Y>j-1)}} \\
&= \log\frac{P(Y = j)}{P(Y > j-1) - P(Y = j)} \\
&= \log\frac{P(Y = j)}{P(Y > j)} \\
&= \log\frac{\pi_j}{\pi_{j+1} + \pi_{j+2} + \cdots + \pi_J}
\end{aligned}
$$