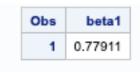
Patrick Fitzgerald ADS 534 Statistical Modeling HW 7

- 1. The dataset lowbwt.sas7bdat contains information for the sample of 100 low birth weight infants born in Boston, Massachusetts. The variable grmhem is a dichotomous random variable indicating whether an infant experienced a germinal matrix hemorrhage. The value 1 indicates that a hemorrhage occurred and 0 that it did not. The infants' five-minute apgar scores are saved under the name apgar5, and indicators of toxemai where 1 represents a diagnosis of toxemia during pregnancy for the child's mother and 0 no such diagnosis under the variable name tox.
- . (a) Write down the equation for a logistic regression model where germinal matrix hemorrhage is the response and five-minute appar score is the predictor, using  $\beta_1$  to represent the regression coefficient of appar score.

grmhem = 
$$\beta_0 + \beta_1$$
apgar5 + $\epsilon$ 

. (b) Fit the logistic regression model in part (a) in SAS. What is  $\hat{\beta}_1$ , the estimated regression coefficient of apgar score? What's the interpretation of  $\hat{\beta}_1$ ?

| Analysis of Maximum Likelihood Estimates |    |          |                   |                    |            |  |
|--|----|----------|-------------------|--------------------|------------|--|
| Parameter                                | DF | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSq |  |
| Intercept                                | 1  | -0.3037  | 0.6191            | 0.2407             | 0.6237     |  |
| apgar5                                   | 1  | -0.2496  | 0.1044            | 5.7206             | 0.0168     |  |



The odds ratio of grmhem associated with a 1 unit increase in appar5 is  $\exp(\beta_1) = \exp(-0.2496) = 0.77911$ . Therefore, higher appar5 is associated with larger odds of grmhem.

. (c) If a particular child has a five-minute appar score of 3, what is the predicted probability that this child will experience a brain hemorrhage?

$$\exp(\beta_0 + \beta_1 * 3) = \exp(-0.3037 - 0.2496*3) = 0.34906$$

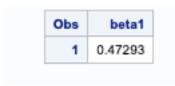
 $Pr(grmhem|apgar5=3) = 0.34906/(1+0.34906) = 0.2587431 = \sim 25.8\%$  that the child will experience a brain hemorrhage.

. (d) What is the estimated odds ratio of suffering a germinal matrix hemorrhage associated with 1 unit increase in five-minute appar score?

From part (a) we see that the odds ratio of grmhem associated with a 1 unit increase in apgar5 is  $\exp(\beta_1) = \exp(-0.2496) = 0.77911$ .

. (e) What is the estimated odds ratio of suffering a germinal matrix hemorrhage associated with 3 units increase in five-minute appar score?

$$\exp(3*\beta_1) = 0.47293$$



. (f) Write down the equation for a logistic regression model where germinal matrix hemorrhage is the response and toxemia status is the predictor, using  $\beta 2$  to represent the regression coefficient of toxemia status.

grmhem = 
$$\beta_0 + \beta_2 tox + \epsilon$$

. (g) Fit the logistic regression model in part (h) in SAS. What is  $\hat{\beta}_2$ , the estimated regression coefficient of toxemia status? What's the interpretation of  $\hat{\beta}_2$ ?

| Analysis of Maximum Likelihood Estimates |    |          |                   |                    |            |  |
|--|----|----------|-------------------|--------------------|------------|--|
| Parameter                                | DF | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSo |  |
| Intercept                                | 1  | -1.5353  | 0.2946            | 27.1530            | <.000      |  |
| tox                                      | 1  | -1.4604  | 1.0662            | 1.8761             | 0.170      |  |

| Obs | beta2   |
|-----|---------|
| 1   | 0.23214 |

The odds ratio of grmhem associated with a 1 unit increase in tox is  $\exp(\beta_2) = \exp(-1.4604) = 0.23214$ . Therefore, higher tox is associated with larger odds of grmhem.

. (h) For a child whose mother was diagnosed with toxemia during pregnancy, what is the predicted probability of experiencing a germinal matrix hemorrhage?

$$\exp(\beta_0 + \beta_2 * 1) = \exp(-1.5353 - 1.4604 * 1) = 0.05$$

Pr(grmhem|tox=1) =  $0.05/(1+0.05) = 0.04761905 = \sim 4.8\%$  that the child will experience a brain hemorrhage.

. (i) What are the estimated odds of suffering a germinal matrix hemorrhage for children whose mothers were diagnosed with toxemia relative to children whose mothers were not?

- 2. The dataset lowbwt.sas7bdat contains information for the sample of 100 low birth weight infants born in Boston, Massachusetts. The variable grmhem is a dichotomous random variable indicating whether an infant experienced a germinal matrix hemorrhage. The value 1 indicates that a hemorrhage occurred and 0 that it did not. The infants' five-minute apgar scores are saved under the name apgar5, and indicators of toxemai where 1 represents a diagnosis of toxemia during pregnancy for the child's mother and 0 no such diagnosis under the variable name tox. First, we fit a logistic regression model where germinal matrix hemorrhage is the response and five-minute apgar score is the predictor, using  $\beta 1$  to represent the regression coefficient of apgar score.
- . (a) At the 0.05 significance level, test the null hypothesis that  $\beta 1$  is equal to 0 using Wald test in SAS. What is the value of the test statistic? What's the distribution of the test statistic (including degrees of freedom)? What do you conclude?

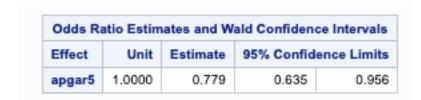
| Testing Global Null Hypothesis: BETA=0 |            |    |            |  |  |
|--|------------|----|------------|--|--|
| Test                                   | Chi-Square | DF | Pr > ChiSq |  |  |
| Likelihood Ratio                       | 5.6150     | 1  | 0.0178     |  |  |
| Score                                  | 6.3451     | 1  | 0.0118     |  |  |
| Wald                                   | 5.7206     | 1  | 0.0168     |  |  |

We see a Wald value of 5.7206, which follows the Chi-Square distribution with 1 degree of freedom. We see that it has a p-value  $(0.0168) < \alpha (0.05)$  thus we reject the null  $(\beta_1 = 0)$  and say that appar5 is significantly associated with grmhem.

We could also check if the 95% CI for apgar5 contains 0 and determine whether to reject / fail to reject. We see the interval (-0.4542, -0.0451) which doesn't contain 0, therefore we would reject the null as previously stated.

| Parameter Estimates and Wald Confidence Intervals |          |              |           |  |  |  |
|---|----------|--------------|-----------|--|--|--|
| Parameter   | Estimate | 95% Confider | ce Limits |  |  |  |
| Intercept   | -0.3037  | -1.5170      | 0.9096    |  |  |  |
| apgar5  | -0.2496  | -0.4542      | -0.0451   |  |  |  |

. (b) Construct a 95% Wald test based confidence interval for the population odds ratio of suffering a germinal matrix hemorrhage associated with 1 unit increase in five-minute apgar score. Does this interval contain the value 1? What does this tell you? Next, fit a logistic regression model where germinal matrix hemorrhage is the response and toxemia status is the predictor, using β2 to represent the regression coefficient of toxemia status.



We see our 95% CI is (0.635, 0.956) which does not contain the value 1 which tells us that the odds ratio is significantly different from 1. This again confirms that appar5 is significantly associated with grmhem.

. (c) At the 0.05 significance level, test the null hypothesis that β2 is equal to 0 using likelihood ratio test in SAS. What is the value of the test statistic? What's the distribution of the test statistic (including degrees of freedom)? What do you conclude?



We see a Likelihood Ratio value of 2.69, which follows the Chi-Square distribution with 1 degree of freedom. We see that it has a p-value (0.1008) >  $\alpha$  (0.05) thus we fail to reject the null ( $\beta_2$  = 0) and say that tox is not significantly associated with grmhem.

. (d) Construct a 95% profile likelihood based confidence interval for the population odds ratio of suffering a germinal matrix hemorrhage for children whose mothers were diagnosed with toxemia relative to children whose mothers were not. Does this interval contain the value 1? What does this tell you?

- 3. A sample of 2500 subjects was involved in a study on the association of patient death rate following admission to an adult intensive care unit (ICU) and hospital discharge of these patients. The data file icu.sas7bdat contains six columns correspond to id, sta (vital status; 0=alive, 1=dead), age (in years), sex (1=female, 0=male), race (1=white, 2=black, 3=other), crn (history of chronic renal failure; 0=no, 1=yes). Let  $\pi$  be the probability of death following admission to ICU.
- . (a) Fit a logistic model of sta on crn and race using dummy variables

$$\log{\{\pi/(1-\pi)\}} = \beta_0 + \beta_1 \operatorname{crn} + \beta_2 \operatorname{race2} + \beta_3 \operatorname{race3}$$

Where race2=1 if race=black and 0 otherwise, and race3=1 if race=other and 0 otherwise. Test for no association between crn and sta (i.e., H0 :  $\beta$ 1 = 0) adjusting for race using likelihood ratio test based on the logistic model (1).

| Analysis of Maximum Likelihood Estimates |    |          |                   |                    |            |
|--|----|----------|-------------------|--------------------|------------|
| Parameter                                | DF | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSq |
| Intercept                                | 1  | -1.2008  | 0.0811            | 219.2711           | <.0001     |
| crn                                      | 1  | 1.3783   | 0.0893            | 238.4814           | <.0001     |
| race2                                    | 1  | -0.1133  | 0.1006            | 1.2683             | 0.2601     |
| race3                                    | 1  | -0.1005  | 0.1150            | 0.7635             | 0.3822     |

We see a p-value (0.0001) <  $\alpha$  (0.05) thus we reject the null hypothesis ( $\beta_1$  = 0) and say that crn is significantly associated with sta.

. (b) Fit a logistic regression model of sta on age, sex, crn and race.

$$\log{\{\pi/(1-\pi)\}} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{crn} + \beta_4 \text{race2} + \beta_5 \text{race3}$$

What is the estimated odds ratio of death associated with crn for those with age=30, sex=female and race=black? What is the odds ratio of death associated with crn for those with with age=50, sex=male and race=white?

| Analysis of Maximum Likelihood Estimates |    |          |                   |                    |            |  |
|--|----|----------|-------------------|--------------------|------------|--|
| Parameter                                | DF | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSq |  |
| Intercept                                | 1  | -2.5990  | 0.1615            | 258.8445           | <.0001     |  |
| age                                      | 1  | 0.0191   | 0.00218           | 77.0498            | <.0001     |  |
| crn                                      | 1  | 1.3213   | 0.0925            | 204.0479           | <.0001     |  |
| sex                                      | 1  | 0.6333   | 0.0937            | 45.6752            | <.0001     |  |
| race2                                    | 1  | -0.1224  | 0.1034            | 1.3999             | 0.2367     |  |
| race3                                    | 1  | -0.1295  | 0.1182            | 1.2000             | 0.2733     |  |

Since the question is asking for odds ratio of *death associated with crn* and there are no crn interaction terms we are only interested in  $\beta_3$ . Thus, both odds ratios will be the same ->  $\exp(\beta_3) = \exp(1.3213) = 3.748291 = \sim 3.75\%$  chance of death.

. (c) Interpret  $\hat{\beta}_1$  in model (2). Estimate the odds ratio of death for every 10 years increase in age, adjusting for other predictors.

We see our  $\beta_1$  term is 0.0191, therefore the estimated odds ratio associated with a 1 year increase in age is  $\exp(\beta_1) = \exp(0.0191) = 1.019284$ . The estimated odds ratio associated with a 10 year increase in age is  $\exp(10^*\beta_1) = \exp(10^*0.0191) = 1.210459$ .

. (d) Based on model (2), calculate the estimated probability of death for an ICU patient with age=50, sex=female, crn=yes and race=black?

```
Pr(sta=1|age=50, sex=female, crn=yes, race=black) = \exp(\beta_0 + 50*\beta_1 + \beta_2 + \beta_3 + \beta_4) / (1 + \exp(\beta_0 + 50*\beta_1 + \beta_2 + \beta_3 + \beta_4)) = \exp(0.1882) / (1 + \exp(0.1882) = 0.5469116
```

. (e) Fit a logistic regression model

$$\log{\{\pi/(1-\pi)\}} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{crn} + \beta_4 \text{race2} + \beta_5 \text{race3} + \beta_6 \text{crn*sex}.$$

What is the estimated odds ratio of death associated with crn for those with age=30, sex=female and race=black? What is the odds ratio of death associated with crn for those with with age=50, sex=male and race=white?

| Analysis of Maximum Likelihood Estimates |            |          |                   |                    |            |  |
|--|------------|----------|-------------------|--------------------|------------|--|
| Parameter                                | DF         | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSq |  |
| Intercept                                | ntercept 1 | -2.4714  | 0.1702            | 210.8364           | <.0001     |  |
| age                                      | 1          | 0.0193   | 0.00219           | 77.9530            | <.0001     |  |
| sex                                      | 1          | 0.3956   | 0.1437            | 7.5743             | 0.0059     |  |
| crn                                      | 1          | 1.0695   | 0.1478            | 52.3663            | <.0001     |  |
| race2                                    | 1          | -0.1253  | 0.1036            | 1.4605             | 0.2269     |  |
| race3                                    | 1          | -0.1391  | 0.1185            | 1.3783             | 0.2404     |  |
| sex*crn                                  | 1          | 0.4094   | 0.1895            | 4.6661             | 0.0308     |  |

This time around there is an interaction term so we include  $\beta_3$  and  $\beta_6$  when determining the odds ratio of death associated with crn for the 30-year-old female but only  $\beta_3$  for the male.

| Odds Ratio Estimates and Wald Confidence Intervals |              |          |             |             |  |  |
|--|--------------|----------|-------------|-------------|--|--|
| Label  | Odds Ratio   | Estimate | 95% Confide | ence Limits |  |  |
| OR1  | crn at sex=1 | 4.388    | 3.475       | 5.541       |  |  |
| OR2  | crn at sex=0 | 2.914    | 2.181       | 3.893       |  |  |

Female:  $\exp(\beta_3 + \beta_6) = \exp(0.3956 + 0.4094) = 4.388116$ 

Male:  $\exp(\beta_3) = \exp(1.0695) = 2.913922$ 

. (f) (5 points) Construct 95% Wald test based confidence intervals for the two odds ratios in part (e). Do these intervals contain the value 1? What do they tell you?

We can see from the previous SAS printout that the Wald CIs for each are (3.475, 5.541) and (2.181, 3.893) respectively. Since neither contains 1 we conclude that they are significantly different from 1 and thus crn is significantly associated with sta (given the other parameters are held constant).

. (g) Based on model (3), we want to test whether sex is significantly associated with sta among those people with history of chronic renal failure (crn=1). Write out the null and alternative hypotheses.

$$H_0$$
:  $\beta_2 + \beta_6 = 0$  vs  $H_1$ :  $\beta_2 + \beta_6 \neq 0$ 

. (h) Conduct a likelihood ratio test for part (g) using SAS proc genmod procedure. Draw a conclusion under significance level 0.05.

| Contrast Results |    |            |            |      |  |
|------------------|----|------------|------------|------|--|
| Contrast         | DF | Chi-Square | Pr > ChiSq | Туре |  |
| test1            | 1  | 43.46      | <.0001     | LR   |  |

We see our p-value (<0.0001) <  $\alpha$  (0.05) thus we reject the null hypothesis and conclude that the effect of sex on sta is statistically significant amongst patients with crn.

. (i) Based on model (3), obtain  $\exp(\hat{\beta}_3)$ . Does it have a simple interpretation? Explain your answer.

We see 
$$\beta_3 = 1.0695$$
 and  $\exp(\beta_3) = 2.913922$ 

This would only explain the odds of death associated with crn for males as there is an interaction term in the model. It was initially stated that males=0 and females=1, therefore we would need the interaction term included in order to draw conclusions on the odds of females.

. (j) Based on model (3), calculate the estimated probability of death for an ICU patient with age=50, sex=female, crn=yes and race=black?

```
Pr(sta=1|age=50, sex=female, crn=yes, race=black) = \exp(\beta_0 + 50*\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_6) / (1 + \exp(\beta_0 + 50*\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_6)) = \exp(0.2428) / (1 + \exp(0.2428) = 0.5604035)
```