**PH 539: Biostatistical Methods II**

**HOMEWORK 3 (Logistic Regression) [Due April 4, 2017]**

**Dr. Fares Qeadan: fqeadan@salud.unm.edu**

|  |
| --- |
| **PART I (MULTIPLE CHOICE AND SHORT ANSWERS QUESTIONS):** |

**Question 1 [1 points]:**

**Which of the following is a logistic regression model? (*p*=probability of an event)**

a. *p*=**0+**1*x*1+ · · · +**k*x*k c. log(*p*)= **0+**1*x*1+ · · · +**k*x*k

b. *p*/(1–*p*)= **0+**1*x*1+ · · · +**k*x*k d. log(*p*/(1–*p*))= **0+**1*x*1+ · · · +**k*x*k

**Answer:**

**Question 2 [1 points]:**

**Which of the following represents an odds ratio in a logistic regression?**

a. estimate of ** c. exp(**)

b. logit difference (logit=log(odds)) d. exp(logit)

**Answer:**

**Question 3 [1 points]:**

**Which of the following yields the correct estimated odds ratio for a 5-unit increase in an explanatory variable?**

a. 5** c. 5e**

b. e 5** d. e5e**

**Answer:**

**Question 4 [1 points]:**

**The Mantel-Haenszel’s odds ratio is often used to account for confounding. When a confounding variable is found to be an effect modifier, why is it not appropriate to report Mantel-Haenszel odds ratios?**

**Answer:**

**Question 5 [1 points]:**

**List situations when Mantel-Haenszel’s method for calculating an adjusted odds ratio is no longer appropriate. What type of analysis is an alternative in these situations?**

**Answer:**

**Question 6 [1 points]:**

**Briefly explain what a positive and negative parameter estimate (**) implies in a logistic regression in terms of the probability of the outcome.**

**Answer:**

**Question 7 [1 points]:**

**How does the magnitude of the parameter estimate (**) in a logistic regression influence the interpretation of the risk factor?**

**Answer:**

**Question 8 [1 points]:**

**Suppose that the estimated odds ratio of asthma is reported to be 0.45 (95% CI: 0.1, 0.8) for children living near farms (measured in kilometers, km), adjusting for age, race, and sex. Interpret.**

**Answer:**

**Question 9 [1 points]:**

**Given that an interaction effect is found significant in the logistic regression model, the main effects associated with the interaction should NOT be interpreted independently. Why? What should be reported/interpreted when an interaction is present?**

**Answer:**

**Question 10 [1 points]:**

**What is the rationale behind using logit (i.e. log(odds)) and not odds or probability of the event as the outcome variable in logistic regression?**

**Answer:**

**Question 11 [1 points]:**

**Three diagnostic tests (Pearson, Deviance, Hosmer-Lemeshow) can be used to evaluate the model fit of logistic regression. Provide the null hypothesis for a test of model fit. What would you conclude if the hypothesis is not rejected?**

**Answer:**

**Question 12 [1 points]:**

**ROC curves are often used to evaluate model fit for a logistic regression. Why is this appropriate?**

**Answer:**

**Question 13 [1 points]:**

**The plot below presents the odds ratio estimates and 95% confidence intervals from three different studies. Interpret.**



**Answer:**

**Question 14 [1 points]:**

**Which of the following is true?**

1. Binomial logistic regression is analogous to multiple regression.
2. Binomial logistic regression can only be used with scores.
3. Binomial logistic regression is the same as multiple regression.
4. Binomial logistic regression is not at all like multiple regression.

**Answer:**

**Question 15 [1 points]:**

**The logit:**

1. is the natural logarithm of the odds ratio.
2. is a logarithm of a digit.
3. is an instruction to record the data.
4. is the cube root of the sample size.

**Answer:**

**Question 16 [1 points]:**

**Logistic regression is used when you want to:**

1. Predict a dichotomous variable from continuous or dichotomous variables.
2. Predict a continuous variable from dichotomous variables.
3. Predict any categorical variable from several other categorical variables.
4. Predict a continuous variable from dichotomous or continuous variables.

**Answer:**

**Question 17 [1 points]:**

**Logistic regression assumes a:**

1. Linear relationship between continuous predictor variables and the outcome variable.
2. Linear relationship between continuous predictor variables and the logit of the outcome variable.
3. Linear relationship between continuous predictor variables.
4. Linear relationship between observations.

**Answer:**

**Question 18 [1 points]:**

**Which of the following is the link function for logistic regression?**

1. The logistic transformation
2. The logit transformation
3. The exponential transformation

**Answer:**

**Question 19 [1 points]:**

**A logistic regression model was used to assess the association between CVD and obesity. P is defined to be the probability that the people have CVD, obesity was coded as 0=non obese, 1=obese. log(P/(1-P)) = -2 + 0.7(obesity) What is the log odds ratio for CVD in persons who are obese as compared to not obese?**

1. 0.7
2. -2
3. 2.7
4. Exp(0.7)
5. Exp(2)

**Answer:**

**Question 20 [1 points]:**

**The lowest possible OR is 1.00.**

1. TRUE
2. FALSE

**Answer:**

**Question 21 [1 points]:**

**If an odds ratio is computed for a particular independent variable, and if that particular OR takes into consideration the other independent variables, then it is called an \_\_\_\_\_ OR.**

**Answer:**

**Question 22 [1 points]:**

**An OR equal to \_\_\_ would mean that membership in the two categories of the dependent variable is unrelated to the independent variable under consideration.**

**Answer:**

**Question 23 [1 points]:**

**Assuming a continuous explanatory variable on the x-axis and a binary proportion (e.g. probability of having a disease) on the y-axis, which of the two curves would be most likely to reflect the relationship?**



**Answer:**

**Question 24 [1 points]:**

**In statistics, logistic regression is used for prediction of the probability of occurrence of an event by fitting data to a logit function logistic curve.**

1. TRUE
2. FALSE

**Answer:**

**Question 25 [1 points]:**

**From the following output for a multiple logistic regression model in which the dependent variable is binary indicating Coronary artery disease (CAD), one could infer that compared with male sex, female sex is protective against this disease:**



1. TRUE
2. FALSE

**Answer:**

**Question 26 [4 points]:**

**Give an interpretation to the three obtained odds ratios from the previous problem?**

**Answer:**

**PART II (STATA+TEXTBOOK):**

**Question 1 [24 points (8 points for each part)]:**

**4.25 Exercises 1-3 only page 199 using the 4.ex.Sepsis.dta data set.**

**Notes:**

1. **For question 1 you should get the following graph (don’t forget to use the if statement when conducting both the logistic regression and graph):**

****

**Figure xx:** The probability of death in treated black patients with Ibuprofen versus Placebo as a function of baseline Apache score.

1. **For question 3, use the followings to get the 95% C.I. for the expected mortality probability:**



**PART III (STATA+TEXTBOOK):**

**Question 1 [42 points (7 points for each part)]:**

**5.38 Exercises 2-7 only pages 283-284 using the 5.ex.InjuryDeath.dta data set.**

**Notes:**

1. **For question 1, to compute the M-H odds ratios please use the following syntax:**

**cc exposure disease, by(confounder)**

1. **For question 4, a multiplicative model doesn’t mean an interaction model. It only mean including the two factors in the model.**
2. **For question 5, to eliminate ambiguities, I suggest that you create new variables (married and pcare) as follows and use them instead of illegit and pnclate:**



**The new variables are more intuitive and less confusing.**

1. **To get the odds ratios for the table in question 5, please use the lincom command as follows:**

**I will demonstrate the first odds ratio:**

**The used model is:**

**Logit(p)=**0+**1*\*Married*+**2*\*Pcare*+**3*\*Married\*Pcare***

**The model corresponding to the numerator is:**

**Logit(p)=**0+**1*\*Married* +**3*\*Married\*Pcare***

 **where Married=1 and Pcare=0.**

**The model corresponding to the denominator is:**

**Logit(p)=**0+**1*\*Married* +**2*\*Pcare*3*\*Married\*Pcare***

 **where Married=1 and Pcare=1.**

 **So, subtracting the denominator from the numerator will give:**



**So the odds ratio for the first case is 1.835 (95% C.I. 1.22-2.749).**

1. **The deviance is equal to:**

 **deviance=2(Log likelihood(F)-Log likelihood(R))**

**and it’s equal to the LRT test statistic (you get it from running: lrtest mr mf)**

1. **For the desired plot, please use the syntax from :**

[**http://biostat.mc.vanderbilt.edu/dupontwd/wddtext/dov2/5.20.EsophagealCa.do**](http://biostat.mc.vanderbilt.edu/dupontwd/wddtext/dov2/5.20.EsophagealCa.do)

 **YOU ONLY NEED TO CHANGE THE Y AND X LIMITS OF THE FIGURE.**