NAME: Hartaig Singh

UNI: hs2937

Homework 6

P8120 Spring 2017

Due: **Monday, April 3rd, 2017 at 5:30pm**.

1. The most common cause of death from diarrhea in children under age 5 is rotavirus. A recent New England Journal of Medicine article (3/23/17; DOI: 10.1056/NEJMoa1609462) reported results from a clinical trial to assess the impact of a new vaccine on diarrhea caused by rotavirus. The researchers reported that there were only 31 cases of rotavirus related diarrhea among the 1,780 children who got the vaccine, while there were 87 among the 1,728 children who got the placebo.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Vaccine | Placebo |  |
| Rotavirus Diarrhea | 31 | 87 | 118 |
| Other | 1749 | 1641 | 3390 |
|  | 1780 | 1728 | 3508 |

1. Fit a logistic regression model using the data provided above. What is your fitted logistic regression model?

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -2.9371 | 0.1100 | 712.7510 | <.0001 |
| **vaccine** | **vaccine** | 1 | -1.0956 | 0.2120 | 26.7141 | <.0001 |

The estimated logistic regression model obtained from SAS is:

1. Provide an appropriate measure of effect to describe the relationship between vaccination and rotavirus related diarrhea.

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **vaccine vaccine vs placebo** | 0.334 | 0.221 | 0.507 |

Since this is a cross sectional study, we can report and odds ratio to describe the relationship between vaccination and rotavirus related diarrhea.

The odds of having rotavirus related diarrhea among those given the vaccine are **.334** times the odds of having rotavirus related diarrhea among those given the placebo.

1. Provide a 95% CI for your measure of association.

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **vaccine vaccine vs placebo** | 0.334 | 0.221 | 0.507 |

Assumptions: Distribution of the log of the estimated odds ratio is approximately normal, therefore we can use a normal approximation

for the true OR = = (.221,.507)

We are 95% confident that the true odds ratio for describing the relationship between vaccination and rotavirus related diarrhea is between **.221 and .507**

1. Use a LRT to assess whether rate of rotavirus related diarrhea varied significantly between those who received the vaccine and those that did not.

| **Model Fit Statistics** | | | |
| --- | --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** | |
| **Log Likelihood** | **Full Log Likelihood** |
| **AIC** | 1034.525 | 1006.178 | 15.514 |
| **SC** | 1040.688 | 1018.504 | 27.839 |
| **-2 Log L** | 1032.525 | 1002.178 | 11.514 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 30.3468 | 1 | <.0001 |
| **Score** | 29.2526 | 1 | <.0001 |
| **Wald** | 26.7141 | 1 | <.0001 |

= the true log odds ratio describing the association between rotavirus related diarrhea and those who received the vaccine compared to those who did not.

Assumptions: Independent observations, correctly specified the model, sufficiently large sample size.

Under Null Hypothesis

Under Alternative Hypothesis

Use a LRT: under Ho

Test Statistic:

p-value < .0001

Decision: Since p-value < .0001 we reject the null hypothesis

We have sufficient evidence in this data to suggest that rotavirus related diarrhea and vaccination status are associated at the 5% significant level.

1. Provide a brief summary (in no more than two sentences) of your findings.

After fitting a logistic regression model on the data we observed a vaccination odds ratio of .334 indicating that getting vaccinated is protective and can prevent against rotavirus related diarrhea. We can also conclude that vaccination status is significantly associated with having rotavirus related diarrhea.

ods rtf file='HW6.rtf';

**data** RotaVirus;

input vaccine events trials;

cards;

0 87 1728

1 31 1780

;

**run**;

**proc** **format**;

value vaccfmt **0** = 'placebo' **1** = 'vaccine';

**run**;

**proc** **logistic** data = RotaVirus;

class vaccine(ref = 'placebo')/param = ref;

model events/trials = vaccine/cl;

format vaccine vaccfmt.;

**run**;

ods rtf close;

1. A survey of high school students was done to examine whether students had ever driven a car after consuming a substantial amount of alcohol (1=yes, 0=no). Data was collected on their sex (male/female), race (black/white), and grade level (9, 10, 11, 12). Researchers realized that the impact of race on consuming alcohol before driving might vary by grade level and decided to fit the following model:

|  |  |
| --- | --- |
| Variable | Coding = 1 if |
| Intercept  Sex (*X*1) | Female |
| Race (*X*2) | Black |
| Grade Level (*X*3- *X*5) | 9th grade  10th grade  11th grade |

1. What is the log odds of driving a car after drinking for a white female in the 9th grade?

(

The log odds of driving a car after drinking for a white female in 9th grade is **-3.5**

1. What is the log odds of driving a car after drinking for a black female in the 9th grade?

(

The log odds of driving a car after drinking for a black female in the 9th grade is **-3.48**

1. What do you notice about the difference (b)-(a)?

-3.5-(-3.48) = .02

There is not much of a difference between the log odds of driving a car after drinking for a white female in the 9th grade and the log odds of driving a car after drinking for a black female in the 9th grade.

1. What is the log odds of driving a car after drinking for a white male in the 9th grade?

(

The log odds of driving a car after drinking for a white male in the 9th grade is **-3.10**

1. What is the log odds of driving a car after drinking for a black male in the 9th grade?

(

The log odds of driving a car after drinking for a black male in the 9th grade is **-3.08**

1. What do you notice about the difference (e)-(d)? How does this compare to your answer in part (c)?

-3.10-(-3.08) = .02

There is not much of a difference between the log odds of driving a car after drinking for a white male in the 9th grade and the log odds of driving a car after drinking for a black male in the 9th grade. This answer is exactly the same as in part c.

1. Compute the OR of drinking before driving for blacks versus non-blacks in the 9th grade, adjusting for gender.

log odds for blacks – log odds for whites

The OR of drinking before driving for blacks versus non-blacks in the 9th grade adjusting for gender is **1.02**

1. Compute the OR of drinking before driving for blacks versus non-blacks in the 10th grade, adjusting for gender.

log odds for blacks – log odds for whites

The OR of drinking before driving for blacks versus non-blacks in the 10th grade adjusting for gender is **.711**

1. Compute the OR of drinking before driving for blacks versus non-blacks in the 11th grade, adjusting for gender.

log odds for blacks – log odds for whites

The OR of drinking before driving for blacks versus non-blacks in the 11th grade adjusting for gender is **.492**

1. Compute the OR of drinking before driving for blacks versus non-blacks in the 12th grade, adjusting for gender.

log odds for blacks – log odds for whites

The OR of drinking before driving for blacks versus non-blacks in the 12th grade adjusting for gender is **.487**

1. Suppose researchers decide to switch their main exposure to grade level and their effect modifier to race. Compute appropriate measures of association to describe the relationship between grade level and drinking before driving, stratified by race and adjusting for gender.

OR for driving after drinking comparing 9th grade to 12th grade among blacks adjusting gender

log odds for 9th grade – log odds for 12th grade

The OR of drinking before driving for 9th graders versus 12th graders among blacks adjusting for gender is **.228**

OR for driving after drinking comparing 10th grade to 12th grade among blacks adjusting gender

log odds for 10th grade – log odds for 12th grade

The OR of drinking before driving for 10th graders versus 12th graders among blacks adjusting for gender is **.35**

OR for driving after drinking comparing 11th grade to 12th grade among blacks adjusting gender

log odds for 11th grade – log odds for 12th grade

The OR of drinking before driving for 11th graders versus 12th graders among blacks adjusting for gender is **.566**

OR for driving after drinking comparing 9th grade to 12th grade among whites adjusting gender

log odds for 9th grade – log odds for 12th grade

The OR of drinking before driving for 9th graders versus 12th graders among whites adjusting for gender is **.109**

OR for driving after drinking comparing 10th grade to 12th grade among whites adjusting gender

log odds for 10th grade – log odds for 12th grade

The OR of drinking before driving for 10th graders versus 12th graders among whites adjusting for gender is **.239**

OR for driving after drinking comparing 11th grade to 12th grade among whites adjusting gender

log odds for 11th grade – log odds for 12th grade

The OR of drinking before driving for 11th graders versus 12th graders among whites adjusting for gender is **.56**

1. Suppose researchers would like to test whether race modifies the association between grade level and drinking before driving.
   1. What would be their null and alternative hypotheses?

at least 1 does equal 0

* 1. If they were to use a Wald Test, where on the SAS output (give table name) should they look for their test statistic? What distribution (with degrees of freedom) would it follow?

Table: Type 3 Analysis of Effect

Distribution: Chi-squared with 3 degrees of freedom

* 1. If they were to use a LRT, what reduced model should they fit? What distribution (with degrees of freedom) would it follow?

Reduced Model:

Distribution: Chi-squared with df = (# parameters under ) – (# parameters under = 8 – 5 = **3**

* 1. Suppose the two tests are in disagreement (one says to reject, the other fail to reject). Which test should they report and why?

The LRT since it is more reliable and more powerful than the Wald Test.

1. Investigators at the University of Massachusetts Aids Research Unit (UMARU) conducted a study to compare treatment programs of different planned durations (3 months vs. 6 months) designed to reduce drug abuse and to prevent high-risk HIV behavior. There were two treatment program sites, A and B. The dichotomous outcome of interest is whether or not the subject remained drug free at the end of the study. Below is a list of the variables in the data set “UMARUdata” which is available in Courseworks:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Name** | **Code** |
| Age | AGE | Years |
| Race | RACE | 0 = White 1 = Other |
| Treatment | TREAT | 0 = Short 1 = Long |
| Treatment Site | SITE | 0 = A 1 = B |
| Returned to Drug Use Prior to Scheduled End of Program | DFREE | 0 = Relapsed 1 = Remained drug free |

**Part 1:** Researchers wish to explore the relationship between remaining drug free and **age, race, and treatment site**.

ods rtf file='HW5(4).rtf';

**proc** **import** datafile = 'C:\Users\PubLibrary\Downloads\UMARUdata.csv'

out = UMARUdata

dbms = csv;

**run**;

**proc** **logistic** data = UMARUdata descending;

class TREAT (ref = '0') / param = ref;

class SITE (ref = '0') / param = ref;

model DFREE = SITE AGE / cl;

**run**;

ods rtf close;

1. Suppose race is associated with both site and remaining drug free. Assess whether race confounds the relationship between site and remaining drug free, controlling for age.

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -1.7654 | 0.5177 | 11.6293 | 0.0006 |
| **SITE** | **1** | 1 | 0.2718 | 0.2038 | 1.7787 | 0.1823 |
| **AGE** |  | 1 | 0.0188 | 0.0153 | 1.4927 | 0.2218 |

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -1.9006 | 0.5250 | 13.1060 | 0.0003 |
| **RACE** | **1** | 1 | 0.4850 | 0.2127 | 5.1999 | 0.0226 |
| **SITE** | **1** | 1 | 0.3115 | 0.2057 | 2.2933 | 0.1299 |
| **AGE** |  | 1 | 0.0185 | 0.0155 | 1.4263 | 0.2324 |

Percent change =

Since the percent difference between the crude estimate for site and the adjusted estimate for site is **12.74** which is greater than 10, we conclude that race is confounding the association between remaining drug free and site.

1. Briefly explain which variables are significant predictors of remaining drug free (no need to carry out the seven steps of hypothesis testing).

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -2.1678 | 0.5412 | 16.0463 | <.0001 |
| **RACE** | **1** | 1 | 0.4502 | 0.2141 | 4.4230 | 0.0355 |
| **TREAT** | **1** | 1 | 0.4271 | 0.1951 | 4.7907 | 0.0286 |
| **SITE** | **1** | 1 | 0.3195 | 0.2067 | 2.3901 | 0.1221 |
| **AGE** |  | 1 | 0.0200 | 0.0155 | 1.6674 | 0.1966 |

The variables that are significant are race and treatment since they have p-values of .0355 and .0286 respectively each of which are less than .05. Variables site and age are not significant because they have p-values greater .05 of .1221 and .1966 respectively.

**Part 2:** Researchers suddenly realize that the relationship between race and remaining drug free might be dependent on which site the participant was enrolled at, holding age constant.

1. Perform a LRT to assess the researchers’ hypothesis.

Under Null

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 655.729 | 653.499 |
| **SC** | 660.083 | 670.916 |
| **-2 Log L** | 653.729 | 645.499 |

Under Alternative

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 655.729 | 648.958 |
| **SC** | 660.083 | 670.729 |
| **-2 Log L** | 653.729 | 638.958 |

= the true log odds ratio describing the interaction between race and site

Assumptions: Independent observations, correctly specified the mode, sufficiently large sample size

Df = # of predictors under alternative - # number of predictors under null = 4 – 3 = 1

This follows a chi-squared distribution with 1 degree of freedom

Use LRT:

p-value = = .011

Since this value is less than .05 we reject the null hypothesis

At a 5% level of significance the observed provides enough evidence to claim that site significantly modifies the relationship between remaining drug free and race holding age constant.

1. Suppose that from past research it is already known that the relationship between race and remaining drug free is dependent on site. For this reason, the researchers plan to report site specific ORs adjusted for age regardless of the results in (b). Provide them with these estimates and interpret.

ods rtf file='HW5(15).rtf';

**proc** **import** datafile = 'C:\Users\PubLibrary\Downloads\UMARUdata.csv'

out = UMARUdata

dbms = csv;

**run**;

**proc** **logistic** data = UMARUdata descending;

class RACE (ref = '0') / param = ref;

class TREAT (ref = '0') / param = ref;

class SITE (ref = '0') / param = ref;

model DFREE = SITE AGE RACE SITE\*RACE / cl;

oddsratio RACE /diff = ref;

**run**;

ods rtf close;

| **Odds Ratio Estimates and Wald Confidence Intervals** | | | |
| --- | --- | --- | --- |
| **Odds Ratio** | **Estimate** | **95% Confidence Limits** | |
| **RACE 1 vs 0 at SITE=0** | 2.269 | 1.394 | 3.695 |
| **RACE 1 vs 0 at SITE=1** | 0.644 | 0.270 | 1.538 |

The odds of remaining drug free for other races is **2.269** times the odds of remaining drug free for whites at site 0 adjusting for age.

The odds of remaining drug free for other races is **.644** times the odds of remaining drug free for whites at site 1 adjusting for age.

1. Provide and interpret 95% CIs for the site specific ORs adjusted for age.

| **Odds Ratio Estimates and Wald Confidence Intervals** | | | |
| --- | --- | --- | --- |
| **Odds Ratio** | **Estimate** | **95% Confidence Limits** | |
| **RACE 1 vs 0 at SITE=0** | 2.269 | 1.394 | 3.695 |
| **RACE 1 vs 0 at SITE=1** | 0.644 | 0.270 | 1.538 |

We are 95% confident that the true odds ratio adjusting for age for race is between **1.394** and **3.695** at site 0.

We are 95% confident that the true odds ratio adjusting for age for race is between **.270** and **1.538** at site 1.

1. Based on your confidence intervals in (f), what would you conclude about the association between race and remaining drug free?

Since 1 is not included in the confidence interval for race at site 0, there is a significant association between race and remaining drug free at site 0.

Since 1 is included in the confidence interval for race at site 1, there is not a significant association between race and remaining drug free at site 1.

1. Which set of results (Part 1 or Part 2) should the researchers present in their manuscript? Why?

The researchers should present both stratum specific odds ratios in their manuscript since presence of interaction can actually lead to greater insight and a more accurate reflection of the observed data.

1. Summarize your findings to **Part 2** in a brief paragraph (no more than 4 sentences).

We investigated the impact of site on the relationship between race and remaining drug free using the LRT which showed significant results indicating site does modify the association between race and remaining drug free at a p-value of .011. Additionally, the site specific odds ratios showed significance at site 0 but not at site 1 with odds of 2.269 and .644 respectively.