**BIS 505b – Problem Set 4 Due 11:59 PM Eastern, 4/11/2021**

**Instructions:**  Follow the homework instructions outlined in the syllabus. Round your answers to 2 decimal places. Perform all tests at the = 0.05-level and follow the steps of hypothesis testing.

**Assignment**

**Question 1:** This question analyzes data from a study conducted to look at the association between high-risk occupation (Yes/No) and presence of bladder cancer (Yes/No). The study reported the following estimates:

* The estimated risk of bladder cancer (Disease) in those with a high-risk occupation (Exposed) = = 0.5983
* The estimated risk of bladder cancer (Disease) in those without a high-risk occupation (Unexposed) = = 0.4329

1. [12] Report the odds of disease (presence of bladder cancer) in the exposed (those with a high-risk occupation). Report the odds of disease (presence of bladder cancer) in the unexposed (those without a high-risk occupation). Report the odds ratio of bladder cancer in those with a high-risk occupation vs. those without a high-risk occupation.

The odds of disease (presence of bladder cancer) in the exposed (those with a high-risk occupation) is 1.49.

The odds of disease (presence of bladder cancer) in the unexposed (those without a high-risk occupation) is 0.76.

The odds ratio of bladder cancer in those with a high-risk occupation vs. those without a high-risk occupation is 1.96.

1. [10] Suppose we were to fit a simple logistic regression model to these data, modeling bladder cancer status where bladder cancer=Yes is the event of interest and high-risk occupation is the only independent variable:

= High risk occupation (0 = Non-high-risk occupation (reference), 1 = High-risk occupation)

This model does not control for any potential confounders. Report the fitted simple logistic regression model. Use your knowledge of what the intercept and slope represent in a logistic regression model along with your unadjusted results from part **(a)** to estimate the intercept and slope parameters of your fitted logistic regression model.

The fitted simple logistic regression model:

**Question 2:** In a study investigating maternal risk factors for congenital syphilis, syphilis is treated as a dichotomous response variable, where 1 represents the presence of disease in a newborn and 0 its absence. The estimated coefficients from a logistic regression model containing the explanatory variables (4 categorical, 1 continuous): cocaine or crack use during pregnancy (1=yes, 0=no (ref)), marital status (1=unmarried, 0=married (ref)), number of prenatal visits to a doctor, alcohol use during pregnancy (1=yes, 0=no (ref)), and level of education (1=less than high school, 0=high school or more (ref)) are listed in the table below.

|  |  |
| --- | --- |
|  | Estimate |
| **Intercept** | 0.080 |
| **Cocaine/Crack Use** | 1.354 |
| **Marital Status** | 0.779 |
| **Number of Prenatal Visits** | -0.098 |
| **Alcohol Use** | 0.723 |
| **Level of Education** | 0.298 |

1. [5] As an expectant mother’s number of prenatal visits to the doctor increases, does the **probability** that her child will be born with congenital syphilis increase or decrease? Explain.

The probability that her child will be born with congenital syphilis will decrease. Because the slope of Number of Prenatal Visits is -0.098 < 0, indicates that larger values of are related to a smaller log-odds, when log-odds decrease, the probability also decreases.

1. [7] Marital status is a dichotomous variable, where the value 1 indicates that a woman is unmarried and 0 indicates that she is married. What are the adjusted relative odds that a newborn will suffer from syphilis for unmarried versus married mothers? Interpret this odds ratio.

The adjusted relative odds that a newborn will suffer from syphilis for unmarried versus married mothers is 2.18, which means that the odds of a newborn will suffer from syphilis for unmarried mother is 118% higher compared to married mother when controlling for cocaine/crack use, number of prenatal visits, alcohol use, and level of education.

1. [7] Cocaine or crack use is also a dichotomous variable; the value 1 indicates that a woman used these drugs during pregnancy and 0 indicates that she did not. What is the estimated adjusted odds ratio that a child will be born with congenital syphilis for women who used cocaine or crack versus those who did not? Interpret this odds ratio.

The adjusted relative odds that a newborn will suffer from syphilis for women who used cocaine or crack versus those who did not is 3.87, which means that the odds of a newborn will suffer from syphilis for women who used cocaine or crack is 287% higher compared to those who did not when controlling for marital status, number of prenatal visits, alcohol use, and level of education.

1. [10] The estimated coefficient of cocaine or crack use has standard error = 0.162. Construct a 95% confidence interval for the population adjusted odds ratio of cocaine/crack use.

95% CI for

95% CI for

The 95% confidence interval for the population adjusted odds ratio of cocaine/crack use is

1. [5] Estimate the probability that a newborn will be born with congenital syphilis if the child’s mother is unmarried, has less than a high school level of education, uses cocaine and alcohol during pregnancy, and has 2 prenatal visits to the doctor. Assume these values are in the range of the data used to fit the model.

The fitted multiple logistic regression model is:

Probability that a newborn will be born with congenital syphilis is

Plug in

The estimated probability that a newborn will be born with congenital syphilis is 0.98 with the stated conditions.

1. [5] Estimate the probability that a newborn will be born with congenital syphilis if the child’s mother is married, has greater than a high school education, does not use crack/cocaine during pregnancy, does not consume alcohol during pregnancy, and has 14 prenatal visits to the doctor. Assume these values are in the range of the data used to fit the model.

Plug in to the same model as question e

The estimated probability that a newborn will be born with congenital syphilis is 0.22 with the stated conditions.

**Question 3:** Suppose you are interested in studying intravenous (IV) drug use among high school students in the United States. Drug use is characterized as a dichotomous variable, where 1 indicates that an individual has injected drugs within the past year and 0 that he/she has not. (4 binary + 1 continuous) Factors that might be related to drug use are: instruction about the human immunodeficiency virus (HIV) in school (where 1 indicates that instruction was received and 0 indicates that instruction was not received (ref)), age of the student (years), sex (1=male, 0=female (ref)), and general knowledge about HIV, including the various modes of transmission and ways to reduce risk (1=possesses good general knowledge, 0=does not (ref)). The estimated coefficients and standard errors from a logistic regression model containing each of these explanatory variables as well as the interaction between HIV instruction and sex are displayed in the table below.

|  |  |  |
| --- | --- | --- |
|  | Estimate | Standard Error |
| **Intercept** | -1.183 | 0.859 |
| **HIV INSTRUCTION** | 0.039 | 0.421 |
| **AGE** | -0.164 | 0.092 |
| **SEX** | 1.212 | 0.423 |
| **HIV KNOWLEDGE** | -0.187 | 0.048 |
| **HIV INSTRUCTION\*SEX** | -0.663 | 0.301 |

1. [10] Determine if the effect of HIV instruction on IV drug use is significantly different in males vs. females.

(1) State the null and alternative hypotheses

vs.

(2) Specify the significance level, α = 0.05

(3) Compute the test statistic

(4) Generate the decision rule

Given α = 0.05,

Reject if or if

(5) Draw a statistical conclusion and state the conclusion in words in the context of the problem.

Conclusion: There is evidence to reject and conclude that the effect of HIV instruction on IV drug use is significantly different in males vs. females.

1. [8] Write the equation of the fitted logit (log odds) of IV drug use for males. Write the equation of the fitted logit (log odds) of IV drug use for females.

The fitted logistic model is:

The fitted logit (log odds) of IV drug use for males

The fitted logit (log odds) of IV drug use for females

1. [10] Report the adjusted OR for the effect of HIV instruction on IV drug use in males. Report the adjusted OR for the effect of HIV instruction on IV drug use in females. In the model above, is HIV instruction significantly associated with IV drug use in females?

The adjusted OR for the effect of HIV instruction on IV drug use in males.

The adjusted OR for the effect of HIV instruction on IV drug use in females.

Hypothesis test of if HIV instruction significantly associated with IV drug use in females:

(1) State the null and alternative hypotheses

vs.

(2) Specify the significance level, α = 0.05

(3) Compute the test statistic

(4) Generate the decision rule

Given α = 0.05,

Reject if or if

(5) Draw a statistical conclusion and state the conclusion in words in the context of the problem.

Conclusion: There is evidence to fail to reject and conclude that HIV instruction is not significantly associated with IV drug use in females

1. [5] Controlling for all the other variables included in the model, as a student becomes older, does the probability that he or she has used intravenous drugs in the past year increase or decrease? Explain.

The probability that he or she has used intravenous drugs in the past year will decrease because the estimated slope of age is -0.164 <0, indicates that larger values of are related to a smaller log-odds, when log-odds decrease, the probability also decreases.

1. [6] Using the model reported above, what is the estimated probability of IV drug use for a 13-year old male who did not receive HIV instruction and who does not possess good general knowledge about HIV? Assume these values are in the range of the data used to fit the model.

The fitted logistic model is:

Probability of IV drug use is

Plug in

The estimated probability of IV drug use for a 13-year old male who did not receive HIV instruction and who does not possess good general knowledge about HIV is 0.11.