Due 11/27 at 11:59 PM ET

Assignment 6 (65 pts)

1. Many studies have examined the association between maternal infection with rubella during pregnancy (exposure) and congenital cataracts (outcome) in newborn babies. The prevalence of congenital cataracts is low (~4 per 10,000 live births).
   1. What sampling plan (cross-sectional, prospective, retrospective) would be most appropriate for a study like this and why? (5 pts)

Retrospective study sampling would be most appropriate.

I choose retrospective study sampling because, the outcome (congenital cataract) of the study is very rare i.e., we found only 4 cases among 10,000 live births of rubella infected mothers.

* 1. Based on you answer in part a), what measure(s) of association could you calculate? (5 pts)

As I have selected retrospective sampling plan, I will choose **Odds ratio.**

**OR=**

1. See the below excerpt from a study comparing survival after lung cancer treatment by race. (*Williams CD et al. Racial Differences in Treatment and Survival among Veterans and Non-Veterans with Stage I NSCLC. An Evaluation of Veteran’s Affairs and SEER Medicare Populations. Cancer Epidemiology Biomarkers and Prevention. 2020. 29(1): 112-118).*  The far-right column shows the p-value for a chi-squared test assessing the association between race and the listed variables.

A picture containing text, receipt

Description automatically generated

* 1. What would be your conclusion about whether a relationship exists between race and histology? (5 pts)

P-value for Race and Histology is <0.01 our P-value is <0.05

P-value = <0.01 (<0.05)

We reject the null hypothesis. We have sufficient evidence to suggest that there is association between survival after lung cancer treatment and race.

* 1. Squamous cell histology is known to be associated with poorer survival. What would be the concern if histology were not accounted for in the analysis? (5 pts)

If histology were not accounted for it will threaten internal validity of the study.

1. Use the *silicosis.csv* dataset to answer the following questions. A cross sectional study was conducted to assess factors associated with silicosis; a lung disease caused by inhaling large amounts of silica dust.

Data Dictionary:

Everdrill: whether the participant ever worked in drilling/mining (1=Yes, 2=No)

Yrsdrill: years of drilling (0=None, 1=1-10, 2= greater than 10)

Smok\_hist: Smoking History (1=Current/past smoker, 2=Never Smoker)

Agecat: Age (1= less than 45, 2=45+)

Silicosis: whether the participant has silicosis (1=Yes, 2=No)

1. Table

   Description automatically generatedRun a chi-squared test to assess the association between everdrill (ever worked in drilling/mining) (exposure) and silicosis (outcome). Write out your null and alternative hypotheses, the p-value, your decision and conclusion (10 pts)

Text

Description automatically generated

H0: Silicosis and everdrill are **independent**

HA: Silicosis and everdrill are **associated**

P-value = 0.0046 (<0.05)

**Decision/Conclusion**

We reject the null hypothesis. We have sufficient evidence to suggest that there is an association between silicosis and everdrill among silicosis.cvs patients

1. Report and interpret the risk ratio and 95% CI for silicosis in those with mining experience, compared to those without (5 pts)

Table

Description automatically generated

Graphical user interface, text

Description automatically generated

**Risk ratio = 2.9006, 95% CI = 1.3443 – 6.2589**

RR>1: Positive Association

* Risk of silicosis is higher in those with mining experience than without mining experience. Mining exposure increases the incidence of the silicosis
* Risk of silicosis is 2.9006 times higher in the exposed vs. the un-exposed
* We are 95% confident the true RR falls between 1.3443 and 6.2589

1. Report and interpret the odds ratio and 95% CI for silicosis in those with mining experience and those without (5 pts).

**Odds ratio = 3.0690, 95% CI = 1.3637 – 6.9071**

OR>1: Positive Association

* Odds of developing silicosis is higher in the people with mining experience than without mining experience. Mining exposure increases the incidence of silicosis
* Odds of silicosis in people exposed to mining is 3.0690 times higher than in people unexposed to mining
* We are 95% confidence that the true OR falls between 1.3637 and 6.9071

1. Based only on the confidence interval for the OR (not using the p-value from the chi-squared test above), would you conclude that there is an association between mining experience and silicosis? Why or why not? (5 pts)

**Odds ratio: 95% CI = 1.3637 – 6.9071**

With the 95% CI of odds ratio, I would conclude that there is an association between mining experience and silicosis.

As odds ratio 95% CI does not contain 1, it indicates there is a strong positive relationship between exposure (mining experience) and outcome (disease: silicosis)

1. We decide to run a logistic regression to assess the relationship between mining and silicosis. However, we think that the number of years mining may give us more insight, so we use the yrsdrill variable, instead of everdrill. (We are modeling silicosis=1).
   1. Based on the output below, report and interpret overall p-value for the yrsdrill variable and the odds ratios for >10 years, compared to 0 years (5 pts).

A picture containing text

Description automatically generated

Table

Description automatically generated

**Yrs drill 2 (greater than 10yrs)**

Β1= 1.9694,

P-value=0.0013

Odds Ratio (95% CI): 7.167 (2.473 – 20.767)

**Interpretation**

The change in log-odds for yrsdrill is significantly different than 0 (p=0.0013, P<0.05). We reject the null hypothesis. We have sufficient evidence to suggest that there is a strong association between people exposed to mining for greater than 10 years and silicosis among silicosis.cvs patients. 95% CI (2.473 – 20.767) does not contain one which suggests a strong positive relationship between increased years of mining (>10yrs) and silicosis. Odds of silicosis in people exposed to mining for greater than 10yrs is 7.167 times higher than in people exposed to mining for less than 10yrs. We are 95% confident the true odds ratio is between 2.473 and 20.767.

* 1. To better explain the occurrence of silicosis, we try including age and smoking status in the model. Based on the below comparison of models, which model would you select and why? Identify the overall p-values for each variable, and interpret the odds ratios and 95% confidence intervals for each variable (For yrsdrill, only interpret thes comparison between >10 years and 0 years). (15 pts).

Model 1: yrsdrill+agecat (AIC: 196.621)

Model2: yrsdrill+agecat+smok\_hist (AIC: 196.172)

Table

Description automatically generatedTable

Description automatically generatedTable

Description automatically generated

I would select model 1

Model 1 is appropriate because, in model-2 the overall P value for smoking is 0.1407 (P>0.05). For model-2 for smok\_hist variable, we do not reject null hypothesis and it indicates that there is no association between smoking (exposure) and the silicosis (outcome). Hence, I selected model-1 is the best fit for this study.

Model-1

**Variable: yrsdrill 2**

Β1=1.9922

P-value=0.0014

Odds Ratio (95% CI): 7.332 (2.497 – 21.528)

**Interpretation**

The change in log-odds for yrsdrill-2 is significantly different than 0 (p=0.0014, P<0.05). We reject the null hypothesis. We have sufficient evidence to suggest that there is a strong association between people exposed to mining for greater than 10 years and silicosis among silicosis.cvs patients. 95% CI (2.497 – 21.528) does not contain one which suggests a strong positive relationship between increased years of mining (>10yrs) and silicosis. Odds of silicosis in people exposed to mining for greater than 10yrs is 7.332 times higher than in people exposed to mining for less than 10yrs. We are 95% confident the true odds ratio is between 2.497 and 21.528

**Variable: agecat**

Β1=0.9568

P-value=0.0385

Odds Ratio (95% CI): 2.603 (1.052 – 6.442)

**Interpretation**

The change in log-odds for agecat is significantly different than 0 (p=0.0385, P<0.05). We reject the null hypothesis. We have sufficient evidence to suggest that there is a strong association between agecat (exposure) and silicosis (outcome) among silicosis.cvs patients. 95% CI (1.052 – 6.442) does not contain one which suggests a strong positive relationship between agecat and silicosis. Odds of silicosis in people aged less than 45 yrs is 2.603 times higher than in people more than 45yrs. We are 95% confident the true odds ratio is between 1.052 – 6.442.