

Homework 5

Due Monday 11/8

Turn in R code and output – try to use R markdown.

Ndeah Terry, Victoria Trapani, and Pedro Vasquez-Perez will lead HW discussion.

1. (60 points) *Comparing COVID-10 Transmission Levels by Poverty Level*

The following table shows high COVID-19 transmission,

$$High\ COVID = I \left[100,000 \times \left(\frac{covid_count_sep17 - covid_count_sep10}{Tot_Population_ACS_14_18} \right) \geq 100 \right],$$

according to whether the county is high poverty

$$High\ Poverty = I[pct_Prs_Blw_Pov_Lev_ACS_14_18 > 20]$$

using the county-level COVID-19 and ACS data from previous HWs.

<i>High Poverty</i>	<i>High COVID</i>	
	Yes	No
No	2297	176
Yes	640	19

- (a) Find all joint and marginal sample probabilities for the 2x2 contingency table.
- (b) Estimate the sample conditional probabilities of *High COVID* given *High Poverty*.
- (c) Find and interpret the difference of proportions. Find and interpret 95% Wald confidence interval for the difference of proportions.
- (d) Find and interpret the relative risk. Find and interpret 95% Wald confidence interval for the relative risk.
- (e) Find and interpret the odds ratio. Find and interpret 95% Wald confidence interval for the odds ratio.
- (f) Is there an association between *High COVID* and *High Poverty*? Justify your answer.

Now consider all four categories of transmission risk, *COVID Transmission*, as in Q2 of HW4, which yields the following 2x4 table

<i>High Poverty</i>	<i>COVID Transmission</i>			
	Low	Moderate	Substantial	High
No	124	11	40	2297
Yes	4	3	12	640

- (g) Find all joint and marginal sample probabilities for the 2x4 contingency table.
 - (h) Calculate the expected counts for each cell, $\hat{\mu}_{ij}$, in the 2x4 contingency table under the assumption of independence.
 - (i) Carry out the Pearson χ^2 test for independence between *COVID Transmission* and *High Poverty*. Make sure to report the null and alternative hypothesis, test statistic, critical value and conclusion. Calculate the test statistic “by hand” in R using the formula in the lecture notes, but you may check your work with the `chisq.test` function.
 - (j) Carry out the likelihood ratio test for independence between *COVID Transmission* and *High Poverty*. Make sure to report the null and alternative hypothesis, test statistic, critical value and conclusion. Calculate the test statistic “by hand” in R using the formula in the lecture notes, but you may check your work with the `GTest` function.
 - (k) Use Pearson standardized residuals to describe the evidence of association.
 - (l) Carry out an analysis of partitioning the χ^2 to describe patterns of association. Comment on and explain any interesting findings.
2. (40 points) *Assessing Association Between COVID-10 Transmission Level, COVID-19 Vaccination Level and Region*

The following table shows *High COVID* (as defined in Q1), region (defined by state as in https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf), and high COVID-19 vaccination rate

$$High\ Vacc = I[Series_Complete_Pop_Pct \geq 50],$$

where *Series.Complete.Pop.Pct* is the percent of the total county population that is fully vaccinated as of 9/10/21 according to <https://covid.cdc.gov/covid-data-tracker/#county-view>. Note that counties with missing vaccination data are excluded from the table.

<i>Region</i>	<i>High Vacc Yes</i>		<i>High Vacc No</i>	
	<i>High COVID</i>		<i>High COVID</i>	
	Yes	No	Yes	No
Midwest	166	5	771	113
Northeast	148	3	66	1
South	90	3	1054	14
West	125	18	271	17

For this question, use the long form of this data posted on Blackboard: `three_way_COVID_table.csv`.

- Calculate the odds ratio for each of the four partial 2x2 tables of *High Vacc* vs. *High COVID* defined by the four regions. Does there appear to be an association between vaccination and COVID transmission level for each of the regions? Provide an interpretation of the odds ratio for the Midwest 2x2 table and the West 2x2 table.
- Fit the saturated loglinear model for the three-way table. Why is the residual deviance zero? Why are all of the Pearson residuals zero?
- Now fit the **joint independence model assuming region is jointly independent of vaccination and COVID transmission level (i.e. only include the *High Vacc/High COVID* interaction)** and test its goodness-of-fit. Make sure to report the null and alternative hypothesis, test statistic, critical value and conclusion.
- What is the reference group in your loglinear model? State and interpret the coefficient on the *High Vacc* and *High COVID* interaction from the **joint independence model in part (c)**.
- Does the association between vaccination rate and COVID transmission level differ by region? **In other words, do differences in COVID-19 transmission levels for low and high vaccination areas differ by region?** Support your answer with the appropriate deviance test. Make sure to report the null and alternative hypothesis, test statistic, critical value and conclusion.
- Fit all possible loglinear models for the three categorical variables. Find the goodness-of-fit deviance test statistic, associated degrees of freedom and p-value for each of the possible loglinear models. Present the results in a table. *Hint: See Table 7.7 in ICDA for an example.*
- Which model would you choose based on the goodness-of-fit results from part (f)? Describe the type of association (e.g. mutual independence, conditional independence, etc.) that your chosen model implies.