HSCI 50 LAB 5: Bivariate Analysis for Categorical Variables

INSTRUCTIONS

The data set that we will be using for the entire course is resampled data from the MIT COVID-19 Beliefs, Behaviors & Norms Survey (https://covidsurvey.mit.edu/api.html). This is a multi-country, online survey that examined different COVID-19 perceptions across time, from July 6, 2020 to March 28, 2021. We will be using data from the Philippines aged 20 to 60.

In this lab, you will learn how to conduct the *z*-test for proportions and chi-square test of independent association of categorical variables.

You have many options to submit this worksheet. Either you work on this by hand and scan/take a clear photo of your submission and save as PDF, or type your responses in a Word processor or PowerPoint presentation. You do not have to copy the questions again, but please number them accordingly.

Let's revisit a cross tabulation we had used in Lab 2 on COVID-19 vaccine acceptance and age group

Table 1: COVID-19 vaccine acceptance in the Philippines by age, March 14-28, 2021

Accept COVID-19 vaccine	Aged 20-30	Aged 31-40	Aged 41-50	Aged 51-60
Yes	179	166	123	110
No	83	63	39	26
Don't Know	157	89	44	32
Vaccinated	18	5	10	1

PART A. *Z*-test for proportions

For this part, we will use a simplified table from Table 1. We will recode the vaccine acceptance variable into a binary category similar to Lab 3 called vaccine hesitancy, where those who do not accept the vaccine or are unsure about it will be considered vaccine hesitant, otherwise they are not vaccine hesitant. We will also recode the age group variable into younger adults, defined in this lab as aged 20 to 40, and older adults, defined in this lab as aged 41 to 60. The table below summarizes that information.

Table 2: COVID-19 vaccine hesitancy in the Philippines by age, March 14-28, 2021

Vaccine hesitant	Younger adult	Older adult
No	368	244
Yes	392	141

1. Calculate the conditional probabilities of vaccine hesitancy by age, i.e. *P*(vaccine hesitant | younger adult) and *P*(vaccine hesitant | older adult)

2. To test for the differences in these two proportions, how would you phrase the null hypothesis for a z-test of two proportions? How about the alternative hypothesis?

3. Calculate by hand the test statistic for this test

4. Calculate the corresponding *p*-value from the test statistic calculated in Question 3 Using the University of Iowa Normal Distribution Applet (https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html)

Note: Matching the variables in the applet and the variables in our lab worksheet: $\mu = p_0$, $\sigma = \mathrm{SE}\ x = z$. Select the 2-sided *p*-value option: $2P(X > |x|)$

5. What do you conclude?

6. Calculate the 95% confidence interval of the difference of the two proportions. Do you arrive at the same conclusions as the hypothesis test? How?

PART B. Chi-square test of independence of associations

1. We will first use the same 2x2 table from Part A to practice calculating expected counts by hand. Calculate the expected counts of each of the cells of Table 2.

2. What is the null hypothesis appropriate for a chi-square test for independence of association? How about the alternative hypothesis?

3. Calculate the chi-squared statistic by hand

4. Calculate the corresponding *p*-value using the University of Iowa chi-square distribution applet (https://homepage.divms.uiowa.edu/~mbognar/applets/chisq.html)

Note that in the applet, $\nu = \text{degrees of freedom}$, $x = \chi_2 \text{ statistic}$, and multiply by 2 whatever output you see on $P(X > x)$ to get the *p*-value

5. What do you conclude?

6. Compare your results and conclusion in Part A with what you have gotten so far in Part B. How does the *z*-test of proportions differ from the chi-square test of independence of association?

7. Finally, return to Table 1 and say you conducted a chi-square test of association using statistical software and you have gotten following statistical output result below

```
##
##  Pearson's Chi-squared test
##
## data:  table1[, c(2, 3, 4, 5)]
## X-squared = 46.53, df = 9, p-value = 4.8e-07
```

Briefly report the appropriate null hypothesis, the chi-squared statistic, the *p*-value, and your conclusion.

8. Are you concerned that the analysis in number 7 met all the assumptions of a chi-squared test for independence of association? Justify.

END OF LAB