

Final Exam**How to Create and Submit This Assignment:**

1. Download this Word Doc and save it as: "Last name, SOCl 15, Final Exam" (That is, I'd save mine as "August, SOCl 15, Final Exam".)
2. Type your answers directly into the document beneath each question.
3. Submit your assignment to Canvas by Friday, December 11 at 11:59 p.m.

Chapter 7 – Estimation

Directions: Answer the questions using the hypothetical data below.

Average Hours of Sleep per Night during COVID-19

Group	N	Mean (hours)	Standard deviation (hours)
Parents	2000	6	2.5
Non-Parents	1000	7.5	2.0
Total	3000	6.5	2.33

Source: These data are made up.

1. In this section, we will follow a four-step process to calculate a confidence interval using the data in the table. The last step of this process is missing. What is it?
 1. Estimate standard error of the mean.
 2. Decide on the level of confidence, and find the corresponding Z value.
 3. Calculate the confidence interval.
 4. _____.

2. Use the formula below to estimate the value of the standard error of the mean for all the parents in the population, based on what we know about this one sample of parents.

$$s_{\bar{Y}} = \frac{s}{\sqrt{N}}$$

where s is the standard deviation from the sample and N is the number of cases in the sample.

3. Suppose we want to construct a 95% confidence interval around the mean number of hours of sleep for the sample of parents. The first thing we'd need to do is look up the z-scores that correspond to a 95% confidence level in the [Standard Normal Table](#). Do this, then report what z-scores we should use.
4. Next, calculate the maximum and minimum values of a 95% confidence interval around the mean number of hours of sleep for parents.

Use this formula for Z-scores to solve for Y :

$$Z = \frac{Y - \bar{Y}}{s}$$

where the values for Z are the ones you looked up in question 2, Y stands for the maximum and minimum values you're solving for, $\bar{Y}$ is the average value of hours of sleep for the sample of parents, and s is the standard error of the mean that you estimated in question 1.

- Interpret your result by stating how confident you can be that the true number of average hours of sleep all the parents in the population got was between the maximum and the minimum values that you calculated.

Chapter 8 – Testing Hypotheses

Directions: Answer the questions using the same hypothetical data in the table.

Average Hours of Sleep per Night during COVID-19

Group	N	Mean (hours)	Standard deviation (hours)
Parents	2000	6	2.5
Non-Parents	1000	7.5	2.0
Total	3000	6.5	2.33

Source: The data in this table are made up.

- The [CDC recommends](#) that adults ages 18-60 get 7 hours of sleep per night. According to the table above, non-parents in the sample got an average of 7.5 hours of sleep per night, with a standard deviation of 2.0 hours. Suppose you wanted to test whether non-parents in the population differ from the recommended 7 hours, in the amount of sleep they get each night. What would be your research hypothesis? (1 point)
- What is the null hypothesis for the test you identified in Question 6?
- In order to test the null hypothesis proposed in question 8, you have to make three assumptions. Two are listed below, but one is missing. What is it?

1. The level of measurement of the variable is interval-ratio
 2. _____
 3. The shape of the population distribution is normal and/or the sample size is larger than 50
9. Let's assume we can make the assumptions listed in Question 9. Next, to test the null hypothesis, we must estimate the value of the standard error of the sampling distribution for non-parents, using what we know about the non-parents sample. Use the equation below to calculate this estimate.

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{N}}$$

where $\sigma_{\bar{Y}}$ is the estimated standard error of the sampling distribution, σ is the standard deviation of the sample, and N is the size of the sample of non-parents.

10. Next, let's calculate how many degrees of freedom are associated with this test, using the equation below.

$$df = N - 1$$

where df is the number of degrees of freedom, and N is the size of the sample of non-parents.

11. Let's assume we want to be 99% confident in our decision to reject the null hypothesis. In the [t-distribution table](#), find the row corresponding to the number of degrees of freedom you calculated in Question 11. Look up the critical value of t- corresponding to a 99% confidence level for that distribution and write it below.

12. Now, let's calculate the obtained t-value for our sample of non-parents. Use this formula:

$$t = \frac{\bar{Y} - \mu}{s / \sqrt{N}}$$

Where t is our obtained value of t that we're solving for, $\bar{Y}$ is the mean value of y for the sample of non-parents, μ is the value proposed in our null hypothesis, and $s/\sqrt{N}$ is the standard error we estimated in Question 10.

13. Compare the obtained value of t you calculated in Question 13 to the critical value of t you looked up in Question 12. Based on this comparison, can you reject the null hypothesis with 99% confidence?
14. Interpret your decision in Question 14 by explaining what we can conclude about the true population of non-parents' number of hours of sleep per night.

Chapter 9 – Bivariate Tables

Directions: Use the data below to answer Questions 15 – 16.

Respondent	Education	Marital status	Respondent	Education	Marital status
1	College degree	Married	6	No degree	Not married
2	No degree	Not married	7	No degree	Married
3	College degree	Not married	8	College degree	Married
4	College degree	Married	9	No degree	Not married
5	No degree	Not married	10	College degree	Married

Source: The data in this table are made up, but let's assume they were collected from respondents chosen at random.

15. Fill in the blank bivariate frequency table below using the information in the above raw dataset. Follow the example in the [Ch. 9 & 10 google slideshow](#), slide 31. Make sure to replace the words "independent variable", "dependent variable" and "Category 1" and "Category 2" for each variable in the blank table with appropriate information from the raw dataset. Then fill in the appropriate values for the independent variable in the columns, the dependent variable in the rows, and the marginal values in the margins. Finally, give your table a title.

Title:

	Independent Variable		
Dependent Variable	Category 1	Category 2	Total
Category 1			
Category 2			
(N)	()	()	()

16. Construct a bivariate percentage table using the information in the frequency table you constructed in Question 16. Follow the example in the [Ch. 9 & 10 google slideshow](#), slide 32. Make sure to replace the words “independent variable”, “dependent variable” and “Category 1” and “Category 2” for each variable in the blank table with appropriate information from the raw dataset. Then fill in the appropriate values for the independent variable in the columns, the dependent variable in the rows, and the marginal values in the margins. Finally, give your table a title.

Title:

	Independent Variable		
Dependent Variable	Category 1	Category 2	Total
Category 1			
Category 2			
Total			

17. According to the [Chapter 9](#) slides, what term is used to describe “a relationship between two variables in which both the independent and dependent variables are influenced by a causally prior control variable, and there is no causal link between them”?

Chapter 10 – The Chi-Square Test and Measures of Association

Directions: Use the data below to answer Questions 18 – 25.

Support for Stricter Gun Laws by Age

	Age Group		
Support for Stricter Gun Laws	Under 35	36 and Older	Total
High	75%	45%	55%
Low	25%	55%	45%
Total	100%	100%	100%
N	(100)	(200)	(300)

Source: The data in this table are made up, but let's assume they were collected from respondents chosen at random.

18. When two variables are not associated, one can say that they are statistically independent. That is, an individual's score on one variable is independent of his or her score on the second variable. Are the variables "age" and "support for stricter gun laws" as summarized in the table above statistically independent?

19. Based on your answer to Question 18, what might you hypothesize about the relationship between "age" and "support for stricter gun laws"? State your research hypothesis, H_1 , below in your own words. (Refer to [Chapter 10 Zoom class](#) slide 16).

20. Given the research hypothesis you stated in Question 19, state the null hypothesis, H_0 .

21. The Chi-square test is an inferential statistical technique designed to test for significant relationships between two nominal or ordinal variables organized in a bivariate table. It does so by comparing the observed frequencies to the expected frequencies for each group. This means that in order to do a Chi-square test, we need to begin by determining the expected frequencies for each of the following 4 groups: younger gun law supporters, older gun law supporters, younger gun law opposers, older gun law opposers. Do so by filling in the blank cells in the table below.

Expected Frequency Table

	Age Group		
Support for Stricter Gun Laws	Under 35	36 and Older	Total
High			55%
Low			45%
Total	100%	100%	100%
N	(100)	(200)	(300)

22. The sampling distribution of Chi-square tells the probability of getting values of chi-square, assuming that no relationship between the independent and dependent variables exists in the population. But the Chi-square distribution is actually a family of distributions that look different depending on the number of degrees of freedom in the bivariate table we're using for our test. Calculate the number of degrees of freedom for this test, using the formula:

$$df = (r - 1)(c - 1)$$

where r is the number of rows and c is the number of columns.

23. Using the [Chi-square Table](#), determine the critical value of χ_c^2 if you want be 95% confident in your decision about whether or not to reject the null (i.e., if you want to reduce the likelihood of falsely rejecting the null to a 5 in 100 chance, or $p < 0.05$).
24. Suppose you used the formula in Chapter 10 and calculated a Chi-square obtained value of **23.05**. After comparing that value to the critical χ_c^2 value you looked up in Question 23, can you reject the null hypothesis with 95% confidence?

Chapter 12 – Regression and Correlation

Directions: Fill in the blanks for Questions 25-26.

25. "A linear prediction model, using one or more independent variables to predict the values of a dependent variable" is called _____.

26. We can use regression analysis to describe the relationship among two or more variables measured at the _____ level.

Directions: Use the hypothetical data in the table to answer Questions 27-30.

Years of Teaching Experience vs. Average Student Test Score

Respondent	Years of Teaching Experience (years)	Average Student Test Score (percent)
Teacher 1	1	72
Teacher 2	10	85
Teacher 3	5	86
Teacher 4	25	89
Teacher 5	3	78
Mean:	8.8	82

27. Draw a scatterplot using the data above. You can either: a.) do this by hand on paper, take a picture, and paste it into this doc. Or, b.) go to the website <https://www.meta-chart.com/scatter#/display>, input the information by following the directions [on the Chapter 12 Zoom class slides](#) #14-15, like we did in class. Make sure to include a title, labels on the X- and Y-axes, and units for your independent and dependent variables.

28. Add to the above scatterplot your best eyeballed or computer-generated estimate of a best-fitting line that minimizes the difference between the predicted average student test scores and the observed average student test scores.

29. Is the slope of the line you drew in #33 positive or negative?

30. Based on your answer for #34, write a sentence explaining your prediction about how the dependent variable will change as the value of the independent variable increases. (Make sure to state this in terms of teachers' experience and their students' average test scores. You can use this format: We predict that, as _____ increases, _____ will increase/decrease.)

