Name _____

Highways Pre-class Homework

Read the attached article, and answer the following questions.

1. Introduction: Explain, citing evidence from the article, how proximity to major traffic roads is related to public health.

2. Introduction: This study looked at all people who lived within 150 meters of a major roadway. Why did the study choose this number (150) as the cutoff?

3. Introduction: What do the authors hypothesize the relationship between socioeconomic status and air pollution exposure is?

4. Results: Which variables that were looked at saw large disparities of percentages living within 150 meters of a major highway? Which variables saw small/no disparities?

5. Discussion: What is the correlation between living in urban areas and people exposure to a major highway? Explain what this correlation value means.

6. Discussion: In light of your answer for (4), what is a possible lurking variable in this study? Explain how this new variable could affect the results.

Highways Classwork 1: Contingency tables, Independence

Use the Highways article to answer the following questions.

1. 11,337,933 people make up 3.7% of the total US population. Using that information, what is the total population of the US?

2. Consider the two populations labeled Non-Hispanic White and Non-Hispanic Black. Of these two, which population had more people living within 150 meters of a major highway? Which population had a larger percentage of people living within 150 meters of a major highway? Which of those statistics do you think is more meaningful in this conversation? Why?

3. Approximately how many people in the US speak primarily Spanish in their households?

4. How do the authors define the "Poor" characteristic? What how many people in the US, in total, are classified as "Poor" under this definition?

5. Using the information from the table found in the article, fill in the blanks to complete the following contingency table, **Table 1**. This compares people's proximity to major highways to the language predominantly spoken in their home. Each percentage shown is of the row total.

Language	Proximity		Total
	< 150 m	> 150 m	
English only	7,513,304 (3.3%)	220,162,575 (96.7%)	227,675,879 (100%)
Spanish	1,805,261 (5.1%)		
Other	1,059,572 ((4.9%)		
Total	10,378,137		

6. Notice that the number you found for "total" in the first column (10,378,137) differs a bit from the "Total" number found in the table in the article (11,337,933). Why do you think this is?

7. Repeat each of the counts from Table 1. Then divide each by the **table total** and express as a percent. Record below in **Table 2**.

Language	Proximity		Total
	< 150 m	> 150 m	
English only	7,513,304 (____ %)	220,162,575 (____ %)	227,675,879 (____ %)
Spanish	1,805,261 (____ %)		
Other	1,059,572 ((____ %)		
Total	10,378,137 ((____ %)		

8. What proportion of the US population primarily speaks Spanish in their home?

9. What proportion of predominantly Spanish-speakers live close to a major highway?

10. Of the people who live close to a major highway, what proportion of them are predominantly Spanish-speakers?

11. Two variables are **independent** if the conditional distribution for one variable remains unchanged with respect to the other variable. We will calculate what the distribution *would be* if we were to **assume independence**. To do this, first enter the marginal distributions (Total row at the bottom, Total column on the right), and the table total from Table 2. Then ignore the other values in Table 2; instead, fill in the value for (Row Total) X (Column Total)/(Table Total) . These are *expected counts* under the assumption of independence. Round values to 1 decimal place. Record below in **Table 3**.

Language	Proximity		Total
	< 150 m	> 150 m	
English only			227,675,879
Spanish			
Other			
Total	10,378,137		

12. Compare the counts in Table 2 and the expected counts in Table 3. Are they reasonably close? Do you think there is some evidence that Language and Proximity are independent? Explain briefly.

Highways Classwork 2: Hypothesis Testing

Recall the article "Residential Proximity to Major Highways in the United States" from earlier this semester. In our initial reading, we explored the relationship between several of the variables by using contingency tables. One major comparison we considered was between proximity to a major highway and language spoken at home. Today, armed with our new statistical technique of hypothesis testing, we wish to continue this discussion with a more quantitative focus.

As a reminder, the following contingency table displayed the language spoken of our participants in the rows, and the participants' distances from a major highway in the columns.

Language	Proximity		Total
	< 150 m	> 150 m	
English only	7,513,304 (3.3%)	220,162,575 (96.7%)	227,675,879 (100%)
Spanish	1,805,261 (5.1%)	33,592,013 (94.9%)	35,397,274 (100%)
Other	1,059,572 ((4.9%)	20,564,346 (95.1%)	21,623,918 (100%)
Total	10,378,137	274,318,934 (96.3%)	284,697,071 (100%)

Approximately 3.7% of the population is classified as living within 150 meters of a major highway. It stands to reason that if the variables were independent, we would expect approximately the same percentages to appear throughout the first column. Instead, we found that approximately 5.1% of native Spanish speakers live close to a major highway. Does this higher percentage imply that there is a considerable difference in the living situation of native Spanish speakers? Or, could this difference be explained by simple (and expected) statistical variance?

1. Let's set up a hypothesis test to evaluate this question. State the Null and Alternate Hypotheses
2. What are our values of a) the hypothesized value of the true population proportion, p_0 and b) the sample proportion $\hat{p}$? c) What is the value of the sample size, n?
3. What does the model for our sampling distribution look like? What is a) its center; b) its spread? c) Sketch and label the sampling distribution model. (See 5. for additions to this figure.)
4. Calculate the test statistic, which is the z-score associated with $\hat{p}$.
5. On the model above, show where $\hat{p}$ appears and shade the tail or tails determined by the alternative hypothesis.
6. Determine the P-value.
7. Describe what your P-value is telling you (using complete sentences).
8. Using a level of significance $\alpha = 0.05$, is our result statistically significant? What can we conclude overall?
9. Interpret your analysis from before. What reasons can you come up with that might explain your answers from before?