# Highways Classwork 2: Hypothesis Testing

Recall the article "Residential Proximity to Major Highways in the United States." 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, $p_0$

   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?