EXERCISE V: HYPOTHESIS TESTING

TOPICS COVERED:

- ❏ Test statistic
- ❏ P-value
- ❏ Conclusion and interpretation
- ❏ Normal hypothesis test for population proportion
- ❏ Hypothesis test for population mean
- ❏ Type I and Type II errors

- ❏ **REQUIRED DATASET: IPUMS-International**
- ❏ **REQUIRED VARIABLES:**
    1. COUNTRY
    2. YEAR
    3. CHBORN (children ever born)
    4. YRSCHOOL (years of schooling)
    5. AGE

[The only preselected variables that are needed in this exercise are COUNTRY and YEAR. Make sure to remove all of the other preselected variables by unchecking the blue boxes next to them. This will reduce the size of your data file and also make it easier to view the data in R.]

- ❏ **RECOMMENDED SAMPLES:**
    1. Armenia 2011
    2. Costa Rica 2011
    3. Uruguay 2011

❖ Question 1

Miguel is interested in studying average years of schooling in various countries around the world. His initial research focused on Costa Rica. He hypothesized that the mean years of school [YRSCHOOL] for people 18 years old or above is higher than 8.69 years.

In order to test his hypothesis, he drew a random sample from the 2011 census of 299,071 people. He found out that the mean number of years of schooling for his sample population is 8.70, with a SD of 4.52. Based on these results, with an alpha of .05, can Miguel reject the null hypothesis and conclude that the mean number of

years of schooling <u>in the population</u> is higher than 8.69? Conduct a full hypothesis testing process, as follows:

[Before conducting the test, examine missing values and the universe for YRSCHOOL. Restrict your sample appropriately]

A. **Write both hypotheses in your own words:**
   Null Hypothesis: The mean number of years of schooling in the population is equal or lower than 8.69 years.
   Research Hypothesis: The mean number of years of schooling in the population is higher than 8.69 years.

B. **Write both hypotheses using the correct symbols:**
   Null Hypothesis: <= 8.69
   Research Hypothesis: > 8.69

C. **Is that a one-tail or two-tail hypothesis? Why?**

   One-tailed/sided test

D. **Write down your sample statistics:**

   Mean (Ybar): 8.70
   SD (sY): 4.52
   N: 299071

E. **Calculate the t-test statistic using the appropriate equation:**

$$t = \frac{\bar{Y} - \mu_Y}{s_Y / \sqrt{N}}$$

   (8.705277 - 8.69) / (4.52/sqrt(299071))= 1.845

F. **Now conduct the appropriate test using R; what is the p-value?**

   P-value: 0.0325

G. What is the relationship between the t-test statistic and p-value stated above? Explain.

The p-value is calculated based on the t-test statistic. In this case, we use the t-test statistic (1.845), and search in the z-table for the matching probability, which is 0.0325 (i.e. the p-value)

H. What is the *meaning* of the p-value? Explain in your own words.

If the mean number of years of schooling in the (hypothetical) population (i.e. all Costa Ricans age 18 and older) is equal or lower than 8.69 years (in other words: if the null is true), then the probability of obtaining a test statistic as or more extreme than the one calculated is 0.0325.

I. What conclusion can Carlos draw from these results?

Since the p-value is less than 0.05, we reject the null hypothesis at the 0.05 level. We have enough evidence to conclude that the mean number of years of schooling in the population (i.e. people in Costa Rica) is higher than 8.69 years.

J. What *is* a Type I error, and what is its probability in our case?

A Type 1 error occurs if the null hypothesis is rejected when, in fact, the null is true. Therefore, the probability of a Type I error is equal to alpha, which in this case is 0.05.

❖ Question 2

When the null hypothesis (H0) is <u>true,</u> the probability of obtaining the value hypothesized on your research hypothesis (H1) or a more extreme value is called:

a) The 95% confidence interval
b) The Alpha value
c) Type I error
d) The P-value

❖ Question 3

Sara hypothesized that the average number of children ever born [CHBORN] to a Uruguayan woman is lower than 1.78. Unfortunately, Sara wasn't able to perform the hypothesis test due to problems with her software; however, she was able to obtain the confidence interval for the beta coefficient [CHBORN], which is: 1.7564 -- 1.7762.

Which of the following statements is true (more than one correct answer is possible)--

a) At a 5% confidence level, Sara can reject the null hypothesis and conclude that the mean in the population is indeed lower than 1.78.
b) At a 1% confidence level, Sara can reject the null hypothesis and conclude that the mean in the population is indeed lower than 1.78.
c) There is no possibility to determine the results of the hypothesis test without having access to the p-value.
d) Based on the confidence interval, we can infer that the sample mean is 1.776267.
