EXERCISE IV: CONFIDENCE INTERVALS

TOPICS COVERED:

- ❏ Unbiased estimators ($\underline{x}$ and $\hat{p}$)
- ❏ Interpretation of intervals
- ❏ Margin of error and standard error
- ❏ Confidence interval for population proportion
- ❏ Confidence interval for population mean
- ❏ t-distribution

- ❏ **REQUIRED DATASET: IPUMS-International**
- ❏ **REQUIRED VARIABLES:**
    1. COUNTRY
    2. YEAR
    3. BIRTHSLYR (number of births last year)
    4. EDATTAIN (educational attainment)

[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. Cambodia [2008]
    3. Portugal [2011]

*****Limit each sample to 10,000 households.**

❖ Question 1

We want a biased point estimator because it does not tend to underestimate or overestimate the true parameter.

a) True
b) False

❖ Question 2

If a statistic is unbiased, then the difference between the sampling distribution and the value of the true parameter is 0.

a) True
b) False

❖ Question 3

Which type of statistic do we prefer to work with when conducting confidence intervals and later on with hypothesis tests?

a) Biased with a small standard error
b) Biased with a large standard error
c) Unbiased with a small standard error
d) Unbiased with a large standard error

❖ Question 4

Which of the following is the definition for the margin of error (MOE)?

a) The margin of error measures how accurate a point estimate is likely to be in estimating a parameter.
b) The margin of error measures how accurate a point estimate is likely to be in estimating a statistic.
c) The margin of error measures how accurate a confidence interval is likely to be in estimating a parameter.
d) The margin of error measures how accurate a confidence interval is likely to be in estimating a statistic.

❖ Question 5

Suppose you have a confidence interval with a point estimate of 2.5 and a MOE of 0.06. Now suppose the MOE increases to 0.15. What happens to the interval and the accuracy of our estimate?

a) The interval decreases and the accuracy of our estimate decreases.
b) The interval decreases and the accuracy of our estimate increases.
c) The interval increases and the accuracy of our estimate decreases.
d) The interval increases and the accuracy of our estimate increases.

❖ Question 6

Select all of the following that are true about the t-distribution.

a) The t-distribution has wider/fatter tails than the normal distribution. (It has a larger spread.)
b) The shape and spread of the t-distribution does not depend on the degrees of freedom.
c) The t-distribution is bell-shaped.
d) The t-distribution is symmetric about 1.

❖ Question 7

Why does the t-distribution get closer to the normal distribution as the degrees of freedom increases?

a) The mean estimate gets better as n decreases.
b) The mean estimate gets better as n increases.
c) The standard deviation estimate gets better as n decreases.
d) The standard deviation estimate gets better as n increases.

❖ Question 8

Laila is interested in the number of births that occur in Cambodia every year. Using the 2008 census, she calculates the mean number of children born in the year before the census.

a) Examine the missing values for BIRTHSLYR. Define the population included in each missing value category.

8 Unknown: Females age 15 to 49 who did not provide a valid response to the census question on births during the last year.

9 NIU (not in universe): All males; Females under age 15 or over age 49.

b) Create and interpret a 95% confidence interval for the true mean number of children born to Cambodian women age 15 to 49 in the year before the 2008 census (exclude missing values). Calculate the interval by hand using the following formula:

$$\overline{X} \pm t \frac{s}{\sqrt{n}}$$

$\overline{x}$= 0.0493521
s = 0.2228004
t = 1.960015
n = 12,502

(0.0493521-1.960015*(0.2228004/sqrt(12,502)), 0.0493521+1.960015*(0.2228004/sqrt(12,502))) = (0.04544652, 0.05325768)

Interpretation: We are 95% confident that the true mean number of children Cambodian women had in the last year lies between 0.04544652 and 0.05325768.

c) Now calculate the interval using the t.test function in R.

95% CI: (0.04544625, 0.05325796)

d) Now create and interpret a 99% confidence interval for the true mean number of children Cambodian women had in the last year. Calculate the interval by hand using the following formula:

$$\overline{X} \pm t \frac{s}{\sqrt{n}}$$

$\overline{x}$ = 0.0493521
t = 2.575936
s = 0.2228004
n = 12,502

(0.0493521-2.575936*(0.2228004/sqrt(12,502)), 0.0493521+2.575936 *(0.2228004/sqrt(12,502))) = (0.04421922, 0.05448498)

Interpretation: We are 99% confident that the true mean number of children Cambodian women had in the last year lies between 0.04421922 and 0.05448498.

e) Now calculate the interval using the t.test function in R.

99% CI: (0.04421865, 0.05448556)

❖ Question 9

Ma used a random sample of 10,000 households from the Portugal 2011 census to calculate the proportion of individuals that had completed University [EDATTAIN].

a) Create and interpret a 98% confidence interval for the true proportion of people from Portugal who completed university. Calculate the interval by hand using the following formula:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$ = 0.1194825
z = 2.326348
n = 25,895

(0.1194825-2.326348*sqrt(0.1194825*(1-0.1194825)/25,895), 0.1194825+2.326348*sqrt(0.1194825*(1-0.1194825)/25,895) = (0.1147934, 0.1241716)

Interpretation: We are 98% confident that the true proportion of people from Portugal who completed university lies between 0.1147934 and 0.1241716.

b) Now calculate the interval using the prop.test function in R.

98% CI: (0.1148538, 0.1242709)

c) Who was asked about educational attainment in the Portugal 2011 census?

All persons

d) Would you expect the proportion of people completing university to be lower or higher if only persons age 15 and older were asked about educational attainment in the Portugal 2011 census?

The proportion of persons completing university would be higher. By excluding persons under age 15, you remove people from the denominator of the proportion but not from the numerator of the proportion because no one under age 15 will have completed university.

❖ Question 10

Manny wishes to draw a sample of Ghana 2010 census data in order to estimate the proportion of people in the population who have a disability. How many people should Manny include in his sample in order to be 95% confident that the margin of error is within 0.01 of the true proportion?

0.01 = 1.96*sqrt(0.5*(1-0.5) / n)
0.01/1.96 = sqrt(0.5*(1-0.5) / n)
(0.01/1.96)^2 = 0.5*(1-0.5) / n
n = (0.5*(1-0.5)) / (0.01/1.96)^2
n = 9604 people