ECOM20001: Econometrics 1

Tutorial 3: Hypothesis Testing of Sample Means, p-values, Confidence Intervals, Testing Differences of Means Between Independent Samples

---

Part 1: Hypothesis Testing and Confidence Intervals for Sample Means in R

A. Getting Started

Please create a Tutorial3 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Tutorial3 folder:

- tute3.R
- tute3_cps.csv

The first file is the R code for Tutorial 3, the second file is the .csv file that contains the dataset for the tutorial.[1] The dataset has the following 5 variables:

- year: year individual was randomly surveyed; either 1992 or 2012
- ahe: individual's average hourly earnings (in real terms, 2012=100)
- bachelor: equals 1 if individual has a bachelor degree, 0 otherwise
- female: equals 1 if individual is female, 0 otherwise
- age: age of the individual at time of survey

In total, the dataset contains this information for 15,052 individuals in the U.S.

B. Go to the Code

With the R file downloaded into your Tutorial3 folder, you are ready to proceed with the tutorial. Please go to the tute3.R file to continue with the tutorial.

---

[1] The reference for these data is the Current Population Survey (CPS) which is collected by the U.S. Department of Labor Statistics and provides individual-level data on the population, employment, and earnings. It is constructed from randomly sampling the U.S. population. For details, see https://www.census.gov/programs-surveys/cps.html

C. Questions

Having worked through the tute3.R code and graphs, please answer the following:

1. What is the sample mean and standard deviation of AHE for males and females? Discuss these numbers and the figure produced in ahe_female.pdf, which reveals what is known as the gender earnings gap.
   - Provide **economic explanation(s)** for your results. Recall from Tutorial 2 that an economic explanation focuses on the costs and benefits of a particular behaviour for explaining empirical patterns.
   - In this example, what are the different economic costs and benefits among males and females in generating household earnings?
2. What is the sample mean and standard deviation of AHE for individuals with and without bachelor degrees? Discuss these numbers and the figure produced in ahe_bachelor.pdf. Provide economic explanation(s) for your results.
3. What is the 95% confidence interval (CI) for AHE? Report the results from the following tests:
   - H0: mean(AHE) = 19.5 vs H1: mean(AHE) != 19.5 (!= means “not equal”)
     H0: mean(AHE) = 19.5 vs H1: mean(AHE) > 19.5
     H0: mean(AHE) = 19.5 vs H1: mean(AHE) < 19.5
4. Discuss the difference in mean AHE for 2012 from a two-sample t-test for the gender earnings gap among males and females without bachelor degrees:
   - H0: mean(AHE_Female_2012_NoBach) = mean(AHE_Male_2012_NoBach) vs H1: mean(AHE_Female_2012_NoBach) != mean(AHE_Male_2012_NoBach)
5. Now discuss the difference in mean AHE for 2012 from a two-sample t-test for the gender earnings gap among males and females with bachelor degrees:
   - H0: mean(AHE_Female_2012_Bach) = mean(AHE_Male_2012_Bach) vs H1: mean(AHE_Female_2012_Bach) != mean(AHE_Male_2012_Bach)

In both questions 4 and 5, report the difference in sample means, test results (assuming 5% level of significance), and the 95% CI for the difference. Provide economic explanation(s) for the difference in your test results in questions 4 and 5. Why do you think the gender earnings gap differs among males and females without and with bachelor degrees? In answering this question, discuss what is happening in figure ahe_female_bachelor_2012.pdf.

Note: Part 2 of this tutorial contains extra practice exercises and will potentially only be partially covered in the tutorial, depending on time remaining. Solutions will be provided for students to work through and follow-up on in consultations.

Part 2: Practice Problems
Hypothesis Testing and Confidence Intervals

Throughout use R to compute p-values in answering the questions if needed using the pnorm() command as described in the comments in tute3.R.

1. Suppose you collected AHE from a random sample n=5,000 of Victorians. You compute the sample mean of $28.25 and sample standard deviation of $10.66.

a. Conduct a two-sided hypothesis test of the null that the population mean is $28 using both the p-value and critical-value approaches to hypothesis testing. Use a 5% level of significance for your test.

b. Construct a 95% CI for the population mean

c. Report the p-value for the two-sided hypothesis test of the null that the population mean is $28, as well as the 95% CI for the following sample sizes:

- n=2,500
- n=5,000
- n=10,000
- n=20,000

Does the p-value for the test go up or down as the sample size rises? Does the 95% CI expand or shrink as the sample size rises? Explain the intuition for your findings regarding sample size, p-values, and confidence interval width.

2. Return to your original sample of n=5,000 Victorians with sample mean AHE of $28.25 and sample standard deviation of $10.66. Suppose you also randomly sampled m=3,000 individuals from NSW and obtained a sample mean AHE of $30.88 and sample standard deviation of $11.22.

a. Construct a 95% CI for the population mean of AHE for the individuals from NSW. Is it wider or more narrow than the 95% CI for AHE from Victoria? Explain.

b. Conduct a two-sample t-test of the null that the difference in mean AHE for individuals in Victoria and NSW is 0. Conduct the test using both the p-value and critical-value approaches. Use a 5% level of significance for your test.

c. Report the 95% CI for the difference in the mean AHE between Victoria and NSW.

3. Suppose you have a random sample of data with a mean ***m***, and you conduct the following hypothesis test:

H0: ***m*** = 10 vs H1: ***m*** != 10 (!= means “not equal”)

Having performed the test, you obtain a p-value of 0.07.

a. Does the 90% CI for the population mean contain ***m*** = 10? Explain.

b. With the information provided in the question, can you determine if ***m*** = 8 is contained in the 90% CI? Explain.