Problem Set #3: Estimation and Inference — November 16, 2020

Quantitative Methods

Daniel Sanchez Moscona

To hand in online before 23:59 on November 23rd.

1. The scores on a literacy test in Bangladesh (out of a 100) are such that the population mean $\mu = 58$ while the population standard deviation is $\sigma = 18$.

   (a) Anna, a researcher gathers data on the scores of a sample of n students from Bangladesh. Let $\bar{X}_n$, the (sample mean of a sample of n observations) be an estimator of μ. The researcher makes the following assertion: "*Given that $\bar{X}_n$ is unbiased, the estimate obtained from a sample of 10 individuals will be identical to the one obtained from a sample of 50 individuals.*" Is her statement correct? Why or why not?

   (b) Ben concludes that: "*Since $\bar{X}_n$ is a consistent estimator, then, with complete certainty, we will obtain an estimate that is closer to μ if the sample size is* 100 *instead of* 10." Is his statement correct? Why or why not?

   (c) Suppose $n = 49$. What is the probability that the researchers obtain an estimate between 55 and 61?

   (d) The researchers are thinking of whether to invest in increasing the sample size to $n = 100$. They judge that the benefits of this outweigh the costs only if the probability of obtaining an estimate between 55 and 61 increases by more than 10 percentage points. Should they make this investment? Justify your answer.

2. Suppose that Greta, a social scientist, wants to estimate the (population) mean number of years of schooling in Guatemala μ.

   (a) She wants to construct a 90 % confidence interval for μ. In her report, she writes the following: "*Once I do the necessary calculations, the interval that I obtain will contain μ with a probability of 0.90*". Is she right in making such a statement? Why or why not?

   Greta knows that the population standard deviation is $\sigma = 5$. After collecting a sample of 10 individuals, she obtains a sample mean estimate of $\bar{x} = 13.58$.

   (b) Calculate the 90, 95 and 99% confidence intervals.

   (c) Suppose that Greta is given the opportunity to increase her sample size. What is the minimum number of observations that she needs if she wants her 99 % confidence interval to be **no longer** than 1 unit?

(d) Suppose that she is limited to her initial 10 observations. What confidence level will let Greta construct a confidence interval of length equal to 1 unit?

(e) Suppose that now σ is an unknown quantity. It turns out that the sample standard deviation she obtains is $s = 5$. Construct the 90, 95 and 99% confidence intervals in this case. How do they compare to the ones you obtained in part (b)? Why is this the case?

3. Suppose that a researcher gathers data on wages in Spain from a sample of 81 observations. She wants to test the hypothesis that the (population) mean monthly wage is $\mu = 1000$ against the alternative that $\mu \neq 1000$. Suppose that she obtains a sample mean estimate $\bar{x} = 940$ and that she knows that the population standard deviation is $\sigma = 100$.

   (a) Carry out the hypothesis test for levels of significance $\alpha \in \{0.1, 0.05, 0.01\}$. Interpret the meaning of α.

   (b) Calculate the p-value and relate the value you obtained to your decision on whether or not to reject the null hypotheses in part (a).

   (c) Calculate the 90, 95 and 99% confidence intervals for μ. Relate your findings to your decision on whether or not to reject the null hypotheses in part (a).

4. (a) Suppose a researcher gathers data on the surface of Spanish homes. She constructs a sample of 35 houses and she wants to test the null hypothesis $H_0 : \mu = 60$ versus the alternative that $H_1 : \mu > 60$. Given that the sample mean she obtained is $\bar{x} = 65$ and that the sample standard deviation is $s = 10$, should she reject the hypothesis at significance levels $\alpha \in \{0.1, 0.5, 0.01\}$? Why or why not? Include the p-value in your analysis.

   (b) The researcher now gathers data on the surface 43 Mexican houses. She wants to test the hypothesis that $H_0 : \mu = 70$ versus the alternative $H_1 : \mu < 70$. Given that the sample mean she obtained is $\bar{x} = 65$ and that the sample standard deviation is $s = 10$, should she reject the hypothesis at significance levels $\alpha \in \{0.1, 0.5, 0.01\}$? Why or why not? Include the p-value in your analysis.

5. In this problem, you will perform a simulation. You will ask a program to simulate data from 100 samples that are extracted from the same population. The program will compute the confidence interval for each sample given a number of observations (n), a value of the true population mean μ and standard deviation $sigma$ and a confidence level. Follow these steps:

   (a) Log on to the website: `https://rdrr.io/snippets/`

   (b) Paste the following two lines of code:

      - library(PASWR2)

- cisim(samples = 100, n = , parameter = , sigma = , conf.level =)

(c) Fill in some values of your choice for n (number of observations in each sample), "parameter" (the value of μ), "sigma" (the value of σ) and "conf.level" (confidence level, expressed in decimal terms - for example, 90% confidence is conf.level=0.90). See Figure 1 for an example.

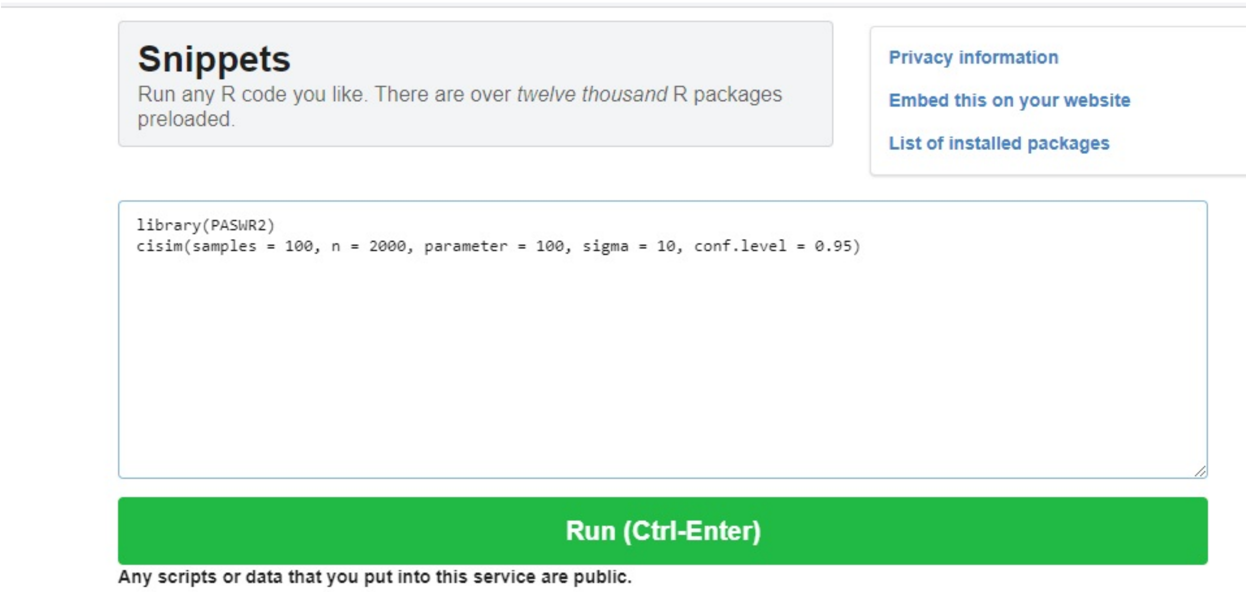

Figure 1: Screenshot of `https://rdrr.io/snippets/`

(d) When you click "Enter", the webpage will return an image of 100 vertical lines. Each of them represents a confidence interval (for a different sampl). The horizontal line in the middle represents the true value of μ. Paste a screenshot of the image you got. Moreover, interpret the output you got from this simulation and relate it to the values you chose in part (c).