Tutorial 2

Applied Stats/Quant Methods 1

Week 2, Fall 2021

1 Bias and Efficiency

This problem requires that you go over the concepts of “bias” and “efficiency” in Kmenta. In `R`, draw a random sample of 1,000 observations from a normal distribution with $\mu = 100$ and $\sigma = 25$ and define these observations as the population of interest. Then draw 50 samples of 50 observations each from this population. Use each of these samples to estimate *mean* and *median* of the population distribution. Then describe the sampling distribution of means and medians after 1, 10, and 50 samples: Are mean and median unbiased estimators of the center of the population distribution? Are they both equally efficient? (Hint: You may want to look at the `R` documentation for functions such as `rnorm`, `sample`, and `replicate`.)

2 The t Distribution

The following questions are about Student’s t distribution. The first questions could be answered from readily-available tables of the t distribution, but please work on them using the relevant `R` functions (`qt()`, `rt()`, and `pt()`.)

(a) Describe the purpose of each of these `R` functions (i.e., `qt()`, `rt()`, and `pt()`).

(b) Consider a random variable t that is distributed Student-t with 20 degrees of freedom. What is the probability that a draw from t would be greater than 1.45?

(c) What is the probability that the absolute value of a draw from t, $|t|$, would be greater than 1.45?

(d) What are the quantiles of a t_5 distribution for a two-tailed hypothesis test at the 89% confidence level?

(e) Call the quantiles in the previous exercise $-c$ and c. Sample $n = 1000$ draws from a t_5-distributed variable. How many of these draws are in fact larger than $|c|$?

3 Hypothesis Testing

(a) (From DeGroot 9.5) Suppose that a random sample $X_1, ..., X_n$ is to be taken from the normal distribution with unknown mean μ and unknown variance σ^2, and the hypotheses to be tested are

$$H_0: \ \mu \le 3, \quad H_1: \ \mu > 3.$$

Suppose also that the sample size n is 17, and it is found from the observed values in the sample that $\bar{X}_n = 3.2$ and $(1/n)\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = 0.09$. Calculate the value of the appropriate statistic, and find the corresponding p-value.

(b) Consider the conditions for (a), but suppose now that the hypotheses to be tested are

$$H_0: \ \mu = 3.1, \quad H_1: \ \mu \neq 3.1.$$

Suppose, as in (a), that the sample size n is 17, and it is found from the observed values in the sample that $\bar{X}_n = 3.2$ and $(1/n)\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = 0.09$. Calculate the value of the appropriate statistic, and find the corresponding p-value.

4 Building Confidence Intervals

Go to this website and take a look at the "confidence interval applet" therein. Write an R function that takes as input user- defined values for μ, n, σ, and α and returns 100 realizations of confidence intervals. In other words, write `R` code to replicate what the applet does.