CGS602A: Basic Statistics, Data Analysis and Inference

Assignment #2:

`Max marks: 90`

`Due on/before: midnight, 27-Nov-2020.` `18-Nov-2020`

1. (a) You are doing a speed survey for vehicles on a stretch of highway. You measure 85 vehicles and get an average speed of 66.3Km/h. You know that $\sigma = 8.3$Km/h. Find the 95% CI for μ.
   (b) A random sample of size $N = 16$ is taken from a normally distributed population with unknown μ and σ. The sample mean is $\bar{x} = 27.9$ and the standard deviation is $s = 3.23$. What is the 95% CI for μ?
   (c) For an infinite normal population the $\sigma = 1.5$. If the 95% CI is to be 2 what should be the sample size?

[6,6,6 =18]

2. Supposing the following data is given to you:
   - The population is normally distributed.
   - $\mu = 100.3$, $\sigma = 6.25$.
   - Based on the treatment you expect a higher μ.

   For each part below show the necessary calculations.

   (a) A sample ($N = 15$) from the treated group is taken and the sample mean is $\bar{x} = 105.6$. If $\alpha = 0.01$ and you do a two tailed test for $H_0 = 100.3$ give the alternative hypothesis and the result of the test.
   (b) Now let $N = 20$, $\bar{x} = 105.0$, $\alpha = 0.05$ and you do a one tailed test for $H_0 = 100.3$. State the alternative hypothesis and the result of the test.
   (c) You want the population mean to be at least 5 more after treatment. What is the probability of a type II error for b) if the population mean after treatment is actually 105.3 rather than the hypothesized 100.3?
   (d) Repeat c if the actual mean of the population after treatment is 101.3.

[8×4=32]

3. In this question you have to do simulations in Python.
   (a) Assume that you have a discrete distribution given below:
   $P(1) = \frac{1}{3}$, $P(2) = \frac{1}{3}$, $P(9) = \frac{1}{3}$. Let M be the number of N-sized samples drawn from the distribution above. Construct the sampling distribution for for $N = 2, 10, 20, 40$ for different values of $M = 100, 10^3, 10^5$. Plot the resulting sample distributions. What do you conclude from this experiment?
   (b) In this simulation you will experiment with bootstrapping. In the general case bootstrapping means repeated resampling of samples of size M from a single sample of size N with replacement where $M \leq N$. Choose a sample of size $N = 15, 20, 40$ from a standard normal distribution. Then do bootstrap sampling for the following values of $M = 5, 10, N$. Do this for 100, 1000 and 10^5 bootstrap samples. Find the critical value $z^\star_{low}$ and $z^\star_{high}$ for $\alpha = 0.5$ by finding the values of z for which the smallest 2.5% of the samples are less than z to get $z^\star_{low}$ and similarly 2.5% are greater than z to get $z^\star_{high}$. Compare with the standard $z^\star$ values. What do you conclude?

[20,20=40]