Stat 2150 - Worksheet 10

Question 1 The `Response_Time` dataset on UMLearn contains measured response times for 200 students. Use bootstrapping to do each of the following.

(**a**) Let S represent the sample standard deviation, used to estimate σ, the population standard deviation. Obtain an approximation of the sampling distribution of S. Also, approximate the bias of S and give a bias-corrected estimate of σ. Finally, approximate the variance of S, and determine a 95% CI for σ.

(**b**) Let H represent the sample IQR, used to estimate the population IQR. Obtain an approximation of the sampling distribution of H. Also, approximate the bias of H and give a bias-corrected estimate of the population IQR. Finally, approximate the variance of H, and determine a 95% CI for the population IQR.

Question 2 Let $X \sim \text{Exp}(0.5)$. The goal is to estimate the population median from the sample median, with a sample of size $n = 40$.

(**a**) Take a sample of size $n = 40$ from X. Use the seed `set.seed(2)`. Use bootstrapping to approximate the bias of $\tilde{X}$, the sample median, and use this to obtain a bias-corrected estimate of the population median.

(**b**) Run simulations to determine the true bias of your estimator.

(**c**) Were you able to remove some bias through bootstrapping, for this particular sample?

(**d**) Repeat this 1000 times to estimate the bias of the corrected estimators, and compare this to the bias of the uncorrected estimators.

Question 3 The dataset `React_85` contains the reaction times for a group of 85 12th-grade High School students.

(**a**) Run a permutation test to see if there is a difference between the reaction times for male and female students, at the 5% level of significance.

(**b**) Repeat part (a), using other test statistics. Do you observe different test outcomes?

Question 4 Let $X_1 \sim \mathrm{N}(\mu_1, \sigma = 4)$, and $X_2 \sim \mathrm{N}(\mu_2, \sigma = 4)$. Our goal is to test the hypotheses

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2$$

with a permutation test, using samples of size $n = m = 15$.

(**a**) Set $\mu_1 = \mu_2 = 0$. Take random samples of size 15 each from X_1 and X_2. Run a permutation test to see what the outcome of the test would be.

(**b**) Set $\mu_1 = 1, \mu_2 = 0$. Take random samples of size 15 each from X_1 and X_2. Run a permutation test to see what the outcome of the test would be. Classify your decision as either correct, a Type I error, or a Type II error.

(**c**) Repeat part (a) 1000 times, and record the P-value each time. It is known that, when H_0 is true, P has a $\mathrm{U}(0, 1)$ distribution. Is this what you observe?

(**d**) Repeat part (b) 1000 times, and record the P-value each time. Calculate how many of these tests lead to a rejection of H_0. This is called the *power* of the test, for an effect size of $\dfrac{\mu_1 - \mu_2}{\sigma} = 0.25$.

(**e**) Repeat part (d), for sample sizes $n = m = 25, 50, 100, 200$. Based on these simulations, what is the minimum sample size such that the power of the test is at least 80% for this effect size?