MTH 372: Final Semester Exam

May 8, 2023

1. Let $X_1, \ldots, X_n$ be i.i.d from Pareto population with its pdf given by

$$f(x_i) = \theta \gamma^{\theta} x_i^{-\theta-1} I(\gamma < x_i < \infty)$$

where $\theta > 0, \gamma > 0$ are unknown parameters. Answer the following

   (a) (3 points) Find the MLE of θ and γ.

   (b) (3 points) Find equal tailed α level likelihood ratio test for $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$.

2. (3 points) Let $X_1, \ldots, X_n$ be i.i.d from $N(\mu_1, \sigma^2)$, $Y_1, \ldots, Y_n$ be i.i.d from $N(\mu_2, 3\sigma^2)$ with $-\infty < \mu_1, \mu_2 < \infty$, $0 < \sigma < \infty$, and $\mathbf{X}, \mathbf{Y}$ be independent. Derive the distribution of $V_n = \frac{1}{2\sigma}\sqrt{n}(\overline{X}_n - \overline{Y}_n - \mu_1 + \mu_2)$, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$.

3. (3 points) An industrial safety program was recently instituted in the computer chip industry. The average weekly loss (averaged over 1 month) in labor-hours due to accidents in 10 similar plants both before and after the program are as follows:

Before	30.5	18.5	24.5	32	16	15	23.5	25.5	28	18
After	23	21	22	28.5	14.5	15.5	24.5	21	23.5	16.5

Determine, at the 5 percent level of significance, whether the safety program has been proven to be effective. Assume that the difference of this random sample comes from Normal population.

4. (1 point) Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance σ^2, unknown. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ. Use $\overline{x} = 9, s^2 = 9.5$.

5. (2 points) Twenty-two volunteers at a cold research institute caught a cold after having been exposed to various cold viruses. A random selection of 10 of these volunteers was given tablets containing 1 gram of vitamin C. These tablets were taken four times a day. The control group consisting of the other 12 volunteers was given placebo tablets that looked and tasted exactly the same as the vitamin C tablets. This was continued for each volunteer until a doctor, who did not know if the volunteer was receiving the vitamin C or the placebo tablets, decided that the volunteer was no longer suffering from the cold. The length of time the cold lasted was then recorded. At the end of this experiment, the data was obtained, that resulted in $\overline{X}$ =6.450, $\overline{Y}$ =7.125 S_x^2 = .581, S_y^2 = .778, where sample X corresponds to those receiving vitamin C and the Y sample to those receiving a placebo.
   Do the information above prove that taking 4 grams daily of vitamin C reduces the mean length of time a cold lasts? Use $\alpha = 0.05$. (Assume that the samples come from Normal population.)

6. (3 points) A randomly chosen group of 20,000 nonsmokers and one of 10,000 smokers were followed over a 10-year period. The following data relate the numbers of them that developed lung cancer during that period.

	Smokers	Nonsmokers	Total
Lung cancer	62	14	76
No lung cancer	9,938	19,986	29,924
Total	10,000	20,000	30,000

Test the hypothesis that smoking and lung cancer are independent. Use the 1 percent level of significance.

7. (3 points) Let X_i denote the length, in centimeters, of a randomly selected pygmy sunfish, $i = 1, 2, \ldots, 10$. If we obtain the following data set:
5.0, 3.9, 5.2, 5.5, 2.8, 6.1, 6.4, 2.6 ,1.7, 4.3
Using hypothesis testing, can we conclude that the median length of pygmy sunfish differs significantly from 3.7 centimeters?

8. The independent random samples of performance IQ scores are obtained from a study involving children who lived near a large ore smelter in Texas. The 121 children were partitioned into independent groups based on their measured blood lead levels. The blood lead level has three different categories: low, medium, and high blood lead levels. Assume that the three populations appear to have distributions that are approximately normal, and the standard deviations that are not dramatically different, so the three population variances appear to be about the same. Make use of the below given table to answer the questions.

Source	df	Sum of Squares (SS)	Mean Sum (MS)	F	p-value
Treatment					0.0195
Error	118	29314.047			
Total		31336.777			

(a) (2 points) Fill in the above table.

(b) (3 points) Use the above given information and a significance level of $\alpha = 0.05$ to test the claim that the three samples come from populations with means that are all equal.

9. (3 points) Consider a simple linear regression model with parameters $\beta_0, \beta_1, \sigma^2$, such that for any fixed value of the independent variable x_i, the dependent variable is a random variable related to x_i through the model equation $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$; $\epsilon_i \sim N(0, \sigma^2)$. The ϵ_i are iid and $i = 1, \ldots, n$.
Find an unbiased estimator of β_0.
(Assume, all the assumptions of simple linear regression are satisfied.)

10. (Bonus Question: 5 points) Suppose a random sample with $X_i, i = 1, \ldots, n$ and pdf

$$f(x_i) = 2(\theta - x_i)\theta^{-2}\, I(0 < x_i < \theta),\ \theta > 0.$$

Find a $(1 - \alpha)$ confidence interval for θ.