MATH 427- Homework 1-Spring 2023

Due: Jan-30-2023

1 Conceptual problems

1. For any random variables X, Y and any constants a, b, c and d. Show that
$COV(a + bX, c + dY) = bdCOV(X, Y)$.

2. Suppose that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $V(\hat{\theta}_1) = \sigma_1^2$, and $V(\hat{\theta}_2) = \sigma_2^2$. Consider the estimator
$\hat{\theta}_3 = \alpha\hat{\theta}_1 + (1 - \alpha)\hat{\theta}_2$.

(a) Show that $\hat{\theta}_3$ is an unbiased estimator for θ

(b) If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, how should α be chosen in order to minimize the variance of $\hat{\theta}_3$?

3. Suppose that X_1, X_2, X_3 denote a random sample from an exponential distribution with density function

$$f(x) = \begin{cases} \left(\frac{1}{\theta}\right) e^{\frac{-x}{\theta}}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

consider the following four estimators of θ:

$\hat{\theta}_1 = X_1$, $\hat{\theta}_2 = \frac{X_1 + 2X_2}{3}$, $\hat{\theta}_3 = \bar{X}$, $\hat{\theta}_4 = min(X_1, X_2, X_3)$.

Which of these estimators are unbiased? (Show your work.)

2 Applied problems-R

1. This exercise relates to the "Credit" data set, which can be found as "Credit.csv" in Canvas.

(a) Use the appropriate function in R to produce a numerical summary of the quantitative variables in the data.

(b) Display a scatter plot matrix between quantitative variables in the data set.

(c) Display histograms of all quantitative variables in one graph except "cards" (side by side). You may use different colors.

(d) Display box-plots of all quantitative variables only in one graph except "cards" (side by side). Make sure to label them.

2. This exercise relates to the "Hwk-data1" data set, which can be found in Canvas.

Operators of gasoline-fueled vehicles complain about the price of gasoline in gas stations. According to the American Petroleum Institute, the federal gas tax per gallon is constant (18.4 cents as of January 13, 2005), but state and local taxes vary from 7.5 cents to 32.10 cents for n = 18 key metropolitan areas around the country.

Use R for part (a), (b), and (c).

(a) Check whether the data are normally distributed by using the Shapiro test or by looking at the QQ plot. (Make sure to display your results).

(b) Use the appropriate R function to find a 90% confidence interval for the average per gallon gas tax in the U.S. (Make sure to display your code and the corresponding result.)

(c) Is there sufficient evidence to claim that the average gas tax is less that 45.2 cents?(Make sure to specify hypotheses and the p-value).

(d) Compute(by hand) a 98% confidence interval for the average per gallon gas tax in the U.S. Compare the length of this interval and the one in part (b). (Hint: the sample standard deviation$s = 7.138$)