Mock Interim Test

MA1202, Introductory Statistics

March 9, 2022

Question 1 [10 points]

The following data given below in Table 1.1 are the mileage ratings of 40 cars of a new car model determined for an environmental survey. The 40 mileage ratings are given in ascending order.

Table 1. The mileage ratings (x) for the 40 cars.

29.8	31.1	32.0	32.1	33.0	33.4	33.5	33.7	33.8	34.1
34.3	35.2	35.2	35.7	35.7	35.8	35.9	36.3	36.7	37.1
37.1	37.1	37.2	37.2	37.2	37.5	38.4	38.8	38.8	39.1
39.3	39.4	39.6	39.9	40.8	40.8	41.0	41.1	42.9	43.0

i) If the first sample moment is $m_1 = 36.765$ and the second sample moment is $m_2 = 1361.579$, find the sample variance and the sample standard deviation for x. (Show your working)

ii) Find the sample median, inter-quartile range, and range for the data. (Show your working)

iii) Construct a table of the frequency distribution for the data with grouping intervals:

$$(0, 32], (32, 34], (34, 36], (36, 38], (38, 40], (40, 42], (42, \infty).$$

iv)Calculate the minimum required size of the sample to estimate true mean value, at level 95%, to be within 0.5 units. Assume that a previous study has shown $\sigma = 3$.

Solution:

i) The sample variance (unbiased) - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \tilde{S}^2 \cdot \frac{n}{n-1}$, where $\tilde{S}^2$ is the biased estimator: $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}((X_i - \bar{X})^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = M_2 - M_1^2$. Hence, the sample variance is :

$$s^2 = (m_2 - m_1^2) \cdot \frac{n}{n-1} = (1361.579 - 36.765^2) \cdot (40/39) = 10.168\ (3d.p.)$$

The sample standard deviation - $s = \sqrt{s^2} = 3.189$.

ii) Median - $m = \frac{x_{(20)} + x_{(21)}}{2} = 37.1$.

IQR - $IQR = Q_3 - Q_1 = \frac{x_{(31)} + x_{(30)}}{2} - \frac{x_{(11)} + x_{(10)}}{2} = 5$.

Range - $r = x_{(40)} - x_{(1)} = 13.2$.

iii)

Range	Frequency	Relative Frequency
$(0, 32]$	3	0.075
$(32, 34]$	6	0.15
$(34, 36]$	8	0.2
$(36, 38]$	9	0.225
$(38, 40]$	8	0.2
$(40, 42]$	4	0.1
$(42, \infty)$	2	0.05

iv) The minimum required size of the sample to estimate true mean value, at level 95%, to be within 0.5 units can be found from:

$$n = \lceil \frac{z_{\alpha/2}^2 \sigma^2}{E^2} \rceil = \lceil \frac{1.96^2 3^2}{0.5^2} \rceil = 139.$$

Question 2 [10 points]

1. [8 points] A random sample $X_1, X_2, ..., X_n$ ($n \geq 2$) of independent observations is taken from a Geometric distribution with parameter $\theta > 0$, whose probability function is given by

$$p(x) = \theta^x (1+\theta)^{-(x+1)}, \quad \text{for} \quad x = 0, 1, 2, ...$$

i) Write down the likelihood function $L(\theta, X_1, X_2, ..., X_n)$ and show the maximum likelihood estimator $\hat{\theta}$ for θ is given by $\hat{\theta} = \bar{X}$.

ii) Using the facts that $\hat{\theta} = \bar{X}$ is an unbiased estimator for θ and that $E(X) = \theta$ and $E(X(X-1)) = 2\theta^2$, obtain the variance of $\hat{\theta}$ and compare it to CRLB for θ. Write your conclusion.

Solution:

i)

$$L(\theta) = \Pi_{i=1}^n f(X_i) = \Pi_{i=1}^n \theta^{X_i} (1+\theta)^{-(X_i+1)} = \theta^{\sum_{i=1}^n X_i} \cdot \theta^{-\sum_{i=1}^n (X_i+1)}$$

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n X_i \cdot \ln(\theta) - \sum_{i=1}^n (X_i + 1) \ln(1+\theta)$$

$$\frac{dl(\theta)}{d\theta} = \frac{\sum_{i=1}^n X_i}{\theta} - \frac{\sum_{i=1}^n (X_i+1)}{1+\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

ii)Obtain $Var(X)$ as $Var(X) = E(X^2) - (E(X))^2$, using $E(X(X-1)) = E(X^2) - E(X) = 2\theta^2$, find that $E(X^2) = 2\theta^2 + \theta$, hence, $Var(X) = 2\theta^2 + \theta - \theta^2 = \theta^2 + \theta$.

$$Var(\hat{\theta}) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n(\theta^2 + \theta) = \frac{\theta^2 + \theta}{n}$$

To obtain CRLB, consider the likelihood function

$$L(\theta) = f(X) = \theta^X (1+\theta)^{-(X+1)}$$

$$l(\theta) = \ln L(\theta) = X \cdot \ln(\theta) - (X+1)\ln(1+\theta)$$
$$\frac{dl(\theta)}{d\theta} = \frac{X}{\theta} - \frac{(X+1)}{1+\theta}$$
$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{X}{\theta^2} + \frac{(X+1)}{(1+\theta)^2}$$

Fisher's Information:

$$E\left(\frac{d^2 l(\theta)}{d\theta^2}\right) = -\frac{E(X)}{\theta^2} + \frac{E(X+1)}{(1+\theta)^2} = -\frac{\theta}{\theta^2} + \frac{\theta+1}{(1+\theta)^2} = -\frac{1}{\theta} + \frac{1}{(1+\theta)} = \frac{-1}{\theta(\theta+1)}$$

Then CRLB is:

$$\left(-n\ E\left(\frac{d^2 l(\theta)}{d\theta^2}\right)\right)^{-1} = \frac{\theta(\theta+1)}{n}$$

Hence, the estimator $\hat{\theta} = \bar{X}$ is efficient estimator for θ.

2. [2 points] A total of 52 yearly measurements of snowfall in Buffalo, NY, are given in the table below. Assume that the standard deviation is 26 inches.

How large does n have to be to guarantee that the width of the 90% confidence interval for mean is less than 10?

Table 2. Annual snowfall, in inches, in Buffalo, NY over 52 years

126.4	77.8	82.4	79.3
78.1	89.6	51.1	85.5
90.9	58.0	76.2	120.7
104.5	110.5	87.4	65.4
110.5	39.9	25.0	40.1
69.3	88.7	53.5	71.4
39.8	83.0	63.6	55.9
46.7	89.9	72.9	84.6
74.6	105.2	83.6	113.7
80.7	124.7	60.3	114.5
79.0	115.6	74.4	49.6
54.7	71.8	49.1	103.9
51.6	82.4	83.6	115.6

Solution:

The confidence interval for μ is:

$$P\left(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}\right) = 0.90$$

From the table we can find that $z_{\alpha/2} = 1.65$, thus we can find the value for n as:

$$\bar{X} + 1.65 \cdot \sigma/\sqrt{n} - \bar{X} + 1.65 \cdot \sigma/\sqrt{n} < 10 \quad \Rightarrow$$

$$\sqrt{n} > \frac{3.30 \cdot \sigma}{10} \quad \Rightarrow \quad n > 73.62.$$

Thus, to construct a 90% confidence interval with the length less than 10 inches we need to have at least 74 observations.