MATH10282

Two hours

THE UNIVERSITY OF MANCHESTER

INTRODUCTION TO STATISTICS

20 May 2021

14.00 – 16.00

Answer ALL FOUR questions in Section A (40 marks in total). Answer TWO of the THREE questions in Section B (40 marks in total). If more than TWO questions from Section B are attempted, then credit will be given for the best TWO answers.

Electronic calculators may be used in accordance with the University regulations

---

SECTION A

Answer **ALL** four questions

A1. (a) Suppose that we have the following sample of observations

-1.3, -0.59, 0.1, -1.4, -0.22, -0.35, -0.76, -0.2, 0.41, 0.32

Compute the following

(i) sample mean;

[1 marks]

(ii) sample variance;

[1 marks]

(iii) sample median;

[1 marks]

(iv) sample first quartile;

[1 marks]

(v) sample third quartile;

[1 marks]

(vi) sample range.

[1 marks]

(b) Let $x_1, x_2, \ldots, x_n$ denote a data set and let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ denote the order statistics in ascending order. Show that

$$\text{sample third quartile} = \begin{cases} x_{(3m)} + \frac{3}{4}\left(x_{(3m+1)} - x_{(3m)}\right), & \text{if } n = 4m, \\ x_{(3m)}, & \text{if } n = 4m-1, \\ x_{(3m-1)} + \frac{1}{4}\left(x_{(3m)} - x_{(3m-1)}\right), & \text{if } n = 4m-2, \\ x_{(3m-2)} + \frac{1}{2}\left(x_{(3m-1)} - x_{(3m-2)}\right), & \text{if } n = 4m-3, \end{cases}$$

where m is an integer greater than or equal to 1.

[4 marks]

[Total 10 marks]

A2. (a) Suppose $\widehat{\theta}$ is an estimator of θ based on a random sample of size n. Define what is meant by the following:

(i) the bias of $\widehat{\theta}$ (written as bias $(\widehat{\theta})$).

[1 marks]

(ii) $\widehat{\theta}$ is an unbiased estimator of θ.

[1 marks]

(iii) $\widehat{\theta}$ is an asymptotically unbiased estimator of θ.

[1 marks]

(iv) the mean squared error of $\widehat{\theta}$ (written as MSE $(\widehat{\theta})$).

[1 marks]

(v) $\widehat{\theta}$ is a consistent estimator of θ.

[1 marks]

(b) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Exp (λ) distribution. Consider the following estimators for $\theta = 1/\lambda$: $\widehat{\theta_1} = (1/n) \sum_{i=1}^{n} X_i$ and $\widehat{\theta_2} = (1/(n+1)) \sum_{i=1}^{n} X_i$.

(i) Find the biases of $\widehat{\theta_1}$ and $\widehat{\theta_2}$.

[2 marks]

(ii) Find the mean squared errors of $\widehat{\theta_1}$ and $\widehat{\theta_2}$.

[3 marks]

[Total 10 marks]

A3. (a) Suppose we wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Define what is meant by the following:

(i) the Type I error of a test;

[1 marks]

(ii) the Type II error of a test;

[1 marks]

(iii) the significance level of a test.

[1 marks]

(b) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, where σ is unknown. State the rejection region for each of the following tests:

(i) $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$.

[1 marks]

(ii) $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma < \sigma_0$.

[1 marks]

In each case, assume a significance level of α.

(c) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, where σ is unknown. Derive an expression for $\mathrm{P}(\text{Reject } H_0 \mid H_1 \text{ is true})$ for the two cases in part (b):

(i) $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$.

[3 marks]

(ii) $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma < \sigma_0$.

[2 marks]

Please express the probability in terms of the distribution function of a chisquared random variable.

[Total 10 marks]

A4. (a) Let $\mathbf{X} = (X_1, \ldots, X_n)$, with $X_1, \ldots, X_n$ an independent random sample from a distribution F_X with unknown parameter θ. Let $I(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$ denote an interval estimator for θ. Define what is meant by the following:

(i) $I(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence interval,

[1 marks]

(ii) coverage probability of $I(\mathbf{X})$,

[1 marks]

(iii) coverage length of $I(\mathbf{X})$.

[1 marks]

(b) Suppose $X_1, \ldots, X_n$ are counts of defective items produced on n different days assumed to be a random sample. The number of defectives on any given day is modeled as a Poisson random variable with parameter λ, which is the unknown population mean defectives per day.

(i) Show that $E\left(\overline{X}\right) = \lambda$ and $Var\left(\overline{X}\right) = \lambda/n$.

[3 marks]

(ii) Assuming that $\left(\overline{X} - \lambda\right) / \left(\sqrt{\lambda/n}\right)$ has the standard normal distribution, derive a $100(1-\alpha)\%$ confidence interval for λ.

[4 marks]

[Total 10 marks]

SECTION B

Answer **TWO** of the three questions

B5. Let X and Y be uncorrelated random variables. Suppose that X has mean 2θ and variance 4. Suppose that Y has mean θ and variance 2. The parameter θ is unknown.

(i) Compute the bias and mean squared error for each of the following two estimators of θ: $\widehat{\theta_1} = (1/4)X + (1/2)Y$ and $\widehat{\theta_2} = X - Y$.

[8 marks]

(ii) Which of the two estimators ($\widehat{\theta_1}$ or $\widehat{\theta_2}$) is better and why?

[4 marks]

(iii) Verify that the estimator $\widehat{\theta_c} = (c/2)X + (1-c)Y$ is unbiased. Find the value of c which minimizes Var $(\widehat{\theta_c})$.

[8 marks]

[Total 20 marks]

B6. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with the common probability mass function (pmf):

$$p(x) = \theta(1-\theta)^{x-1}$$

for $x = 1, 2, \ldots$ and $0 < \theta < 1$.

(i) Write down the likelihood function of θ.

[4 marks]

(ii) Find the maximum likelihood estimator of θ.

[4 marks]

(iii) Find the maximum likelihood estimator of $\psi = 1/\theta$.

[4 marks]

(iv) Determine the bias, variance and the mean squared error of the maximum likelihood estimator of ψ.

[4 marks]

(v) Is the maximum likelihood estimator of ψ unbiased? Is it consistent?

[4 marks]

[Total 20 marks]

B7. Consider the linear regression model with zero intercept

$$Y_i = \beta X_i + e_i$$

for $i = 1, 2, \ldots, n$, where $e_1, e_2, \ldots, e_n$ are independent and identical normal random variables with zero mean and variance σ^2 assumed known. Moreover, suppose $X_1, X_2, \ldots, X_n$ are known constants taking distinct values.

(i) Write down the likelihood function of β.

[4 marks]

(ii) Show that the maximum likelihood estimator of β is

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

[4 marks]

(iii) Find the bias of the estimator in part (ii). Is the estimator unbiased?

[4 marks]

(iv) Find the mean square error of the estimator in part (ii).

[4 marks]

(v) Find the exact distribution of the estimator in part (ii).

[4 marks]

[Total 20 marks]

END OF EXAMINATION PAPER