Two hours

Statistical tables to be provided

THE UNIVERSITY OF MANCHESTER

INTRODUCTION TO STATISTICS

20 May 2019

14.00 – 16.00

Answer **ALL FOUR** questions in Section A (10 marks each) and **TWO** of the **THREE** questions in Section B (20 marks each). 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

1.75, 0.77, 0.39, 0.34, 0.07, 0.58, -0.54, -2.78, 0.75, -0.68

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]

For computing the quartiles, use the fact that the pth quartile is

$$Q(p) = x_{(r')} + \left\{ p(n+1) - r' \right\} \left\{ x_{(r'+1)} - x_{(r')} \right\},$$

where $r' = [p(n+1)]$ and n is the sample size.

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

[1 marks]

(ii) the bias of $\widehat{\theta}$ (written as $\text{bias}(\widehat{\theta})$);

[1 marks]

(iii) the mean squared error of $\widehat{\theta}$ (written as $\text{MSE}(\widehat{\theta})$);

[1 marks]

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

[1 marks]

(b) Suppose $X_1, \ldots, X_n$ are independent $\text{Uniform}(0, \theta)$ random variables. Let $\widehat{\theta} = \max(X_1, \ldots, X_n)$ denote a possible estimator of θ.

(i) Derive the bias of $\widehat{\theta}$;

[3 marks]

(ii) Derive the mean squared error of $\widehat{\theta}$;

[1 marks]

(iii) Is $\widehat{\theta}$ an unbiased estimator for θ? Justify your answer;

[1 marks]

(iv) Is $\widehat{\theta}$ a consistent estimator for θ? Justify your answer.

[1 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) $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

[1 marks]

(ii) $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.

[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) $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

[3 marks]

(ii) $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.

[2 marks]

You may express the probability in terms of the distribution function of a Student't t 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) that $I(\mathbf{X})$ 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 $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$. Derive a $100(1-\alpha)\%$ confidence interval for μ if

(i) σ is known.

[1 marks]

(ii) σ is not known.

[1 marks]

(c) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from Uniform $[0, a]$.

(i) Show that the cumulative distribution function $\min(X_1, X_2, \ldots, X_n) = Z$ say, is

$$F_Z(z) = 1 - \left[1 - \frac{z}{a}\right]^n$$

for $0 < z < a$.

[2 marks]

(ii) Use the result in (i) to derive a $100(1-\alpha)\%$ confidence interval for a.

[3 marks]

[Total 10 marks]

SECTION B

Answer **TWO** of the three questions

B5. Suppose X_1, X_2 and X_3 are independent $\text{Exp}(1/\lambda)$ random variables. Let $\widehat{\theta}_1 = a\,(X_1 + X_2 + X_3)$ and $\widehat{\theta}_2 = b\sqrt{X_1 X_2}$ denote possible estimators of λ, where a and b are constants.

(i) Show that $a = 1/3$ if $\widehat{\theta}_1$ is to be an unbiased estimator of λ;

[4 marks]

(ii) Show that $b = 4/\pi$ if $\widehat{\theta}_2$ is to be an unbiased estimator of λ;

[4 marks]

(iii) Determine the variance of $\widehat{\theta}_1$;

[4 marks]

(iv) Determine the variance of $\widehat{\theta}_2$;

[4 marks]

(v) Which of the estimators ($\widehat{\theta}_1$ and $\widehat{\theta}_2$) is better with respect to mean squared error and why?

[4 marks]

[Total 20 marks]

B6. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution specified by the probability density function

$$f_X(x) = \frac{x^2}{\sigma^3} \exp\left(-\frac{x^3}{3\sigma^3}\right)$$

for $x > 0$.

(i) Write down the likelihood function of σ^3.

[4 marks]

(ii) Show that the maximum likelihood estimator of σ^3 is

$$\widehat{\sigma^3} = \frac{1}{3n} \sum_{i=1}^{n} X_i^3.$$

[4 marks]

(iii) Deduce the maximum likelihood estimator of σ.

[4 marks]

(iv) Show that the estimator in part (ii) is an unbiased estimator of σ^3.

[4 marks]

(v) Show that the estimator in part (ii) is a consistent estimator of σ^3.

[4 marks]

[Total 20 marks]

B7. An electrical circuit consists of four batteries connected in series to a lightbulb. We model the battery lifetimes X_1, X_2, X_3 as independent and identically distributed Uniform $(0, \theta)$ random variables. Our experiment to measure the operating time of the circuit is stopped when any one of the batteries fails. Hence, the only random variable we observe is $Y = \min(X_1, X_2, X_3)$.

(i) Determine the cumulative distribution function of the random variable Y.

[4 marks]

(ii) Write down the likelihood function of θ based on a single observation of Y.

[4 marks]

(iii) Derive the maximum likelihood estimator of θ.

[4 marks]

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

[4 marks]

(v) Find the mean squared error of the estimator in part (iii).

[4 marks]

[Total 20 marks]

END OF EXAMINATION PAPER