Two hours

Statistical tables to be provided

THE UNIVERSITY OF MANCHESTER

INTRODUCTION TO STATISTICS

20 May 2020

14.00 – 16.00

Answer **ALL FOUR** questions in Section A and **TWO** of the **THREE** questions in Section B.

Electronic calculators may be used in accordance with the University regulations

---

SECTION A

Answer **ALL** four questions

A1. (a) 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. Define what is meant by the following:

(i) sample median;

(ii) sample first quartile;

(iii) sample third quartile;

(iv) sample inter quartile range.

(b) Show that

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

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

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 $\text{bias}(\widehat{\theta})$);

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

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

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

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

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

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

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

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;

(ii) the Type II error of a test;

(iii) the significance level of a test.

(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 $H_1 : \mu \neq \mu_0$.

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

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 $H_1 : \mu \neq \mu_0$.

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

You may express the probability in terms of the distribution function of a Student't t random variable.

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.

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

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

(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 σ.

(c) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution specified by the cumulative distribution function

$$F_X(x) = 1 - \exp(\theta - x)$$

for $x > \theta$.

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

$$F_Z(z) = 1 - \exp(n\theta - nz)$$

for $z > \theta$.

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

SECTION B

Answer **TWO** of the three questions

B5. Suppose $X \sim \text{Binomial}(m, p)$ and $Y \sim \text{Binomial}(n, p)$ are independent random variables. Consider the following estimators for p:

$$\widehat{p_1} = \frac{X}{2m} + \frac{Y}{2n}$$

and

$$\widehat{p_2} = \frac{X+Y}{m+n}.$$

(i) Calculate the bias of $\widehat{p_1}$.

(ii) Calculate the bias of $\widehat{p_2}$.

(iii) Calculate the mean squared error of $\widehat{p_1}$.

(iv) Calculate the mean squared error of $\widehat{p_2}$.

(v) Which of the estimators ($\widehat{p_1}$ and $\widehat{p_2}$) is better with respect to bias and why?

(vi) Which of the estimators ($\widehat{p_1}$ and $\widehat{p_2}$) is better with respect to mean squared error and why?

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

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

for $x > 0$.

(i) Write down the likelihood function of σ^2.

(ii) Show that the maximum likelihood estimator of σ^2 is

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

(iii) Deduce the maximum likelihood estimator of σ.

(iv) Show that the estimator in part (ii) is an unbiased estimator of σ^2.

(v) Show that the estimator in part (ii) is a consistent estimator of σ^2.

B7. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution specified by the probability mass function

$$p_X(x) = \binom{x+r-1}{x}(1-p)^r p^x$$

for $x = 0, 1, \ldots$ with the properties

$$E(X) = \frac{pr}{1-p}$$

and

$$\text{Var}(X) = \frac{pr}{(1-p)^2}.$$

Furthermore, assume r is known but p is unknown.

(i) Write down the likelihood function of p.

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

$$\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{nr + \sum_{i=1}^{n} X_i}.$$

(iii) Deduce the maximum likelihood estimator of $p/(1-p) = \psi$ say.

(iv) Show that the estimator in part (iii) is an unbiased estimator of ψ.

(v) Show that the estimator in part (iii) is a consistent estimator of ψ.

END OF EXAMINATION PAPER