UNIVERSITY OF LEICESTER

MA1202
All candidates

Semester 2 Examinations 2021

DO NOT OPEN THE QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY THE CHIEF INVIGILATOR	
School	Mathematics and Actuarial Science
Module Code	MA1202
Module Title	Introductory Statistics
Exam Duration	4 hours + 45 mins upload time
CHECK YOU HAVE THE CORRECT QUESTION PAPER	
Number of Pages	3
Number of Questions	3
Instructions to Candidates	All questions carry equal weight. Answer all questions.
FOR THIS EXAM YOU ARE ALLOWED TO USE THE FOLLOWING:	
Calculators	Yes
Books/Statutes provided by the University	No
Are students permitted to bring their own Books/Statutes/Notes?	Yes
Additional Stationery	Yes

1. *Let $X_1, X_2, ..., X_n$ be a random sample of n iid random variables from a distribution with probability density function:*

$$f_X(x,\gamma) = \frac{1}{B\gamma} e^{-x/(B\gamma)}, \quad x \geq 0,$$

where $\gamma > 0$ is a parameter, and $B > 0$ is a constant.

Using the formulae for the first two population moments $E(X) = B\gamma$, $E(X^2) = 2B^2\gamma^2$, answer the following questions:

a) **[10 marks]** Consider the estimator $\hat{\gamma} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Is it a sufficient estimator for γ? Justify your answer.

b) **[5 marks]** Consider the estimator $\hat{\gamma}_\alpha = \alpha\bar{X}$. For what value of α is the estimator $\hat{\gamma}_\alpha$ unbiased for γ? Justify your answer.

c) **[10 marks]** For the value α from part $b)$, is the estimator $\hat{\gamma}_\alpha$ an efficient estimator for γ? Justify your answer.

d) **[5 marks]** Provide the definition of consistent estimator and explain its meaning. Is estimator $\hat{\gamma}_\alpha$ a consistent estimator? Justify your answer.

Total: 30 marks

2. *Let $X_1, X_2, ..., X_n$ be a random sample of n iid random variables from $N(\mu_X, \sigma^2)$ distribution and let $Y_1, ..., Y_m$ be a random sample of m iid random variables from $N(\mu_Y, \sigma^2)$ distribution. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ be the corresponding sample means.*

Answer the following questions:

a) **[10 marks]** If $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$ are the respective sample variances, show that

   i) the sample variance (for example, $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$) is an unbiased estimator for σ^2;

   ii) the pooled estimator

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{(n+m-2)}$$

   is also unbiased for σ^2.

b) **[5 marks]** Determine the distribution the following random variable :

$$D = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

and state all assumptions that need to be satisfied.

c) **[10 marks]** Two random samples of 16 male professors and 13 female professors from a large university have a sample mean salary for male professors of £58200 with a standard deviation of £3600 and a sample mean salary for female professors of £55600 with a standard deviation of £3500. Assume that the salaries of male and female professors are both normally distributed with equal standard deviations.

   i) At 5% significance level test the hypothesis that there is no difference in mean salaries between male and female professors against the alternative that there is a difference.

ii) Calculate the corresponding p-value to support your conclusion.

iii) Comment on your results.

d) **[5 marks]** Explain how to estimate the confidence interval for the difference of two population means using the large random samples. Please state all relevant theorems.

Total: 30 marks

3. *The data given below in Table 1 are 40 long-duration monthly average river discharge measurements, D, in cumecs (cubic metres per second) of a river basin system in the UK from 2010–13.*

Table 1. Monthly average river discharge measurements, D, in cumecs

11.825	12.75	13.38	14.575	14.95	14.95	15.525	15.825	16.25	16.275
16.575	16.575	17.15	17.475	18.20	18.35	18.525	18.525	18.9	19.2
19.325	19.52	19.55	19.8	20.325	20.325	20.4	20.475	20.775	21.3
21.3	21.925	22.425	22.65	22.8	23.85	24.375	25.9	26.4	27.675

Answer the following questions:

a) **[7 marks]** Let $Y_1, Y_2, ..., Y_n$ be a sample of n iid random variables with $N(\mu, \sigma^2)$ distribution, show that the maximum-likelihood estimator (MLE) for the population mean, if σ^2 is assumed to be known, is equal to $\frac{1}{n}\sum_{i=1}^{n} Y_i$.

b) **[5 marks]** Given that, for the data in Table 1,

$$\sum_{i=1}^{n} Y_i = 766.875 \text{ and} \sum_{i=1}^{n} Y_i^2 = 15232.764,$$

calculate the sample mean and the sample standard deviation (unbiased).

c) **[3 marks]** Construct a frequency table for the data, taking the 6 grouping intervals:

$$(0,15],\ (15,17],\ (17,19],\ (19,21],\ (21,23],\ (23,\infty).$$

d) **[15 marks]** Test the hypothesis, at 5% significance level, that the river discharge measurements have normal distribution, if the sample mean and the sample standard deviation equals to the values calculated in part b):

i) using z-transform: $Z = \frac{X - \bar{x}}{s_x}$, find the lower and upper boundaries for each range for z;

ii) find the expected frequencies using table for cumulative standard normal distribution;

iii) find the value of the appropriate test statistics;

iv) what is your conclusion? Comment on your answer.

Total: 30 marks

[Total 90 marks]