Strathmore UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)
MASTERS OF SCIENCE IN STATISTICAL SCIENCES
END OF SEMESTER EXAMINATION
STA 8102: STATISTICAL INFERENCE

DATE: 6th December, 2023 **TIME**: 3 Hours

INSTRUCTIONS

1. This examination consists of **FOUR** questions.
2. Answer Question **ONE (COMPULSORY)** and any other **TWO** questions.
3. You may use a **SIMPLE CALCULATOR**. No **MOBILE PHONES** in the exams room.

Question One (30 Marks)

(i) Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations

X	0	1	2	3
P(X)	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{(1-\theta)}{3}$

were taken from such a distribution: (3,0,2,1,3,2,1,0,2,1). What is the maximum likelihood estimate of θ. (3 marks)

(ii) Let $(X_1, \ldots, X_n)$ be a simple random sample of a random variable X with Exp(α) distribution. Is $\bar{X}$ is sufficient for α? (3 marks)

(iii) A physician is interested in the proportion of men that smoke and develop lung cancer. The physician wants to select a sample of smokers and observe whether they develop cancer or not. What has to be the sample size so that with a 95% probability the difference between the sample proportion and the true proportion is less than 0.02? (3 marks)

(iv) Assuming that $X_i \sim N(\mu, \sigma^2)$

(a) Which of the statistics below are unbiased estimators of μ? (4 marks)

$$(i) \quad \hat{\mu}_1 = \frac{X_1 + X_2 + X_3 + X_4}{4} \quad (ii) \quad \hat{\mu}_2 = \frac{2(X_1 + X_2)}{6} + \frac{X_3 + X_4}{6}$$
$$(iii) \quad \hat{\mu}_3 = \frac{X_1 - X_2 + X_3 - X_4}{4}.$$

(b) Among all the unbiased estimators in (a), which one is the most efficient? Which one is the most consistent among all the three estimators? (7 marks)

(v) Using the identity (3 marks)

$$(\hat{\theta} - \theta) = \left(\hat{\theta} - E[\hat{\theta}]\right) + \left(E[\hat{\theta}] - \theta\right) = \left(\hat{\theta} - E[\hat{\theta}]\right)$$

show that

$$\mathrm{MSE}[\hat{\theta}] = E\left[(\hat{\theta} - \theta)^2\right] = \mathrm{Var}[\hat{\theta}] + \left(\mathrm{Bias}[\hat{\theta}]\right)^2$$

(vi) One source of water pollution is gasoline leakage from underground storage tanks. In Mombasa, a random sample of $n = 74$ gasoline stations is selected and the tanks are inspected; 10 stations are found to have at least one leaking tank. Calculate a 95 percent confidence interval for p, the population proportion of gasoline stations with at least one leaking tank. (4 marks)

(vii) A university has found over the years that out of all the students who are offered admission, the proportion who accept is 0.70. After a new director of admissions is hired, the university wants to check if the proportion of students accepting has changed significantly. Suppose they offer admission to 1200 students and 888 accept. Is this evidence at the $\alpha = .05$ level that there has been a real change from the status quo? How about at the 0.02 level? (3 marks)

Question Two (15 Marks)

(i) Let $X_1, X_2, \ldots, X_n$ be gamma random variables with parameters α and θ so that the probability density function is:

$$f(x_i) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

what are the method of moments estimators of α and θ? (7 marks)

(ii) Let $(X_1, \ldots, X_n)$ be a simple random sample of a random variable with exponential distribution of parameter $1/\theta$, that is, with p.d.f.

$$f(x;\theta) = \begin{cases} \theta e^{-\theta x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Obtain the sampling distribution of the sum statistic $T(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$. (3 marks)

(iii) Let $(X_1, \ldots, X_n)$ be a simple random sample of $X \sim \mathrm{Pois}(\lambda)$, that is, with p.m.f.

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \ldots, \quad \lambda > 0.$$

Consider the parameter $\theta = p(0;\lambda) = e^{-\lambda}$. Obtain the Uniformly minimum-variance unbiased estimator (UMVUE) of θ. (5 marks)

Question Three (15 Marks)

(i) Consider a radar system that uses radio waves to detect aircraft. The system receives a signal and, based on the received signal, it needs to decide whether an aircraft is present or not. Let X be the received signal. Suppose that we know

$$X = W, \qquad \text{if no aircraft is present.}$$
$$X = 1 + W, \qquad \text{if an aircraft is present.}$$

where $W \sim N(0, \sigma^2 = \frac{1}{9})$. Thus, we can write $X = \theta + W$, where $\theta = 0$ if there is no aircraft, and $\theta = 1$ if there is an aircraft. Suppose that we define H_0 and H_1 as follows:

$$H_0(\text{null hypothesis}): \text{ No aircraft is present.}$$
$$H_1(\text{alternative hypothesis}): \text{An aircraft is present}$$

Design a level 0.05 test ($\alpha = 0.05$) to decide between H_0 and H_1. Use likelihood ratio test (approach). (9 marks)

(ii) Find the probability of type II error, β, for the above test. Note that this is the probability of missing a present aircraft. (6 marks)

Question Four (15 Marks)

(i) A new training method for an assembly operation is being tested at a factory. For that purpose, two groups of nine employees were trained during three weeks. One group was trained with the usual procedure and the other with the new method. The assembly time (in minutes) that each employee required after the training period is collected in the table below. Assuming that the assembly times are normally distributed with both variances equal to 22 minutes, obtain a confidence interval at level 0.05 for the difference of average assembly times for the two kinds of training. (4 marks)

Procedure	Measurements								
Standard	32	37	35	28	41	44	35	31	34
New	35	31	29	25	34	40	27	32	31

(ii) With the current level of communication resources for an online bookstore and their projected growth over the next 6 months, a company will be able to provide satisfactory service if the average connection time per customer is no more than 13.5 minutes. Based on a random sample of 45 connections yielding s a sample mean of 15.3 minutes with a sample standard deviation of 6.7 minutes, would you recommend that the company upgrades their communication resources? (Perform a one-sided test at a 5% significance level.) (6 marks)

(iii) An airline wants to evaluate the depth perception of its pilots over the age of 50. A random sample of $n = 14$ airline pilots over the age of 50 are asked to judge the distance between two markers placed 20 feet apart at the opposite end of the laboratory. The sample data listed here are the pilots' error (recorded in feet) in judging the distance.

2.7	2.4	1.9	2.6	2.4	1.9	2.3
2.2	2.5	2.3	1.8	2.5	2.0	2.2

Use the sample data to test the hypothesis that the average error in depth perception for the company's pilots over the age of 50 is 2.00 at $\alpha = 0.05$ confidence level on μ. (5 marks)

END