STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)

MASTERS OF SCIENCE IN STATISTICAL SCIENCES

END OF SEMESTER EXAMINATION

STA 8102: STATISTICAL INFERENCE

DATE: 9th January, 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) Explain the properties of point estimators. (3 marks)

(ii) 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

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

(iii) Let $(X_1, \cdots, X_n)$ be a simple random sample of a random variable with $N(\mu, \sigma^2)$ distribution. Find the sufficient statistics for σ^2, if μ is known and μ if σ^2 is known and μ and σ^2. (6 marks)

(iv) Explain and give examples of three types of random sampling. (6 marks)

(v) 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)

(vi) The summary statistics given below from two catalysts types in which 8 samples in the pilot plant are take from each are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, the 1st catalyst is currently in use, but the 2nd catalyst is acceptable.

Observation number	Catalyst 1	Catalyst 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75

Construct a confidence interval for the ratio variance of yields. Use $\alpha = 0.05$. (4 marks)

(vii) 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)

Question Two (15 Marks)

(i) Explain five steps that must be performed in an expression that defines the Cramer-Rao lower bound (CRLB). (5 marks)

(ii) Show that $\bar{x}$ is a uniformly minimum variance unbiased estimator for independent Bernoulli random variable with unknown success probability θ. (3 marks)

(iii) Assuming that $X_i \sim N(\mu, \sigma^2)$, which of the statistics below are unbiased estimators of μ?

$$(a)\quad \hat{\mu}_1 = \frac{X_1 + X_2 + X_3 + X_4}{4} \qquad (b)\quad \hat{\mu}_2 = \frac{2(X_1 + X_2)}{6} + \frac{X_3 + X_4}{6} \qquad (c)\quad \hat{\mu}_3 = \frac{X_1 - X_2 + X_3 - X_4}{4}.$$

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

Question Three (15 Marks)

(i) In certain food experiment to compare two types of baby foods A and B, the following results of increase in weight (lbs) we observed in 8 children as follows.

Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	53	50	54	54	53

Examine the significance of increase in weight of children due to food B. (4 marks)

(ii) 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(null hypothesis) : No aircraft is present.

H_1(alternative hypothesis) : An aircraft is present

(a) Write the null hypothesis, H_0, and the alternative hypothesis, H_1, in terms of possible values of the parameter θ. (2 marks)

(b) Design a level 0.05 test ($\alpha = 0.05$) to decide between H_0 and H_1. Use likelihood ratio test (approach). (6 marks)

(c) Find the probability of type II error, β, for the above test. Note that this is the probability of missing a present aircraft. (3 marks)

Question Four (15 Marks)

(i) Suppose the samples are simple random samples taken from normal populations. Calculate the confidence intervals for the ratio of the two population variances and the ratio of standard deviations given the following information. (4 marks)

$$\alpha = 0.05; \quad n_1 = 30; \quad s_1 = 16.37; \quad n_2 = 39; \quad s_2 = 9.88.$$

(ii) A gunpowder manufacturer developed a new formula that was tested in eight bullets. The resultant initial velocities, measured in feet per second, were

3005, 2925, 2935, 2968, 2995, 3005, 2937, 2905.

Assuming that the initial velocities have normal distribution with $\sigma = 39$ feet per second, find a confidence interval at level $\alpha = 0.05$ for the initial mean velocity of the bullets that employ the new gunpowder. (3 marks)

(iii) A random sample of 20 nominally measured 2mm diameter steel ball bearings is taken and the diameters are measured precisely. The measurements, in mm, are as follows:

2.02, 1.94, 2.09, 1.95, 1.98, 2.00, 2.03, 2.04, 2.08, 2.07
1.99, 1.96, 1.99, 1.95, 1.99, 1.99, 2.03, 2.05, 2.01, 2.03.

Assuming that the diameters are normally distributed with unknown mean, μ, and unknown variance, σ^2, find a two-sided 95% confidence interval for the variance, σ^2. (4 marks)

(iv) Acute exposure to cadmium produces respiratory distress and kidney and liver damage (and possibly death). For this reason, the level of airborne cadmium dust and cadmiumoxide fume in the air, denoted by X (measured in milligrams of cadmium per m^3 of air), is closely monitored. A random sample of $n = 35$ measurements from a large factory are given below:

0.044,	0.030,	0.052,	0.044,	0.046,	0.020,	0.066,	0.052,	0.049,	0.030,	0.040,
0.045,	0.039,	0.039,	0.039,	0.057,	0.050,	0.056,	0.061,	0.042,	0.055,	0.037,
0.062,	0.062,	0.070,	0.061,	0.061,	0.058,	0.053,	0.060,	0.047,	0.051,	0.054,
0.042,	0.051.									

Find a 99 percent confidence interval for μ, the mean level of airborne cadmium. (4 marks)

END