



**NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY**

FACULTY OF HEALTH AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

| | |
|---|--------------------------------|
| QUALIFICATION: BACHELOR OF SCIENCE HONOURS IN APPLIED STATISTICS | |
| QUALIFICATION CODE: 08BSHS | LEVEL: 8 |
| COURSE CODE: SAT802S | COURSE: SAMPLING THEORY |
| SESSION: November 2022 | PAPER : THEORY |
| DURATION: 3 Hours | MARKS: 100 |

| | |
|---|------------------|
| FIRST OPPORTUNITY EXAMINATION QUESTION PAPER | |
| EXAMINER | Mr. J. J. SWARTZ |
| MODERATOR: | Dr. I. NEEMA |

| |
|---|
| INSTRUCTIONS |
| <ol style="list-style-type: none">1. Answer ALL the questions in the booklet provided.2. Show clearly all the steps used in the calculations.3. All written work must be done in blue or black ink and sketches must be done in pencil. |

PERMISSIBLE MATERIALS

1. Calculator
2. Pen and Clean Paper for calculations

THIS QUESTION PAPER CONSISTS OF 4 PAGES (Including this front page)

Question 1 [25 marks]

1.1 Write a short description on the importance of the normal distribution in sampling theory. [4]

1.2 Provide six basic steps in developing a sampling plan. [6]

1.3 For the 200 managers and 800 engineers of a corporation, the standard deviations of the number of days a year spent on research were presumed to be 30 and 60 days, respectively. Find the sample size needed for proportional allocation to estimate the population mean with the S.E. of the estimator not exceeding 10 and its allocation for the two groups. [5]

1.4 Among 100 Retailers in Namibia, the average of employee sizes for the largest 10 and smallest 10 corporations were known to be 300 and 100, respectively. For a sample of 20 from the remaining 80 retailers, the mean and standard deviation were 250 and 110, respectively. For the total employee size of the 80 retailers, find the

1.4.1 Estimate for the total, [2]

1.4.2 S.E. of the estimate, and [3]

1.4.3 95% confidence limits. [5]

Question 2 [25 marks]

2.1. The Ministry of Health and Social Services (MoHSS) wants to estimate the rate of incidence of respiratory disorders among the middle-aged male and female smokers in Namibia. How large a sample should be taken to be 95% confident that the error of estimation of the proportion of the population with such disorders does not exceed 0.05? The true value of p is expected to be near 0.30. [5]

2.2.

We propose to estimate the mean \bar{Y} of a characteristic y by way of a sample selected according to a simple random design without replacement of size 1000 in a population of size 1 000 000. We know the mean $\bar{X} = 15$ of an auxiliary characteristic x . We have the following results:

$$s_y^2 = 20, s_x^2 = 25, s_{xy} = 15, \hat{X} = 14, \hat{Y} = 10$$

2.2.1. Estimate \bar{Y} by way of Horwitz – Thomson, difference, ratio and regression estimators. Estimate the variances of these estimators. [15]

2.2.2. Which estimator should we choose to estimate \bar{Y} ? [5]

Question 3 [25 marks]

3.1. The Namibian 25, 2001, summarized the results of a survey conducted by Yellow Express on 2000 lawyers on sexual advances in the office. Between 85 and 98% responded to the questions in the survey; 49% of the responding women and 9% of the responding men agreed that some sorts of harassment exist in the offices. Assume that the population of lawyers is large and there are equal numbers of female and male lawyers, and ignore the nonresponse; that is, consider the respondents to be a random sample of the 2000 lawyers.

3.1.1 Find the standard errors for the above percentages. [5]

3.2. A forest resource manager is interested in estimating the total number of dead trees in a 400 acre area of heavy infestation. She subdivides the area into 200 plots of equal sizes and uses photo counts to find the number of dead trees in 18 randomly sampled plots. She then randomly samples 8 plots out of these 18 plots and conducts a ground count on these 8 plots. Let x denote the number of dead trees in the plot by photo count and y the number of dead trees by ground count. The data are given as:

| Plot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|------|---|---|----|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| x' | 5 | 7 | 10 | 6 | 7 | 9 | 3 | 6 | 8 | 11 | 5 | 9 | 12 | 13 | 3 | 20 | 15 | 4 |

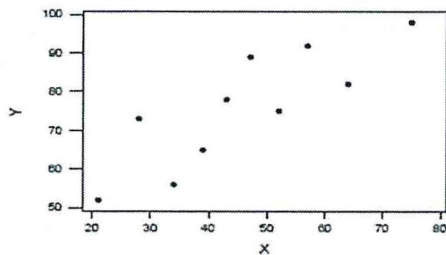
Out of these 18 plots, 8 are randomly selected and a ground count is conducted.

| Plot | 2 | 3 | 5 | 6 | 12 | 15 | 16 | 17 |
|--------|--------|--------|--------|---------|---------|--------|--------|---------|
| x | 7 | 10 | 7 | 9 | 9 | 3 | 20 | 15 |
| y | 9 | 13 | 10 | 11 | 10 | 4 | 25 | 17 |
| $y-rx$ | 0.3375 | 0.6250 | 1.3375 | -0.1375 | -1.1375 | 0.2875 | 0.2500 | -1.5625 |

- 3.2.1 Estimate the total number of dead trees in the 400-acre area. [6]
 3.2.2 Compute the ratio estimate for the population total. [6]
 3.2.3 Compute the estimated variance of the ratio estimator [8]

Question 4 [25 marks]

4.1 A mathematics achievement test was given to 486 students prior to entering a certain college who then took a calculus class. A simple random sampling of 10 students are selected and their calculus score recorded. It is known that the average achievement test score for the 486 students was 52. The scatterplot of the 10 samples are given below and the data follow.



The scatter plot shows that there is a strong positive linear relationship.

$$\hat{\mu}_L = \bar{y} + b(\mu_x - \bar{x}) = a + b\mu_x$$

| Student | Achievement test score X | Calculus score Y |
|---------|--------------------------|------------------|
| 1 | 39 | 65 |
| 2 | 43 | 78 |
| 3 | 21 | 52 |
| 4 | 64 | 82 |
| 5 | 57 | 92 |
| 6 | 47 | 89 |
| 7 | 28 | 73 |
| 8 | 75 | 98 |
| 9 | 34 | 56 |
| 10 | 52 | 75 |

The regression equation is
 $Y = 40.8 + 0.766 X$

| Predictor | Coef | StDev | T | P |
|-----------|--------|--------|------|-------|
| Constant | 40.784 | 8.507 | 4.79 | 0.001 |
| X | 0.7656 | 0.1750 | 4.38 | 0.002 |

S = 8.704 R-Sq = 70.5% R-Sq(adj) = 66.8%

| Analysis of Variance | | | | | |
|----------------------|----|--------|--------|-------|-------|
| Source | DF | SS | MS | F | P |
| Regression | 1 | 1450.0 | 1450.0 | 19.14 | 0.002 |
| Resid. Err | 8 | 606.0 | 75.8 | | |
| Total | 9 | 2056.0 | | | |

4.1.1 Using the results from the output above, calculate the regression estimate. [3]

4.1.2 What is the variance of the regression estimate? [5]

4.1.3 Calculate the approximate 95% CI for μ . [7]

4.2 A population of 20000 farms were divided into 30 clusters. Sample 3000 farms from 10 clusters using Probability Proportional to size (PPS) by completing the table below in your answer sheet. [10]

| A | B | C | D | E | F | G | H |
|---------|------|----------------|------------------|--------|-------------------------|--------|----------------|
| Cluster | Size | Cumulative sum | Clusters sampled | Prob 1 | Individuals per cluster | Prob 2 | Overall weight |
| 1 | 1028 | | 905 | | 300 | | |
| 2 | 555 | | | | | | |
| 3 | 390 | | | | | | |
| 4 | 1309 | | 2905 | | 300 | | |
| 5 | 698 | | | | | | |
| 6 | 907 | | | | | | |
| 7 | 432 | | 4905 | | 300 | | |
| 8 | 897 | | | | | | |
| 9 | 677 | | | | | | |
| 10 | 501 | | 6905 | | 300 | | |
| 11 | 867 | | | | | | |
| 12 | 867 | | 8905 | | 300 | | |
| 13 | 1002 | | | | | | |
| 14 | 1094 | | 10905 | | 300 | | |
| 15 | 668 | | | | | | |
| 16 | 500 | | | | | | |
| 17 | 835 | | 12905 | | 300 | | |
| 18 | 396 | | | | | | |
| 19 | 630 | | | | | | |
| 20 | 483 | | | | | | |
| 21 | 319 | | 14905 | | 300 | | |
| 22 | 569 | | | | | | |
| 23 | 987 | | | | | | |
| 24 | 598 | | 16905 | | 300 | | |
| 25 | 375 | | | | | | |
| 26 | 387 | | | | | | |
| 27 | 465 | | | | | | |
| 28 | 751 | | 18905 | | 300 | | |
| 29 | 365 | | | | | | |
| 30 | 448 | 20000 | | | | | |

*****END OF EXAMINATION*****