



MACHAKOS UNIVERSITY

University Examinations 2019/2020

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

FOURTH YEAR SPECIAL/SUPPLEMENTARY EXAMINATION FOR
BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)
SST 408: DESIGN AND ANALYSIS OF SAMPLE SURVEY II

DATE: 22/1/2021

TIME: 2.00-4.00 PM

INSTRUCTIONS:

Answer question ONE and any other TWO questions

QUESTION ONE (30 MARKS)

- a) Give five elements of a sample survey design (5 marks)
- b) What is panel design as used in sample survey? (2 marks)
- c) Discuss the two types of non-response. (4 marks)
- d) Briefly explain four possible means of handling non-coverage error in household surveys. (4 marks)
- e) Define the term imputation as a way of dealing with non-response and give three types of imputation (5 marks)
- f) Give three variants on the prospective panel design (3 marks)
- g) A research was conducted to estimate prevalence of psychiatric disorders. Participants in a phase I sample were given the General Health Questionnaire (GHQ) and classified into three strata based on their GHQ score. The stratification was used to take a stratified random sample of 250 persons for the phase II sample; the Composite International Diagnostic Interview (CIDI), considered to be a more accurate diagnostic tool, was administered to each person in the phase II sample. The CIDI score was used to classify the phase II sample members as having at least one psychiatric disorder (case) or having no psychiatric disorder (non-case).

The results are given in the following table. The counts of cases and non-cases are from the phase II sample.

Stratum	n_h	m_h	Non-case	Case
GHQ ≤ 3 (low)	1049	60	33	27
GHQ=4, 5 (medium)	237	48	14	34
GHQ ≥ 6 (high)	272	142	23	119
Total	1558	250		

Calculate the Phase II sampling weight $w_i^{(2)}$ for each stratum. (4 marks)

h) Give three considerations when designing a household questionnaire (3 marks)

SECTION B: ANSWER ANY TWO QUESTIONS IN THIS SECTION

QUESTION TWO (20 MARKS)

- a) Differentiate between Multiphase and Multistage Sampling. (4 marks)
- b) Explain any four challenges in using Multiphase Design (8 marks)
- c) Give any quantities that are held fixed and that are allowed to be random under the model-based approach (5 marks)
- d) What are the three advantages of replication method as sampling method (3 marks)

QUESTION THREE (20 MARKS)

- a) Give three factors that affect the sample survey design (5 marks)
- b) Explain brief any four types of household surveys that you know (4 marks)
- c) A linguist wants to study the depth and breadth of language acquisition between ages 3 and 15. However, the linguist does not have 12 years to conduct the investigation. The linguist could recruit four panels of children who are 3, 5, 7, and 9 years old and follow them all for a period of 3 years.
 - i) Which type of panel design will the linguist used? (1 mark)
 - ii) Why is the linguist not collecting data on 3-,6-,9- years old for three years? (3 marks)
 - iii) Which panels will have both data on language acquisition during the fifth year of life? (2 marks)
- d) Give five special problems with panel designs (5 marks)

QUESTION FOUR (20 MARKS)

- a) State any four sources of survey errors? (4 marks)
- b) Briefly explain four approaches in dealing with Non-response. (4 marks)
- c) Explain the two phases used in two phase sampling (4 marks)
- d) Differentiate between designed based and model based sampling (4 marks)
- e) What are the basic features of a traditional population and housing census? (4 marks)

QUESTION FIVE (20 MARKS)

- a) Briefly explain any two advantages of imputation as a way of dealing with non-response error (2 marks)
- b) State any four ways on how to avoid Non-Response bias (4 marks)
- c) Briefly explain five considerations when to choose Multiphase design. (5 marks)
- d) What are the recommended reasons in Replicated Sampling for periodic, large-scale, agricultural surveys? (4 marks)
- e) An investigator takes a two-phase sample to estimate the prevalence of diabetes in a population. In phase I, a Simple Random Sample of size 1000 is taken from the population of size 100,000, and each individual is asked demographic information and whether he or she has diabetes. It is known that some demographic groups are more at risk for diabetes than others; in addition, the self-report of diabetes may be inaccurate. Therefore, each individual in the phase II sample is given a medical exam to determine diabetes status. The phase I sample is divided into 4 strata. Stratum h has n_h observations in the phase I sample and m_h observations in the phase II sample. After the medical exam, r_h persons in stratum h of the phase II sample were determined to have diabetes.

Stratum	n_h	m_h	r_h
High risk group and reports diabetes	241	96	86
High risk group and does not report diabetes	113	45	17
Low risk group and reports diabetes	174	35	29
Low risk group and does not report diabetes	472	47	8

Estimate the total number of persons with diabetes in the population, along with its standard error. (5 marks)