University of Kelaniya – Sri Lanka
Centre for Distance & Continuing Education
Bachelor of Science (General) External
Second year second semester examination - 2024 (July2026)
(New Syllabus)
Faculty of Science

Statistics

STAT 27542 - Survey Methods and Sampling Techniques

Time allowed: **TWO Hours** Answer **ALL** questions.

Calculators are allowed.

(1) A government health department plans to conduct a nationwide survey to estimate the prevalence of diabetes among adults in Sri Lanka.

(a) Discuss the principal steps involved in planning and executing this sample survey. Explain why each step is important and describe the potential consequences if any step is overlooked. **[10 marks]**

(b) Identify and explain three possible sources of error that may arise during this survey. **[10 marks]**

(c) Discuss practical measures that can be taken to minimize these errors and improve the accuracy and reliability of the survey results. **[10 marks]**

(d) Evaluate the suitability of stratified random sampling for this survey. Explain why it may be preferred over simple random sampling. **[10 marks]**

(2) (a) The Ministry of Health wishes to estimate the total number of hours nurses spend each week on administrative paperwork rather than direct patient care in government hospitals. There are 600 nurses employed across the hospitals. A simple random sample of 40 nurses was selected and each nurse recorded the number of hours spent on administrative paperwork during one week. The sample results showed an average of 7.8 hours per nurse per week with a sample variance of 1.96 hours^2. Estimate the total number of nurse-hours spent on administrative paperwork each week by all nurses. Calculate the bound on the error of estimation at approximately the 95% confidence level. **[20 marks]**

(b) The Department of Agriculture wishes to estimate the total amount of fertilizer used during a growing season by $N = 2{,}500$ rice farmers in a district. The department wants an estimate of total fertilizer usage with an error bound of 2,000 kg at approximately the 95% confidence level. Based on previous agricultural surveys, the population variance of fertilizer usage per farmer is estimated to be 100 kg^2. Determine the minimum sample size required for the survey. **[20 marks]**

(3) (a) A hospital administrator plans to conduct a survey to estimate the average waiting time (in minutes) for patients before they see a doctor in the outpatient department. Since waiting times are expected to differ between clinics, the patients are divided into two strata: General Medicine and Specialist Clinics. There are 180 patients in the General Medicine stratum and 120 patients in the Specialist Clinics stratum. Previous hospital records indicate that the waiting time ranges from 24 minutes for General Medicine patients and 16 minutes for Specialist Clinic patients. Assuming that the waiting times are approximately bell-shaped, the population standard deviation may be approximated as one-fourth of the range. The administrator wishes to estimate the overall average waiting time with a bound on the error of estimation of 2 minutes. Further, assuming that the cost of sampling is the same for both strata, determine the total sample size required for the survey and the number of patients that should be selected from each stratum using optimum allocation. **[30 marks]**

(b) A university administration wishes to estimate the proportion of undergraduate students who support the introduction of a new online learning management system (LMS). The university has 6,000 registered undergraduate students and a one-in-five systematic sample is selected from the student registration list after choosing a random starting point. Each sampled student is asked whether they support the new LMS, where a response of "Yes" is coded as 1 and a response of "No" is coded as 0. From the systematic sample, 1,200 students are selected and 840 students respond "Yes". Using the information provided, estimate the proportion of undergraduate students who support the introduction of the new LMS, calculate the bound on the error of estimation at approximately the 95% confidence level and interpret the results. **[20 marks]**

(c) The Department of Education wishes to estimate the average annual expenditure per student on educational materials in a district. The district consists of 300 schools (clusters). A one-stage cluster sample of 15 schools is selected by simple random sampling and all students within the selected schools are included in the survey. The following summary statistics are obtained from the selected schools: $\sum_{i=1}^{15} m_i = 750$, where m_i is the number of students in the i^{th}school, $\sum_{i=1}^{15} y_i = 3{,}375{,}000$, where $y_i$is the total annual expenditure on educational materials (in dollars) for the i^{th}school and the estimated residual variance is $s_r^2 = \frac{\sum_{i=1}^{15}(y_i - \bar{y}m_i)^2}{n-1} = (18{,}000)^2$. Using the information provided, estimate the average annual expenditure per student, calculate the bound on the error of estimation at approximately the 95% confidence level and interpret the results. **[20 marks]**

(04) (a) An external auditor is assigned to verify the total value of accounts receivable in a large retail company. The company's accounting system provides the recorded value of each customer account (x), but the auditor suspects that some recorded values may not reflect the actual collectible amounts (y). Since verifying every account is costly and time-consuming, the auditor decides to use a ratio estimator based sampling approach. A simple random sample of 12 customer accounts is **[30 marks]**

selected from a total of $N = 2400$ accounts. For each selected account, both recorded values (x) and actual verified values (y) are obtained. The total recorded value from the accounting system is $\tau_x = \$60{,}000$. From a pilot study, the following sample summary statistics are obtained: $\bar{x} = 18$, $\bar{y} = 17.1$ and the estimated variance of the ratio residuals is $s_r^2 = 4.84$.

The auditor wishes to estimate the total actual receivables (τ_y) with a bound on the error of estimation of \$600. Using the relationship between ratio estimation and residual variance, determine the required sample size.

(b) A team of public health researchers is conducting a study to understand the experiences of young adults who have recovered from substance addiction through peer-support groups. Because many individuals are reluctant to publicly disclose their past addiction and there is no complete list of people who meet the study criteria, it is difficult for researchers to identify potential participants directly. The researchers decide to begin by interviewing a few known participants and then ask each participant to recommend other eligible individuals who may be willing to participate in the study.

(i) Based on the given scenario, identify the most appropriate sampling technique for the study and justify your selection. **[10 marks]**

(ii) Discuss two advantages and one limitation of the sampling technique identified in part (i). **[10 marks]**

Equation sheet:

1. For simple random sampling:
   Ratio estimation:
   Estimator of the population ratio (R): $r = \frac{\bar{y}}{\bar{x}}$
   Estimated variance of r: $\hat{V}(r) = \left(1 - \frac{n}{N}\right)\left(\frac{1}{\mu_x^2}\right)\frac{s_r^2}{n}$ $\quad where\ s_r^2 = \frac{\sum_{i=1}^{n}(y_i - rx_i)^2}{n-1}$
   Sample size required to estimate R: $n = \frac{N\sigma^2}{ND+\sigma^2}$ $\quad where\ D = \frac{B^2\mu_x^2}{4}$
   Sample size required to estimate τ_y: $n = \frac{N\sigma^2}{ND+\sigma^2}$ $\quad where\ D = \frac{B^2}{4N^2}$

2. For Stratified sampling:
   Neyman allocation: $n_i = n\left(\frac{N_i\sigma_i}{\sum_{k=1}^{L} N_k\sigma_k}\right)$ $\quad n = \frac{\left(\sum_{k=1}^{L} N_k\sigma_k\right)^2}{N^2D + \sum_{i=1}^{L} N_i\sigma_i^2}$
   Proportional allocation: $n_i = n\left(\frac{N_i}{\sum_{k=1}^{L} N_k}\right)$ $\quad n = \frac{\sum_{k=1}^{L} N_k\sigma_K^2}{ND + \frac{1}{N}\sum_{k=1}^{L} N_k\sigma_k^2}$
   where $D = \frac{B^2}{4}$ when estimating μ and $D = \frac{B^2}{4N^2}$ when estimating τ

3. For Cluster sampling:
   Estimator of the μ: $\bar{y} = \frac{\sum_i^n y_i}{\sum_i^n m_i}$ *and* estimated variance of $\bar{y}$: $\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s_r^2}{n\bar{M}^2}$

4. For systematic sampling:
   Estimator of the μ: $\hat{y}_{sy} = \frac{\sum_i^n y_i}{n}$ *and* estimated variance of $\hat{y}_{sy}$: $\hat{V}(\hat{y})_{sy} = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$

End of Paper