
Department of: Statistics and Informatics

College of: Administration & Economics

University of: Salahaddin University - Erbil

Subject: High Diploma

Course Book –

Lecturer's name: Nazeera S. Kareem (Assist.Prof. Dr)

Academic Year: 2023/2024

Course Book

| | |
|---|---|
| **1. Course name** | **Sampling Theory** |
| **2.Lecturer in charge** | **Nazeera S.Kareem** |
| **3.Department/ College** | **Statistics and Informatics/ Administration & Economics** |
| **4. Contact** | **e-mail: nazeera.kareem@su.edu.krd**<br>**nazeera.barznji@gmail.com**<br>**Tel: (optional) 07504812550** |
| **5. Time (in hours) per week** | **Theory: 1 hours**<br>**Practical: 2 hour** |
| **6. Office hours** | **20 hours per week** |

| | |
|---|---|
| **7. Course code** | **SAE104** |
| **8.Teacher's academic profile** | -BSC(Statistics department)( College Administration & Economics) from (The University of Sulaimania)<br>-MSC (Statistics department)( College Administration & Economics) from (The University of Salahaddin-Erbil)(2001)<br>-PHD (Statistics department)( College Administration & Economics) from the(University of Salahaddin-Erbil) (2015)<br>-Assistant researcher( College Administration & Economics)1980<br>-Assistant lecturer (The University of Salahaddin -Erbil) (2006)<br>-Lecturer(2015)(The University of Salahaddin -Erbil)<br>-Assistant Prof.(2019) )(The University of Salahaddin -Erbil)<br>-Teaching from (43) years at the University of Salahaddin -Erbil)<br><br>**The Subjects that I taught:**<br>- **Sampling Theory**(Master- Statistics and Informatics department )<br>-**Non Parametric Statistical test** (Master- Statistics and Informatics department)<br>**-Numerical Analysis with R programing Language** --2nd Stage Statistics and Informatics department<br>-**Numerical Analysis** 2nd Stage Statistics and Informatics department)<br>-**Advanced Statistics** 2ndStage Economics department<br>-**Principle of Statistics** 1st Stage Economics department |

| | |
|---|---|
| | -**Principle of Statistics** 1st Stage Administration department<br>-**Computer- M.S. Excel** 2nd Stage Administration department<br>-**Principle of Mathematics** 1st Stage Finance and Banks department<br>-**Academic Debate** 1st Stage Statistics and Informatics department<br>-**Methodical Research** 4st Stage Statistics and Informatics department<br><br>• **The researches that I had accomplished**<br>1-PhD (A Dissertation) about [Genetic Effects using R-QTL Statistical Analysis after Chemical Attack on Survivors in Halabja- kurdistan<br>2- MSC( Thesis) Statistical Study in Analyzing the Chemical Structure of Some Carbonic Rocks in Kurdistan-Iraq<br>3- (Women and education)for-conference about the reality of women (in the global women's day) it took 3 days (3 / 11-8 / 2008)<br>4- (University leadership) in conference the Ministry of Higher Education and Scientific research 2008<br>5- (Using Entropy in Kurdish poetry(homeland) from poet Fayaq Bekas in work shop for World Statistics Day(20/10/2015)<br>6- construction robust simple linear regression profile monitoring Simulation study (7/3/2017)<br>7-Logistic Regression and Discriminant analysis to identify the risk of Diabetes<br>8- De-noise data by using Multivariate Wavelets |

| | |
|---|---|
| | in the Path analysis with application<br>9-Constructing Mathematical Models, by Interpolation Methods, of people's interest to listening to Quran's voice or<br>10-التنبؤبالرقم القياسي ومعدل التضخم لإقليم كوردستان العراق بإستخدام نموذج حركي للشبكات العصبية مع السلاسل الزمنية<br>11- Assessing by Model Fitting Criteria and comparing between Stirling and Aitken interpolation formulas to determine the preferred treatment of choice, (surgery) or (surgery with chemotherapy), for lung cancer in Kurdistan 2024<br>12- Utilizing Cox Regression Analysis and Bootstrapping Test for Indicating the Risk of Smoking on Health in Kurdistan2024<br>13- Estimating Parameters and Predicting the Risk of Miscarriage, Infertility, and Psychological Disorders Among Chemical Attack Survivors of Kurdistan by Using Multinomial Logistic Regression Analysis 2024<br>14-<br>**Languages**<br>-Kurdish the mother language<br>-English<br>-Arabic |
| **9. Keywords** | Sampling Theory, Probability Sampling , Nonprobability Sampling |

**10.Course overview:**

Sampling theory refers to the statistical branch dealing with the selection, gathering, and evaluation of data obtained from a sample population. It provides efficient and trusted methods to estimate and draw inferences, plus estimate characteristics of the whole population along with testing hypotheses or it is a branch of statistics that provides a framework for making inferences about a population based on a subset of that population, called a sample. Its types include simple random, systematic, stratified, cluster, non-probability, convenience, judgmental, snowball, and quota sampling.

- Sampling theory is a branch of statistics that provides a framework for making inferences about a population based on a subset of that population, called a sample.
- Its types include simple random, systematic, stratified, cluster, non-probability, convenience, judgmental, snowball, and quota sampling.
- Moreover, businesses, marketers, financiers, educators, economists, quality control specialists, agriculturists, forest managers, healthcare professionals, environmental researchers, and social scientists all apply it.
- Hence, it is essential for efficient data collection, cost-effectiveness, informed decision-making, population representation, practicality in complex situations, and statistical analysis support.

**Sampling Theory in Statistics Explained**

Sampling theory in statistics means the practice of choosing a subset entity or individual out of a larger population to collect data to make inferences about the whole population. Moreover, it works by selecting a representative sample of market participants or financial data. The samples could contain investor behavior, stock prices, or bond yields.

Furthermore, researchers employ a random or systematic sampling method to ensure unbiased pictures. They then analyze these samples and draw appropriate conclusions concerning the larger financial landscape. Additionally, sampling theory bias arises when the

sample selection is more or less likely to include specific individuals, resulting in inaccurate or skewed results.

Statistical sampling theory provides a powerful theoretical framework for generalizing from samples to corresponding populations and is most relevant when generalizing to populations of units and settings (external validity question 1) that can be enumerated and are under the control of the researchers. In research practice, however, random sampling from designated universes of students, classrooms, schools, etc. is uncommon in most studies that involve causal associations and random assignment of experimental units to different treatment conditions. The relevance of sampling theory to generalizations about the treatments and outcomes and to other questions of external validity is less clear. This situation leaves researchers in a difficult position: If sampling theory is not often used or is not able to answer important generalizability questions, how should we justify the external validity of causal relationships? What is needed is a viable new theory of generalization that can guide the design of primary studies and the design of research syntheses.

The principles introduced by Cook (1991, 1993) and further elaborated by Shadish *et al.* (2002) can serve as a starting point. They borrow strategies and techniques commonly used to justify construct-validity claims and have been applied in programmatic approaches to support the transfer of research and technology from labs to applications (US FDA, 1998). They provide practical guidelines for exploring and building arguments around generalizability claims and can be applied very well to research syntheses (Matt and Cook, 2009).

Meta-analysis provides the most promising research design to explore generalizability claims using Cook's five principles as no single study provides as many opportunities as research synthesis to examine the robustness and moderating conditions of causal associations. Matt and Cook (2009) examine, in detail, the validity of generalized causal inferences based on research syntheses.

Another promising direction for a comprehensive theory of generalization comes from a branch of mathematics called group theory. In fact, group theory is a broadly applicable and highly abstract mathematical theory designed to describe generalizations of all kinds (Livio, 2006). A mathematical group is defined as any set whose members obey certain rules with respect to particular operations, and therefore, act similarly in response to such operations. The members of a group may include numbers, words in the English language, a pair of jeans, or a causal association. Operations can be as diverse as 'rotate 120°,' 'do nothing,' 'turn inside out,' and 'followed by,' 'add,' or 'multiply.' A group exists if the operations are applied to the members and the following properties hold: (1) closure, (2) associativity, (3) identity element, and (4) inverse.

In contrast to sampling theory, group theory does not characterize a generalization based on the objects it involves (e.g., a set of exemplars defining a universe settings), but by whether an operation changes it or leaves it unchanged. Similar to Cook's pragmatic principles, group theory relies on studying the invariance to transformation (i.e., robustness) and the identification of moderators (i.e., discriminant validity). Translating the tenets of group theory to the generalizability of causal association could provide important insights into the design of primary studies and the analysis of meta-analyses that facilitate generalizable proposition. Based on Cook's five principles and the cursory introduction of group theory, drawing generalizable inferences will benefit from efforts to examine causal associations under many diverse conditions, both substantively irrelevant as well as substantive relevant characteristics of units, treatments, outcomes, and settings. Research synthesis and programs of research bear most promise for providing the empirical warrants for such generalized inferences.

**11. Course objective:**

Statistical Sampling can provide a valid and defensible methodology but it is important to match the type of sample needed to the type of analysis required. It is usually not cost effective or practicable to collect and examine all the data that might be available.

Two basic purposes of sampling are

1-To obtain the maximum information about the population without examining every unit of the population.

2-To find the reliability of the estimates derived from the sample, which can be done by computing the standard error of the statistic.

**12. Student's obligation**

The role of students and their obligations

1-The student attendance to lecture at the time and place as scheduled

by the head of department

2- Preparing the home work

3- Solving the problem(exercise) on the whiteboard by the students

4- The student is ready for the sudden exam on the material is described before (Most of the homework exercises will come after explaining theorems or applications).

5- Enquiry the student of topics mysterious and unintelligible in leisure time of lecturer

6- Preparing the student for daily tests (quiz) after clarify and explain Article The main technique will be used in the lectures. At the beginning of each lecture I will inquire students if they were reading the previous lecture by 5 Minutes (quiz).

7-After complementing explaining the lecture I encourage student to ask questions if they are unable to ask their questions in class, then they may ask outside of class in any time I am in the office.

**13. Forms of teaching**

To accomplish acceptable outcome, the lecturer, use several methods to explain and clarify the lecture

Power point presentation for, title of theorem, definitions, graph, results general formula, Exercises.

2- Use Data Show to view PowerPoint representation.

3-White board using to prove theories and solutions for examples or exercise.

**14. Assessment scheme**

The students Supposed

to do the final exam on 50 degrees

and The lecturer will divide the other 50 degrees as follow:

Midterm Exam, 20%

Quiz, 10%

Seminar, (presentation) 10%

Other Activity, Daily students activity(, homework , attendance, Student participation in solving problems on the whiteboard, ) 10%

| | | | | |
|---|---|---|---|---|
| Total 50% | Midterm Exam, 20% | Other Activity, 10% | Quiz, 10% | Seminar, 10% |

**15. Student learning outcome:**

Sample surveys are an important source of statistical data. A great many published statistics on demographic, economic, political and health related characteristics are based on survey data. Simple random sampling is a well-known method of sampling but, for reasons of efficiency and practical constraints, methods such as stratified sampling and cluster sampling are typically used by statistical authorities such as the Australian Bureau of Statistics and by market research organisations. This course is concerned with the design of sample surveys and the statistical analysis of data collected from such surveys. Topics covered are:

experiments and surveys, steps in planning a survey; randomisation approach to sampling and estimation, sampling distribution of estimator, expected values, variances, generalisation of probability sampling; prediction approach, inadequacies of approach, decomposition of population total, concomitant variables; regression through the origin, estimation by least squares, ratio estimation, variance formulae; balance and robustness; best fit sample; stratified sampling, estimation, allocation, construction of strata, stratification on size variables, post-stratification; two-stage sampling, estimation, allocation, cluster sampling.

Students who successfully complete the course should:

1. understand the principles underlying sampling as a means of making inferences about a population,
2. understand the difference between randomization theory and model based analysis,
3. understand the concepts of bias and sampling variability and stragies for reducing these,
4. be able to analyse data from multi-stage surveys,
5. have an appreciation of the practical issues arising in sampling studies.

## 16. Course Reading List and References:

1-Sampling:Design and AnalysisSharon L. LohrArizona State University

2-Skinner, C. J. 1989. Domain means, regression, and multivariate analysis. In Analysis of complex surveys. Edited by C. J. Skinner, D. Holt, and T. M. F. Smith, 59-88. New York:Wiley.

3-Snedecor. G. W. 1939. Design of sampling experiments in the social sciences. Journal of Farm Economics 21: 846-855.

.

Statistical process control of sampling frames. Survey Methodology 21:
185-190.

Squire, P. 1988. Why the 1936 Literary Digest poll failed. Public Opinion Quarterly 52:
125-133.

Stasny, E. A. 1991. Hierarchical models for the probabilities of a survey classification and nonresponse. Journal of the American Statistical Association 86: 296-303.

StataCorp. 1996. Stata statistical software: Release 5.0. College Station, Tex.: Stata Corporation.

Statistics Canada. 1993. Standards and guidelines for reporting of nonresponse rates. Ottawa, Canada.
Stein, C. 1945. A two-sample test for a linear hypothesis whose power is independent of the variance. Annals of Mathematical Statistics 37: 36-50.
Stockford, D. D., and W. F. Page. 1984. Double sampling and the misclassification of Vietnam
service. ASA Proceedings of the Social Statistics Section, 261-264.
Strunk, W.. and E. B. White. 1959. The elements of style. New York: Macmillan.
Stuart, A. 1984. The ideas of sampling. New York: Oxford University Press.
Student. 1908. The probable error of the mean. Biometrika 6: 1-25.
Sudman, S. 1976. Applied .sampling. San Diego: Academic Press.
Sudman, S., and N. M. Bradburn. 1982. Asking questions: A practical guide to questionnaire design. San Francisco: Jossey-Bass.
Sudman, S., N. M. Bradburn, and N. Schwarz. 1995. Thinking about answers: The application of cognitive processes to survey methodology. San Francisco: Jossey-Bass.
Sudman, S., M. G. Sirken, and C. D. Cowan. 1988. Sampling rare and elusive populations. Science 240: 991-996.
Sukhatme, P. V., B. V. Sukhatme, S. Sukhatme, and C. Asok. 1984. Sampling theory of surveys
with applications. 3rd ed. Ames: Iowa State University Press.
Tanur, J., ed. 1993. Questions about questions: Inquiries into the cognitive bases of surveys. New York: Sage.
Taylor, B. M. 1989. Redesign of the National Crime Survey. Washington, D.C.: Department of
Justice.
Taylor, H. 1995. Horses for courses: How different countries measure public opinion in very different ways. Public Perspective 6: 3-7.
Teichman, J., D. Coltrin, K. Prouty, and W. Bir. 1993. A survey of lead contamination in soil along Interstate 880, Alameda County, California. American Industrial Hygiene Association 54: 557-559.
Thomas. D. R. 1989. Simultaneous confidence intervals for proportions under cluster sampling.
Survey Methodology 15: 187-201.
Thomas, D. R., and J. N. K. Rao. 1987. Small-sample comparisons of level and power for simple
goodness-of-fit statistics under cluster sampling. Journal of the American Statistical Association 82: 630-636.
Thomas, D. R., A. C. Singh, and G. R. Roberts. 1996. Tests of independence on two-way tablesunder cluster sampling: An evaluation. International Statistical Review 64: 295-311.

| | |
|---|---|
| **17. The Topics:** | **Lecture (14weeks)** |
| **Chapter One**<br>**1.Introduction**<br>**1.2 Some Concepts**<br>**Statistics**<br>**Data**<br>**Population**<br>**Types of statistical population**<br>**The population size**<br>**Sample**<br>**Sample size**<br>**Representative sample:**<br>**1.3Sources of data and information**<br>**1.3.1-Historical Sources**<br>**1.3.2-Field Sources**<br>**1.4Methods of data collection**<br>**1.4.1- Complete enumeration or census**<br>**1.4.2-Sampling**<br>**-Sampling Frame**<br>**- Sampling unit**<br>**1.5 Advantages and disadvantages of sampling Method**<br>**1.5.1 Advantages of Sampling over Census:**<br>**1.5.2 Disadvantage of sampling Method**<br>**1.6Sampling Error vs. Non-sampling Error**<br>**1.6.1 Sampling Errors** | |

| | |
|---|---|
| **1.6.2Non-Sampling Errors**<br>**1.7 Steps in Sampling Process**<br>**1.8 Techniques of Sampling:**<br>**First== probability(Random) sampling**<br>**Second==Non-probability sampling(Non-random Sample or purposive sample):**<br>**Chapter two**<br>**First== probability(Random) sampling**<br>**2.1Definition of random sampling**<br>**2.2-Simple Random Sampling (SRS)**<br>**2.3Types of Simple Random Sampling**<br>**2.3.1. Simple Random Sampling Without Replacement (SRSWOR)**<br>**2.3.2. Simple Random Sampling With Replacement (SRSWR)**<br>**2.4Steps to perform Simple Random Sampling**<br>**Step 1: Define the Population (Make a List)**<br>**Step 2: Assign a Sequential Number**<br>**Step 3: Choose Sample Size**<br>**Step 4: Use a Random Number (Random Number Techniques)**<br>**2.5Probability of drawing a Sample:**<br>**2.5.1 Simple Random Sampling Without Replacement (SRSWOR)**<br>**2.5.2 Simple Random Sampling With Replacement (SRSWR)]:**<br>**2.6. Probability of drawing a unit**<br>**2.6.1 Probability of drawing a unit SRSWOR**<br>**2.7. Examples**<br>**2.8. Estimation of population mean**<br>**2.8.1. Estimation of population mean (SRSWOR)**<br>**2.8.2Estimation of population mean SRSWR** | |

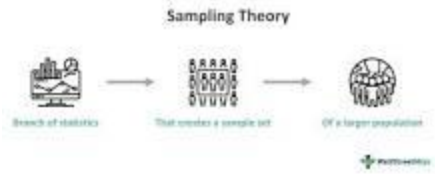
| | |
|---|---|
| **2.9. Estimation of population variance (Estimation of variance from a sample)**<br>**2.9.1 In the case of SRSWOR**<br>**2.9.2In the case of SRSWR**<br>**2.10 Efficiency of under SRSWOR over SRSWR**<br>**2.11Confidence limits for the population mean**<br>**2.12 Determination of Sample Size**<br>**2.12.1. Pre-specified variance**<br>**2.12.2 Pre-specified estimation error**<br>**2.12.3. Pre-specified width of the confidence interval**<br>**2.12.4. Pre-specified coefficient of variation**<br>**2.12.5. Pre-specified relative error**<br>**2.12.6. Pre-specified cost**<br>**2.13Corollary (1)**<br>**2.14Theorem 2.2**<br><br>**Chapter Three**<br>**First== probability(Random) sampling**<br>**3-Stratified Random Sampling**<br>**3.1Types of stratified sampling**<br>**3.2Steps to select stratified random sample:**<br>**3.3Procedure of stratified sampling**<br>**3.4Estimation of population mean**<br>**3.5Estimation of population variance**<br>**Chapter Four**<br>**First== probability(Random) sampling**<br>**4-Systematic Sampling** | |

| | |
|---|---|
| **4.1Arranging Systematic Sampling**<br>**4.2Types of systematic sampling**<br>**4.2.1Linear systematic sampling**<br>**4.2.2Circular systematic sampling**<br>**4.3 Advantages of systematic sampling**<br>**4.4Estimation of population mean**<br>**4.5Estimation of variance**<br>**Chapter Five**<br>**First== probability(Random) sampling**<br>**5-Cluster Sampling**<br>**5Estimation of population mean**<br>**Estimate of variance:**<br>**Chapter Six**<br>**Second== Non-random sample or purposive sample:**<br>**6-Convenience sampling**<br>**Chapter Seven**<br>**Second== Non-random sample or purposive sample:**<br>**7-Quota sampling**<br>**Chapter Eight**<br>**Second== Non-random sample or purposive sample:**<br>**8-Self-selection (volunteer) sampling**<br>**Chapter Nine**<br>**Second== Non-random sample or purposive sample:**<br>**9-Snowball sampling**<br>**Chapter Ten**<br>**Second== Non-random sample or purposive sample:**<br>**10-Purposive (judgmental) sampling** | |

**19. Examinations:**

**20. Extra notes:**

What is sampling theory short notes?




Sampling theory is a branch of statistics that provides a framework for making inferences about a population based on a subset of that population, called a sample. Its types include simple random, systematic, stratified, cluster, non-probability, convenience, judgmental, snowball, and quota sampling.

21. Peer review: