ACADEMIC PROGRAMME: BSCS COMPUTER SCIENCE

COURSE CODE AND TITLE: BSCS 305: COMPUTING RESEARCH METHODS

LECTURER'S NAME: PHILIP BTTOK

LECTURER'S CONTACTS: Phone No.: 0723683409 **Email**: philip.bittok@gmail.com

METHOD SECTION

Expected Learning Outcomes:

By the end of this lesson, you should be able to:

i. Understand the sampling procedures
ii. Understand how to computer sample sizes

Introduction

- The method section of a research study describes the procedures that have been followed in conducting the study
- At this stage, techniques, of obtaining data are developed and data is actually collected to test hypotheses, if any
- The steps involved in in conducting the study should be described in detail
- This helps other researchers in understanding one's study, especially where replication is desired
- This class will focus on populations, samples, samling methods, sampling bias, types of variables, and measurement scales

Populations and Samples

- We defined a population as a complete set of individuals, cases or objects, with some common observable characteristics
- A particular population has some characteristics that differentiate it from other populations
- A subset of a particular population is called a sample
- Samples are useful because it is hard to study an entire target population
- Researchers therefore identify and define an experimentally accessible population – also called survey population
- A target population is a population that a researcher would like to generalize the results of a study

- When defining the accessible population, one should not be influenced by convenience
- The researcher should define the population as consistently as possible with the purpose of the study
- In research, consistency supersedes convenience
- There should be a rationale when defining accessible population from the target population
- Example. One wants to study the success factors for online teaching for students in public universities in Kenya
- The target population is students in public universities in Kenya
- The accessible population might be students in a public university in Kenya, like Kenyatta University
- The target population and the accessible population should be comparable on many grounds – this is referred to as population validity
- The accessible population should be carefully defined as that others will assess the extent in which the results can be generalized
- After identifying the target and the accessible population, then next step is to define the sample from the accessible population
- The sample must be large enough to represent the salient characteristics of the accessible population and thus the population – it is assumed that population validity exists
- One factor that is considered by many researchers is the sample size
- The sample size depends on many factors like the number of variables, the type of research design, the method of data analysis, and the size of accessible population
- It has been argued by researchers that for correlational research, a sample size of 30 should be used, for descriptive studies, 10 percent of the accessible population is enough for experimental studies, and at least 30 cases are required per group

Computing the Sample Size

- If allowed by resources and time, a bigger sample size is encouraged
- With a large sample, the researcher is confident that if another sample of the same size was to be selected, there is a high degree of similarity in results
- The risk that is associated with small samples is that they do not reproduce the salient features of the accessible population to an acceptable degree
- If allowed by resources and time, a bigger sample size is encouraged

- With a large sample, the researcher is confident that if another sample of the same size was to be selected, there is a high degree of similarity in results
- The risk that is associated with small samples is that they do not reproduce the salient features of the accessible population to an acceptable degree
- The discrepancy between the sample characteristics and the population characteristics is referred to as sampling error
- The smaller the sample, the bigger the sampling error.

Situations when large samples are required

1. When many variables are held constant in the study
2. When hypothesized relationships or differences between variables are very small in real life – when anticipated effect size is small
3. When the study requires the sample to be broken into sub-groups – for example users using smart phones, desktops, or laptops
4. When it is expected that many subjects or cases in the sample will not respond, drop out or die in the case of animals or plants
5. When a very high level of statistical significance is required – for example a significance level of 0.001
6. When the accessible population is highly heterogenous on the variable under study (if it were possible for the population to be wholly homogenous, studying one case would be enough)

Sample size

- Many researchers find it hard to determine the sample sizes of their studies
- The rule to follow is to always obtain a big a sample as possible
- Resources and time limit the extent in which samples can be enlarged

Sampling procedures

- After deciding the sample size, the next step will be to determine the procedure to be used to select participants to be included in the sample
- To select a a representative sample, one will be required to have a sampling frame
- Sampling frame is a list, directory or index of cases from which a sample can be selected
- Subjects or cases selected from the sampling frame form the units of observation in a study

- Examples of sampling frames include national census list, list of car owners from Registrar of Motor Vehicles, list of approved plans from the City Commission, list of Information Systems registered in a county government
- The degree of generalization of a study depends on the accuracy of the sampling frame from which the sample was selected
- If a sample is drawn from an incomplete sampling frame, the findings from that study may not be generalised beyond the sample or the sampling frame from which the sample was drawn
- In some cases, a sampling frame may not exist or it would require much time and resources to come up with one
- In such a case, the researcher will be required to decide on any other method of selecting the sample as long this is described in detailed in the research report
- In other cases, the target population is too small that selecting a sample would be meaningless – one would take the whole population
-

Sampling Techniques

- The following are the sampling techniques that have been described
- Sampling techniques fall in two categories:
    - Probability sampling
        - Simple random sampling
        - Systematic random sampling
        - Stratified random sampling
        - Cluster sampling
    - Biased/Non-probability sampling
        - Purposive sampling
        - Quota sampling
        - Snowball sampling

Probability sampling

- The main goal of probability sampling is to select a reasonable number of subjects, objects or cases that represent the target population

- Probability sampling can provide accurate information about groups that are too large to study in its entity
- It provides an efficient way of capturing, in a small group, the variations or heterogeneity that exist in the target population

Random sampling

- Random sampling is the key to obtaining a representative sample
- In random sampling, every sample of a given size in the accessible population has an equal chance of being selected
- Random sampling allows generalizability to a larger population with a margin of error that is statistically determinable
- It also allows the use of inferential statistics – statistics calculated on the sample can be calculated to determine the degree to which they accurately represent the population parameters
- **Methods of random sampling**
- There are four methods of random sampling:
    - Simple random sampling
    - Systematic random sampling
    - Stratified random sampling
    - Cluster samplig

Simple random sampling

- This method involves giving a number to every subject or member of the accessible population, placing the numbers in a container and then picking any number at random
- The subjects corresponding to the numbers picked are included in the sample
- Another strategy involves the use of a table of random numbers
- Tables of random numbers can be found in statistics sources or from computer programs

Systematic random sampling

- In this case, every K^{th} case is in the population frame is selected for inclusion in the sample
- To obtain a truly random sample using this method, the list of all the members in the sampling frame should be randomised
- A list that is arranged in alphabetical order is not random

- Cases or objects arranged in numerical order may follow a certain pattern, for example, size, or color, and therefore picking every K^{th} subject may eliminate cases of a certain size entirely
- If a list of individuals is arranged in alphabetical order, a subgroup of the population may be concentrated within one section of the alphabet because some subgroups have names that often start with a particular letter or alphabet
- In such a case, section of the population may be left out
- Alternatively, the majority of subjects in a particular sub-group may be included in the sample
- This introduces a systematic error in sampling
- Once the population frame is randomized, the researcher then decides on the sampling interval
- A sampling interval is the distance between the cases that are selected for the sample
- To rule of thumb in deciding the sampling interval is to divide the total population by the sample size
- If a research is to be done on all first year students in public universities in Kenya, and there are 12000 first year students, and the sample size is 500, then if using systematic random sampling, the sampling interval is 12000/500 which is 24.
- We will pick every 24^{th} student from the sample

Stratified random sampling

- The goal of stratified random sampling is to achieve desired representation from various subgroups in the population
- In stratified random sampling, subjects are selected in such a way that the existing subgroups in the population are more or less reproduced in the sample
- This means that the sample will consist of two or more sub-groups
- There is a difference between stratified sampling and random assignment
- In random assignment, a random sample is first selected by whatever method
- Once a random sample is selected, subjects within the sample are randomly assigned to two or more subgroups or treatment groups
- In stratified sampling, the population is divided into two or more groups using a given criterion and then given number of cases are randomly selected from each population subgroup
- To use stratified sampling, one will first decide on the criteria under which the population and hence the sample will be stratified

- Example of criterion used for stratification could be gender, income level, size, color, type of smart phone, type of system, and type of teaching method
- The next step is to determine the size of each stratum or subgroup in the sample
- If we use the first year students in public universities, where we indicated that there are 12000 first year students
- If we decide to use gender as the variable for stratum
- If we have 8000 male students and 4000 female students, and a desired sample is 600 students
- Then we could design the stratums to reflect what we have in the target population such that the strata will be 400 male students and 200 female students
- The actual method of sampling from each subgroup of the population can be simple random sampling or systematic random sampling

Steps in stratified random sampling

1. Identify the population
2. Define the criterion for stratification
3. List the population according to the defined strata or subgroups
4. Determine the required sample size and the appropriate representation in each stratum. This can be proportionate or equal
5. Select, using random numbers, an appropriate number of subjects for each stratum

Advantage of stratified random sampling

- One thing to note is that stratification can be done on more than one variable
- An advantage of stratified random sampling is that it ensures inclusion, in the sample, of subgroup, which otherwise would be omitted entirely by other sampling methods because of their small numbers in the population

Cluster sampling

- Cluster sampling is used when it is not possible to obtain a sampling frame because the population is either very large or scattered over a large geographical area
- Cluster sampling involves selection of an intact group
- All the members of such an intact group are then included in the sample and each member becomes a unit of observation

Example of cluster sampling

- A doctor may want to study patients suffering from malaria in Kenya

- It would be time-consuming and expensive for the doctor to try and list all malaria patients who are hospitalized in Kenya
- The doctor may consider it logical to confine himself or herself to district hospitals.
- A list of all district hospitals would be compiled and one or two district hospitals would be randomly selected, depending on the required sample size
- All malaria patients hospitalised in the selected district hospital(s) would then be included in the sample
- It is evident that in cluster sampling, it is the groups or clusters that are randomly selected and not the individuals or cases.
- In this type of sampling, it is assumed that the clusters have similar characteristics
- Examples of clusters include schools, towns, cities, hospitals, large-scale industries, and government ministries
- Cluster sampling takes less time and is convenient
- The degree of generalization of the research findings depends on the degree to which the clusters within the population are similar
- Where the clusters are similar to a very high degree, cluster validity is said to exist

Steps in cluster sampling

1. Identify the population
2. Define the cluster forming the population
3. Determine the required sample size
4. List all the clusters in a random order
5. Using the table of random numbers, select the required number of clusters according to the sample size required
6. All members in the selected clusters are included in the samples as units of observation

Cluster sampling – multi-stage cluster

- Cluster sampling can be multi-stage, where a researcher can do cluster sampling in cycles
- Example is if a researcher wants to find out the government's ICT rollout in rural areas in Kenya
- The researcher can start by sampling at random, one county from the 47 counties in Kenya

- From the selected county, one can then select two sub-counties by listing all the sub-counties in the county
- The researcher can then select 4 wards from each subcounty, making 8 wards

Biased sampling/Non-probability sampling

- Non-probability sampling is used when a researcher is not interested in selecting a sample that is representative of the population
- Mostly used in qualitative studies because the focus is on in-depth information and not making inferences or generalizations
- The methods in biased sampling include purposive sampling, maximum variation versus homogenous sampling, snowball sampling, quota sampling, and convenient/accidental sampling

Purposive sampling

- This type of sampling allows the researcher to use cases that have the required information with respect with their objectives of the study
- Cases here are handpicked because they are informative
- You have to specify the criteria one used in picking the cases – for example, a certain age, educational level, or religious sect
- One disadvantage of this method of sampling is that the results might be dismissed because of being extreme

Quota sampling

- This strategy is similar to stratified random sampling and the objective is to include various groups or quotas of the population in the study based on some criteria
- Example is a researcher might want to include a certain religion or social class in the sample and therefore picks quotas from each
- The researcher purposively selects subjects to fit the quotas identified
- The selection of actual participants is not random since subjects are picked as they fit into identified quotas

Snowball sampling

- In this method, initial subjects with the identified characteristics are identified using purposeful sampling technique
- The identified subjects name others that they know have the required characteristics until the researcher gets the number of cases they require
- Useful in cases where the population that has the characteristics is now well known
-

Further E-Resource Readings

Lune, H., & Berg, B. L. (2016). *Qualitative research methods for the social sciences*. Pearson Higher Ed.

7.0 REFERENCE JOURNALS

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.