

**Name : Haseeb Ullah**

**ID : 13401**

**Paper : Research methodology**

**SECTION A**

1. May remember that three years ago there was a multistate outbreak of illnesses caused by a specific and unusual strain of *Listeria monocytogenes*. As part of the investigation of this outbreak, CDC workers checked the food histories of 20 patients infected with the outbreak strain and compared them with the food histories of 20 patients infected with other *Listeria* strains. This study design is best described as which one of the following:
  - A. Analytical, experimental
  - B. **Observational, case-control**
  - C. Analytical, observational,
  - D. Cohort Descriptive
  
2. A published study follows a large group of women with untreated dysplasia of the uterine cervix, documenting the number who improve, stay unchanged, or progress into cervical cancer. This study design is best described as which one of the following:
  - A. Analytic, experimental
  - B. Analytic, observational, cohort
  - C. Analytic, observational, case/control
  - D. **Descriptive, observational**
  
3. A community assesses a random sample of its residents by telephone questionnaire. Obesity is strongly associated with diagnosed diabetes. This study design is best described as which one of the following:
  - A. Case-control
  - B. Cohort
  - C. **Cross-sectional**
  - D. Experimental
  
4. Based on a list of residents from election rolls, 2/3 of men in a large city are invited (including repeated educational urgings) and 1/3 of men are not invited to be screened by PSA blood test

for prostate cancer. Over the next 10 years the two groups are compared as to the rate of death from prostate cancer. This study design is best described as which one of the following:

- A. Case-control
  - B. Cohort
  - C. Cross-sectional
  - D. **Experimental**
5. In a case-control study of alcohol intake and bladder cancer, cases and matched controls are each interviewed by interviewers who are not blinded as to whether the subject is a case or a control. Many of the interviewers are in fact convinced that drinking alcohol is a cause of bladder cancer. Is this likely to represent a bias?
- A. No, because the interviewers can't affect whether the subjects are considered cases or controls; that's already decided
  - B. Yes, but it's hard to predict the direction of the bias.
  - C. **Yes, and would predispose to a rejection of the null hypothesis.**
  - D. Yes, and would predispose to an acceptance of the null hypothesis.
6. Interviewing all members of a given population is called:
- A. A sample.
  - B. A Gallup poll.
  - C. **A census.**
  - D. A Nielsen audit.
7. Sampling means following a sequence of stages. Which ONE of the following stages should come before the others?
- A. Proceed with the fieldwork.
  - B. Find suitable source for the population members.
  - C. Define the people of interest.
  - D. **Examine the objective of the study.**
8. Which ONE of these sampling methods is a probability method?
- A. Purposive.
  - B. Judgement.

- C. Convenience.
  - D. **Simple random.**
9. Which ONE of the following is the benefit of using simple random sampling?
- A. **We can calculate the accuracy of the results.**
  - B. The results are always representative.
  - C. Interviewers can choose respondents freely.
  - D. Informants can refuse to participate.
10. Which ONE of the following is the main problem with using non-probability sampling techniques?
- A. The expense.
  - B. The results are never representative.
  - C. **Human judgement error.**
  - D. Informants can refuse to participate.
11. Which ONE of the following is the best – but an often unused – way to decide on sample size?
- A. By using industry standards.
  - B. **By calculation.**
  - C. By ‘building blocks’.
  - D. By budget available.
12. Which ONE of the following methods is generally used in qualitative sampling?
- A. Random digit dialing.
  - B. **Purposive.**
  - C. Stratified random.
  - D. Simple random.
13. The median of 7, 6, 4, 8, 2, 5, 11 is
- A. **6**
  - B. 12
  - C. 11

- D. 4
14. Number which occurs most frequently in a set of numbers is
- A. Mean
  - B. Median
  - C. Mode
  - D. None of above
15. The mode of 12, 17, 16, 14, 13, 16, 11, 14 is
- A. 13 B. 11
  - C. 14
  - D. 14 and 16

## SECTION B

### Q: 1

Explain cohort study and types of cohort study design in detail?

**Ans :**

#### **Cohort study:**

A cohort study is a particular form of longitudinal study that samples a cohort (a group of people who share a defining characteristic, typically those who experienced a common event in a selected period, such as birth or graduation), performing a cross-section at intervals through time. It is a type of panel study where the individuals in the panel share a common characteristic.

Cohort studies typically observe large groups of individuals, recording their exposure to certain risk factors to find clues as to the possible causes of disease.

The cohort study design is the best available scientific method for measuring the effects of a suspected risk factor.

#### **Types of cohort study design:**

Two types of cohort study design which are given below.

- Prospective study.

- Retrospect study.

### **Prospective study:**

A prospective cohort study is a longitudinal cohort study that follows over time a group of similar individuals who differ with respect to certain factors under study, to determine how these factors affect rates of a certain outcome. For example, one might follow a cohort of middle-aged truck drivers who vary in terms of smoking habits, to test the hypothesis that the 20-year incidence rate of lung cancer will be highest among heavy smokers, followed by moderate smokers, and then nonsmokers.

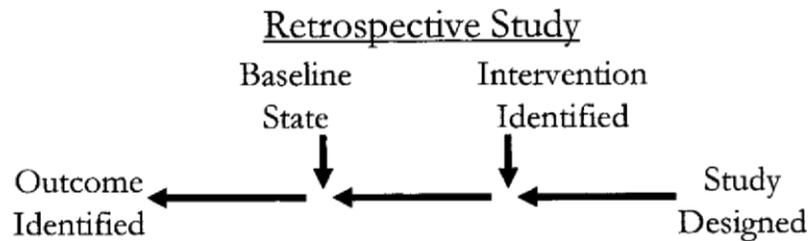
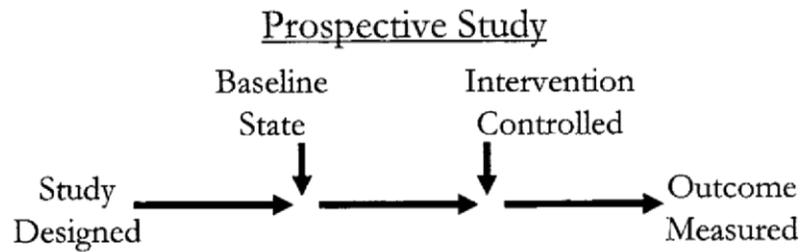
One of the advantages of prospective cohort studies is they can help determine risk factors for being infected with a new disease because they are a longitudinal observation over time, and the collection of results is at regular time intervals, so recall error is minimized.

After baseline information is collected, subjects in a prospective cohort study are then followed “longitudinally,” i.e. over a period of time, usually for years, to determine if and when they become diseased and whether their exposure status changes outcomes. In this way, investigators can eventually use the data to answer many questions about the associations between “risk factors” and disease outcomes.

### **Retrospect study:**

A retrospective cohort study, also called a historic cohort study, is a longitudinal cohort study used in medical and psychological research. A cohort of individuals that share a common exposure factor is compared with another group of equivalent individuals not exposed to that factor, to determine the factor’s influence on the incidence of a condition such as disease or death. Retrospective cohort studies have existed for approximately as long as prospective cohort studies.

Retrospective cohort studies try to compare the risk of developing a disease to some already known exposure factors, a case-control study will try to determine the possible exposure factors after a known disease incidence. Both the relative risk and odds ratio are relevant in retrospective cohort studies, but only the odds ratio can be used in case-control studies. Although most case-control studies are retrospective, they can also be prospective when the researcher still enrolls participants based on the occurrence of a disease as new cases occur.



**Q: 2**

Define sample and explain types of non probability sampling?

**Ans:**

**Sample:**

A sample refers to a smaller, manageable version of a larger group. It is a subset containing the characteristics of a larger population. Samples are used in statistical testing when population sizes are too large for the test to include all possible members or observations. A sample should represent the population as a whole and not reflect any bias toward a specific attribute.

- A sample refers to a smaller, manageable version of a larger group or subset of a larger population.

- Using samples allows researchers to conduct their studies easily and in a timely fashion.
- In order to achieve an unbiased sample, the selection has to be random so everyone from the population has an equal and likely chance of being added to the sample group.
- In simple random sampling, every entity in the population is identical, while stratified random sampling divides the overall population into smaller groups.

Derived data involves using existing data points, often from different data sources, to create new data through some sort of transformation, such as an arithmetic formula or aggregation. For example, combining area and population data from the Twin Cities metro area to create population density data. While this type of data can usually be replaced if lost, it may be very time-consuming (and possibly expensive) to do so.:

1. Convenience sampling.
2. Consecutive sampling.
3. Quota sampling.
4. Judgement sampling.
5. Snowball sampling.

### **1. Convenience sampling:**

is probably the most common of all sampling techniques. With convenience sampling, the samples are selected because they are accessible to the researcher. Subjects are chosen simply because they are easy to recruit. This technique is considered easiest, cheapest and least time consuming.

### **2. Consecutive sampling:**

Is very similar to convenience sampling except that it seeks to include ALL accessible subjects as part of the sample. This non-probability sampling technique can be considered as the best of all non-probability samples because it includes all subjects that are available that makes the sample a better representation of the entire population.

### **3. Quota sampling:**

Is a non-probability sampling technique wherein the researcher ensures equal or proportionate representation of subjects depending on which trait is considered as basis of the quota. For example, if basis of the quota is college year level and the researcher needs equal representation, with a sample size of 100, he must select 25 1<sup>st</sup> year students, another 25 2<sup>nd</sup> year students, 25 3<sup>rd</sup> year and 25 4<sup>th</sup> year students. The bases of the quota are usually age, gender, education, race, religion and socioeconomic status.

### **4. Judgement sampling:**

Is more commonly known as purposive sampling. In this type of sampling, subjects are chosen to be part of the sample with a specific purpose in mind. With judgmental sampling, the researcher believes that some subjects are more fit for the research compared to other individuals. This is the reason why they are purposively chosen as subjects.

### **5.Snowball sampling:**

Is usually done when there is a very small population size. In this type of sampling, the researcher asks the initial subject to identify another potential subject who also meets the criteria of the research. The downside of using a snowball sample is that it is hardly representative of the population.

### **Q: 3**

Define and explain Data and types of Data?

### **Ans:**

#### **Data:**

Research data is any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.

Data is also define as

Data are values of the observation recorded for variables ( e.g. age, weight, sex) Research

data can take any form:

- documents, spreadsheets
- laboratory notebooks, field notebooks, diaries
- questionnaires, transcripts, codebooks
- audiotapes, videotapes
- photographs, films
- test responses
- slides, artefacts, specimens, samples
- collections of digital outputs
- data files
- database contents (video, audio, text, images)
- models, algorithms, scripts
- contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- methodologies and workflows
- standard operating procedures and protocols **Types of Data:**

Data may be grouped into four main types based on methods for collection: observational, experimental, simulation, and derived. The type of research data you collect may affect the way you manage that data. For example, data that is hard or impossible to replace (e.g. the recording of an event at a specific time and place) requires extra backup procedures to reduce the risk of data loss. Or, if you will need to combine data points from different sources, you will need to follow best practices to prevent data corruption.

1. Observational Data.
2. Experimental Data.
3. Simulation Data.
4. Derived Data.

### **1.Observational data:**

Observational data are captured through observation of a behavior or activity. It is collected using methods such as human observation, open-ended surveys, or the use of an instrument or sensor to monitor and record information – such as the use of sensors to observe noise levels at the Mpls/St Paul airport. Because observational data are captured in real time, it would be very difficult or impossible to re-create if lost.

### **2.Experimental data:**

Experimental data are collected through active intervention by the researcher to produce and measure change or to create difference when a variable is altered. Experimental data typically allows the researcher to determine a causal relationship and is typically projectable to a larger population. This type of data are often reproducible, but it often can be expensive to do so **2.3.**

### **3.Simulation data:**

Simulation data are generated by imitating the operation of a real-world process or system over time using computer test models. For example, to predict weather conditions, economic models, chemical reactions, or seismic activity. This method is used to try to determine what would, or could, happen under certain conditions. The test model used is often as, or even more, important than the data generated from the simulation.

### **4.Derived data:**

Derived data involves using existing data points, often from different data sources, to create new data through some sort of transformation, such as an arithmetic formula or aggregation. For example, combining area and population data from the Twin Cities metro area to create population density data. While this type of data can usually be replaced if lost, it may be very time-consuming (and possibly expensive) to do so.