INSTRUCTIONS

All trainees are required to read this material before attending RMB & DW workshop so that the learning in the workshop is optimum and does really help in writing the synopsis and dissertation.
CHAPTER 1: AN INTRODUCTION TO RESEARCH

1. RESEARCH:
   A process of systematic and scientific data collection, analysis; & interpretation so as to find solutions to a problem.

2. QUANTITATIVE RESEARCH:
   • Research based on collection of data having clear measurable entity through close ended questions which is amenable to statistical testing.

3. QUALITATIVE RESEARCH:
   • Qualitative research seeks out the ‘why’, not the ‘how’ of its topic through the analysis of unstructured information – things like interview transcripts, open ended survey responses etc. Focus groups, in-depth interviews, content analysis are among the many formal approaches that are used. It doesn’t rely on statistics or numbers, which are the domain of quantitative researchers.

4. RESEARCH OBJECTIVE:
   A researcher’s intent written in clear measurable terms

5. RATIONALE OF THE STUDY:
   Rationale of the study indicates the specific reason for doing the research on the topic. If a topic has previously been researched the Rationale clearly states why then the need for “re-search”. What was missing in or were limitations /draw backs of previous studies and what new will the current research presenting. In the event of a new research the bases of the hypothesis /idea is clearly with references, mentioned in the Rationale section

6. OPERATIONAL DEFINITIONS:
   Means of measurement/determination of the exposure and outcome variables of interest in context to objective in a particular study

7. LITERATURE:
   Written work, books and writings published on a particular subject.

8. LITERATURE SEARCH --- WHY?
   Allows one to search in a purposeful and systematic manner, through a range of literature or information i.e. (journal article, Research projects, conference, congress &meetings proceedings, internet) relevant to ones particular field, and to hone in on material relevant to ones interest and objectives.
CHAPTER 2: RESEARCH AND STUDY DESIGNS

9. STUDY DESIGN
   A study design is a specific plan or protocol for conducting a study, which allows the
   investigator to translate the conceptual hypothesis into an operational one.

10. DESCRIPTIVE STUDIES:
    A study that tries to reveal patterns of a specific disease without an emphasis on pre-
    specified hypotheses looking for association. Sometimes these types of studies are
    called hypothesis generating studies in contrast to hypothesis testing analytical studies.

11. CASE REPORT
    A study of an unusual disease in a single person

12. CASE SERIES:
    A study of several unusual cases all with similar conditions

13. CROSS-SECTIONAL STUDY
    A survey of a defined population at a single point in time or over a specified time also
    called as prevalence study. Examines the relationships between diseases and exposures
    as they exist in a defined population at or over a specified time but cannot establish
    temporal relationship. Helps to determine burden of disease.

14. TEMPORAL RELATIONSHIP:
    The timing of the relationship between a factor and an outcome. It is one of the criteria
    used to assign causality to a relationship. To confirm causality exposure should precede
    outcome.

15. BURDEN OF DISEASE:
    Disease burden is the burden a particular disease process has in a specific area as
    measured by cost, morbidity, and mortality.

16. ANALYTICAL STUDY:
    Done for testing a hypothesis, determines association between a disease (outcome) and
    possible causes of the disease(exposure). Analytical studies always have at least two
    groups.
17. COMPARISON GROUP
A group in an analytic study (e.g., a cohort or case-control study) with whom the primary group of interest (exposed group in a cohort study or case-patients in a case-control study) is compared. The comparison group provides an estimate of the background or expected incidence of disease (in a cohort study) or exposure (in a case-control study).

18. OBSERVATIONAL STUDY:
A study in which the investigator observes rather than influences exposure and disease among participants. Case-control and cohort studies are observational studies.

19. COHORT:
A well-defined group of people who have had a common experience or exposure, who are then studied for the development of new diseases or events.

20. COHORT STUDY (PROSPECTIVE)
The investigator assembles the exposed but not diseased non exposed and non-diseased groups in the present time, collects baseline data on them and then continues to collect data for a period that can last many hours to years.

21. COHORT STUDY (RETROSPECTIVE)
The investigator selects the exposed (but not diseased) and non exposed (and non-diseased) groups from the past records and then studies subsequent development of disease in the groups and thereby determine the associations between the exposure and outcome(s).

22. CASE-CONTROL STUDY
A study that compares people with a specific disease or outcome of interest (cases) to people from the same population without that disease or outcome (controls), to find associations between the outcome and prior exposure to particular risk factors.

23. EXPERIMENTAL STUDY:
A study in which the investigator assigns the exposure for each person (clinical trial) or community (community trial) then assesses the health status to determine the effects of the exposure on the outcome.
24. PREVENTIVE TRIALS:
   Refers to experimental trials done to find better ways to prevent disease in people who
   have never had the disease or to prevent a disease from returning. These approaches
   may include medicines, vaccines, vitamins, minerals, or lifestyle changes.

25. THERAPEUTIC TRIALS:
   Refers to trials which test new treatments, new combinations of drugs, or new
   approaches to surgery or radiation therapy in patients suffering from disease.

26. RANDOMIZED CONTROLLED TRIAL (RCT)
   (Synonym: randomized clinical trial)
   Individuals are randomly allocated to at least two groups. One group is subjected to an
   intervention, while the other group(s) is not.

27. SINGLE-BLIND STUDY
   A clinical trial design in which one party, usually the participant, is unaware of what
   medication assigned. Also called single-masked study

28. DOUBLE-BLIND STUDY
   A clinical trial design in which neither the participating individuals nor the study
   Researcher knows which participants are receiving the experimental drug and which are
   receiving a placebo (or the therapy).

29. TRIPLE-BLIND STUDY
   A clinical trial design in which participants, investigators, and data analyst all remain
   unaware of the intervention assignments throughout the trial.

30. PLACEBO
   Refers to a pharmacologically inactive agent that investigators administer to participants
   in the control group of a trial
CHAPTER 3: SAMPLE SELECTION & SAMPLING PROCEDURE

31. SAMPLE
A sample is a sub set of the population, with all its inherent qualities. Inferences about the population can be made from the measurements taken from a sample, if the sample is truly representative of the population.

32. SAMPLING TECHNIQUES
Methods to enroll subjects/participants in studies so as to ensure validity and generalizability.

33. PROBABILITY SAMPLING TECHNIQUE
A sampling technique in which all units/participants have an equal chance to be included in the sample, this type of technique is amenable to statistical testing.

TYPES OF PROBABILITY SAMPLING

34. SIMPLE RANDOM
A type of probability sampling technique in which elements are selected into the sample randomly by using a random numbers table or a computer-generated random sequence or a lottery method when the sampling frame is available. Prerequisite in this type of sampling is a sampling frame.

35. SAMPLING FRAME
The list of all possible units/participants that might be drawn in a sample.

36. SYSTEMATIC SAMPLING TECHNIQUE
A type of probability sampling technique which require a sampling frame (if available). however it can be done without the sampling frame too. All one requires is an estimate sample size and estimated population size to determine

\[ K = \frac{N}{n} \]

Following which every Kth patient/respondent is included in the sample.

37. STRATIFIED SAMPLING
A type of probability sampling technique which aims to produce a sample that is representative of different strata in a given population. E.g. contains representation based on ethnicity, gender etc.
38. **CLUSTER SAMPLING:**
When a list of the entire area is not available and it is not physically possible to visit the entire area (e.g. the city, or country) one can divide the area into several equal size clusters or units. e.g.: Mohallas, Apartment Buildings, Villages, Schools One can select (randomly) only a few clusters, number all the units within it and draw either:

1. A random sample or
2. A systematic sample

39. **NON PROBABILITY SAMPLING TECHNIQUE**
A sampling technique in which there is no assurance that each element will have the same chance of being included in the sample

40. **CONVENIENCE SAMPLING**
It is the process of taking those members of the accessible population who are easily available. Sample is selected in a haphazard fashion. This type of sampling technique is fraught with biases.

41. **CONSECUTIVE SAMPLING:**
It involves taking every patient who meets the selection criteria over a specified time interval or number of patients. It is the best of the non probability techniques and regarded by some as a facsimile of probability sampling.
CHAPTER 4: VARIABLES, DATA AND ITS PRESENTATION

42. VARIABLE
   A Variable is a characteristic of a person, object or phenomenon that can take on different values.

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

44. DEPENDENT VARIABLE:
   The variable that is used to describe or measure the problem under study (outcome) is called the dependent variable.

45. INDEPENDENT VARIABLE:
   The variables that are used to describe or measure the factors that are assumed to cause or at least to influence the problem are called the Independent (exposure) variables.

46. QUALITATIVE VARIABLE:
   Qualitative variable are those that express a qualitative attribute such as hair color, ethnicity, religion, gender, and so on. Qualitative variables are sometimes referred to as categorical variables, as they categorize the qualitative attribute. e.g. Male/Female, Good/Bad

47. QUANTITATIVE VARIABLE:
   Quantitative variables are those variables that are measured in terms of numbers. Some examples of quantitative variables are height, weight, and shoe size.

   TYPES OF QUANTITATIVE VARIABLES

48. NOMINAL SCALE:
   A measurement scale consisting of qualitative categories whose values have no inherent statistical order or rank (e.g., categories of race, ethnicity, religion, or country of birth).

49. ORDINAL SCALE:
   A measurement scale consisting of qualitative categories whose values have a distinct order but no numerical distance between their possible values (e.g., stage of cancer, I, II, III, or IV).
PRESENTATION OF DATA

QUALITATIVE DATA

50. FREQUENCY TABLE:
   In statistics, a frequency distribution is an arrangement of the values that one or more variables takes in a sample. Each entry in the table contains the frequency or count of the occurrences of values within a particular group or interval, and in this way, the table summarizes the distribution of values in the sample.

51. CHARTS:
   A chart is a graphical representation of data, in which "the data is represented by symbols, such as bars in a bar chart, lines in a line chart, or slices in a pie chart.

52. GRAPHS:
   A diagram showing the relation between typically two variable quantities, each measured along one of a pair of axes at right angles.

53. PIE CHART:
   A pie chart (or a circle graph) is a circular chart divided into sectors, illustrating proportion. In a pie chart, the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents.

QUANTITATIVE DATA

54. HISTOGRAM:
   A graphical representation, of a quantitative variable similar to a bar chart but without any space between bars, that organizes a group of data points into user-specified ranges, and also gives information on Mean and SD.
CHAPTER 5: SUMMARIZATION OF DATA

55. MEASURES OF CENTRAL TENDENCY:
   Measure of central tendency, of a data set is a measure of the "middle" value of the data set. Many different descriptive statistics can be chosen as a measure of the central tendency of the data items. These include the arithmetic mean, the median, and the mode.

56. MEAN:
   The measure of central location, commonly called the average, calculated by adding all the values in a group of measurements and dividing by the number of values in the group.

57. MEDIAN:
   Median the measure of central location that divides a set of data into two equal parts, above and below which lie an equal number of values. To calculate median first ordering of data in ascending or descending order is required.

58. MODE:
   The value of a numerical set that appears with the greatest frequency.

59. MEASURES OF VARIATION:
   Measure of variation is a measure that describes how spread out or scattered a set of data is. It is also known as measures of dispersion or measures of spread. There are three measures of variation: The range, the variance, and the standard deviation.

60. RANGE:
   In statistics, the difference between the largest and smallest values in a distribution; in common use, the span of values from smallest to largest.

61. VARIANCE:
   A measure of the spread in a set of observations, calculated as the sum of the squares of deviations from the mean, divided by the number of observations minus 1.

   Variance \( (S^2) = \sum (X - \mu)^2 / (N - 1) \)

   Where,
   \[ X = \text{individual observation} \]
   \[ \mu = \text{Mean of sample} \]
62. **STANDARD DEVIATION:**

It is the square root of the variance. The standard deviation is a measure, which describes how much individual measurements differ, on the average, from the mean.

\[ \text{Variance} = S^2 = \sum (x_i - \text{sample mean})^2 / (n - 1) \]

Where \( x_i \) = Individual sample observation
\( n \) = Total sample size

63. **SAMPLING VARIATION:**

When we draw a sample from study population and compute its sample mean it is not likely to be identical to the population mean. If we draw another sample from same population and compute its sample mean, this may also not be identical to the first sample mean. It probably also differs from the true mean of the total population from which the sample was drawn this phenomenon is called sampling variation.

64. **STANDARD ERROR:**

The standard error gives an estimate of the degree to which the sample mean varies from the population mean and this measures is used to calculate CI.

65. **CONFIDENCE INTERVAL:**

A confidence interval (CI) is a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is a range of values so defined that there is a specified probability that the value of a parameter lies within it.

66. **CONFIDENCE LEVEL:**

The confidence level of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter. The commonly used confidence levels are 95% and 99%.

67. **THE NORMAL DISTRIBUTION:**

This is a bell shaped curve with most of the values clustered near the mean and a few values out near the tails.

68. **BELL SHAPED CURVE:**

It is the frequency curve that resembles the outline of a bell, as a normal curve. It is the symmetrical curve of normal distribution.
CHAPTER 6: ESTIMATION AND HYPOTHESIS TESTING

70. ESTIMATION:
The process of using sample information to draw conclusion about the value of a population parameter is known as estimation.

71. POINT ESTIMATE:
A specific numerical value estimate of a parameter.

72. INTERVAL ESTIMATE:
An interval estimate of a parameter is an interval or a range of values used to estimate the parameter also known as CI.

73. HYPOTHESIS:
A testable theory, or statement of belief used in evaluation of a population parameter of interest e.g. Mean or proportion.

74. NULL HYPOTHESIS (H₀):
A proposition that undergoes verification to determine if it should be accepted or rejected in favor of an alternative proposition. Often the null hypothesis is expressed as "There is no relationship between two quantities.

75. ALTERNATE HYPOTHESIS:
In hypothesis testing, an alternate hypothesis a proposition that is accepted if the null hypothesis is rejected. There is an association between or among study variables.

76. SIGNIFICANCE LEVEL:
The significance level is the risk we are willing to take that a sample which showed a difference was misleading. 5% significance level means that we are ready to take a 5% chance of wrong results.

77. P-VALUE
The probability (ranging from zero to one) that the results alone observed in a study (or results more extreme) could have occurred by chance if in reality the null hypothesis was true. P Value: indicates the probability or likelihood of obtaining a result at least as extreme as that observed in a study by chance alone, assuming that there is truly no association between exposure and outcome under consideration. By convention the p value is set at 0.05 level. Thus any value of p less than or equal to 0.05 indicates that there is at most a 5% probability of observing an association as large or larger than that found in the study due to chance alone given that there is no association between exposure and outcome. If p value is > 0.05 do not reject the null hypothesis.
78. STEPS OF HYPOTHESIS TESTING:
   a. Statement of research question in terms of statistical hypothesis (null and alternate hypothesis)
   b. Selection of an appropriate level of significance.
   c. Choosing an appropriate statistics e.g t test (if the sample size is 30 or more then 30.), z test (if the sample is more than 30) for continuous data when groups are two and chi square for proportions etc.
   d. Generate p value using test statistics.
   e. Drawing conclusions, based on p value (reject in null hypothesis if the p value is less than the set significance level).

79. TYPE I ERROR (ALPHA ERROR)
   Mistakenly rejecting the null hypothesis when it is actually true

80. TYPE II ERROR (BETA ERROR)
   Mistakenly accepting (not rejecting) the null hypothesis when it is false

81. POWER
   The probability of rejecting null hypothesis when it is false. In other words, it is the probability of avoiding a type II error.
CHAPTER 7: MEASURES OF DISEASE FREQUENCY

82. RATIO:
   It is obtained by simply dividing one quantity by another without implying any specific relationship between the numerator and denominator. In ratio, the numerator & denominator are mutually exclusive.

83. PROPORTION:
   A proportion is a type of ratio in which those who are included in the numerator must also be included in the denominator.

84. RATE:
   A rate is a proportion with specifications of time. There is a distinct relationship between the numerator and denominator with a measure of time being an intrinsic part of the denominator.

85. PREVALENCE:
   Quantifies the proportion of individuals in a population who have the disease at a specific instant or period calculated as number of existing cases of a disease/ total population at a given point in time or over a specified period.

86. POINT PREVALENCE:
   The status of the disease in a population at a point in time and as such is also referred to as point prevalence.

87. PERIOD PREVALENCE:
   It represents the proportion of cases that exist within a population during a specified period of time.

88. INCIDENCE:
   It quantifies the number of new events or cases of disease that develop in a population of individuals at risk during a specified time interval.
CHAPTER 8: MEASURES OF ASSOCIATION

89. CORRELATION COEFFICIENT:
An association between two variables can be measured by correlation coefficient ‘r’ which is a measure of degree to which a dependent variable varies with an independent variable. ’r’ varies between +1 and –1.

90. RELATIVE RISK (RR):
Measures the strength of association between exposure and outcome and is calculated as the incidence of the outcome in the exposed group (Ie) divided by the incidence of the outcome in the unexposed group. It is used for Cohort studies.

Relative Risk = \( \frac{\text{Incidence in Exposed}}{\text{Incidence in Unexposed}} \)

Relative Risk = \( \frac{a}{(a+b)} \cdot \frac{c}{(c+d)} \)

91. ODDS RATIO:
Measures the association between disease and exposure, calculated as a ratio of the odds of a case group being exposed divided by the odds of a control group being exposed. Calculated in a case control studies, where incidence cannot be calculated as individual enrolled are diseased as well as exposed.

Odds Ratio = \( \frac{ad}{bc} \)
CHAPTER 9:  
FACTOR AFFECTING STUDY OUTCOMES  
BIAS, CONFOUNDING, CHANCE

93. BIAS:  
Any systematic error that results in an incorrect estimate of the association between the exposure and outcome. It is usually introduced by the experimenter or the researcher himself due to non-standardized measuring techniques. These are bias, confounder & chance

TYPES OF BIASES

94. SELECTION BIAS:  
Systematic difference in the enrollment of participants in a study that leads to an incorrect measure (e.g., relative risk or odds ratio) or inference.

95. RECALL BIAS:  
Systematic error due to differences in accuracy or completeness of recall to memory of past events or experiences.

96. INTERVIEWER’S BIAS:  
The influence of the interviewer on the interviewee. This may result from several factors, including the physical and psychological characteristics of the interviewer, poor questionnaire, and inadequate interviewer training which may cause differential responses from interviewees.

97. LOST TO FOLLOW UP BIAS:  
If a patient can no longer be followed for the outcome of interest, e.g., a patient who is unwilling or unable to return to the clinic for follow-up examinations, or a patient who cannot be located for subsequent follow-up and as a result the outcome cannot be measured leads to bias. The bias resulting is referred to as Lost to follow up bias.

98. CONFOUNDING:  
Confounding can be thought of as mixing of the effect of the exposure under study on the disease with that of an extraneous factor. This external factor or variable must be associated with the exposure and, independent of the exposure must be a risk factor for the disease.
99. EFFECT MODIFIER:
Effect modifiers are variables that bring about a change in the magnitude of an effect. Unlike confounders, effect modifiers do not require to be related to both exposure and outcome variable.

100. CHANCE:
Effect of chance can be determined by calculation of:
   a. Confidence interval
   b. Hypothesis testing.
CHAPTER 10: SAMPLE SIZE CALCULATION

101. SAMPLE SIZE ESTIMATION:
A process to estimate the number of subjects/participants required to achieve objectives of a particular study. It is a prerequisite for the validity and reliability of the outcome of epidemiological studies.

Prerequisite for sample size calculation depends on type of study designs

<table>
<thead>
<tr>
<th>DEScriptIvE sTUDIES</th>
<th>ANALYTICAL sTUDIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnitude of the outcome of interest derived from previous studies i.e. prevalence/mean±SD/Sensitivity &amp; specificity of the test</td>
<td>Magnitude of the outcome of interest derived from previous studies for both groups</td>
</tr>
<tr>
<td>Level of confidence or significance</td>
<td>Level of significance</td>
</tr>
<tr>
<td>Margin of error</td>
<td>Power of the test</td>
</tr>
</tbody>
</table>

102. SIGNIFICANCE LEVEL:
The significance level is the risk we are willing to take that a sample which showed a difference was misleading. 5% significance level means that we are ready to take a 5% chance of wrong results.

103. MARGIN OF ERROR (D):
The margin above and below the estimated prevalence that we want our sample size to capture.

104. PRECISION:
Precision is a measure of how close an estimator is expected to be to the true value of a parameter.

105. POWER:
The probability of rejecting null hypothesis when it is false. In other words, it is the probability of avoiding a type II error and is denoted by 1-β. The commonly used power values are 80%, 90%, 95% and 99%. The more powerful a study the less chance of missing out a difference between group if one actually exists.
CHAPTER 11: SCREENING TEST

106. SCREENING:
Screening for disease control can be defined as the examination of asymptomatic people in order to classify them as likely or unlikely to have the disease that is objective of screening.

107. VALIDITY:
The ability of a test to distinguish between who has disease and who do not.

108. SENSITIVITY:
It is the ability of a test to detect people who do have disease. If a Test is always positive for all diseased persons then sensitivity of the Test will be 100%.

109. SPECIFICITY:
It is the ability of a Test to detect people who don’t have disease. Thus a Test which is always negative in non- diseased individuals with have 100% specificity.

110. POSITIVE PREDICTIVE VALUE:
The positive predictive value, or precision rate is the proportion of subjects with positive test results who are correctly diagnosed. It is a critical measure of the performance of a diagnostic method.

\[
PPV = \frac{a}{a + b} \times 100
\]

111. NEGATIVE PREDICTIVE VALUE:
The negative predictive value (NPV) is a summary statistic used to describe the performance of a diagnostic testing procedure. It is defined as the proportion of subjects with a negative test result who are correctly diagnosed.

\[
PPV = \frac{d}{c + d} \times 100
\]
CHAPTER 12: QUESTIONNAIRE DEVELOPMENT

112. QUESTIONNAIRE:
A set of printed or written questions with a choice of answers, designed for the purpose of seeking specific information from the respondent.

113. TYPES OF QUESTIONNAIRE:
1. Self administrated---------------------Answered by respondents themselves
2. Administered by interviewers----------------Questions asked by interviewer

114. TYPES OF QUESTIONS
1. Closed ended questions
2. Open ended questions

115. CLOSED ENDED QUESTIONS
In a closed ended question the responded is provided with a list of pre-determined response option. They are more common and form the basis for most standardized measures. These questions asked respondents to chose from pre selected answers

116. OPEN ENDED QUESTIONS
Open ended question elicit detailed responses and provide no preselected options. These type of questions are the hallmark of qualitative research.

117. LIKERT SCALES
It is difficult to quantitatively assess abstract concepts, such as quality of life, from single questions. Therefore abstract characteristics are commonly measured by generating score from a series of questions that are organized in to a scale e.g. strongly agree, agree, disagree and strongly disagree. Likert scales are commonly used to quantify attitudes behavior and domains of health related quality of life these scales provide respondents with a list of statements or questions and asked them to select a response that the best represents the rank of their answer.