DANTULURI NARAYANA RAJU COLLEGE(A)
BHIMAVARAM
DEPARTMENT OF PG MICROBIOLOGY

STUDY MATERIAL
SEMESTER-IV
MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

DEPARTMENT OF PG MICROBIOLOGY

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

RESEARCH METHODOLOGY

Research methodology is a crucial aspect of any scientific inquiry, providing a structured framework for conducting systematic investigations and acquiring knowledge. It encompasses the principles, procedures, and techniques that guide researchers in their quest to answer questions, test hypotheses, and explore phenomena. This essay explores the essential componentsof research methodology, its significance in various disciplines, and the key methods employed.

Introduction to Research Methodology

Research methodology refers to the systematic process of planning, conducting, and analyzing research studies. It involves defining the research problem, selecting appropriate methods, collecting and analyzing data, and drawing valid conclusions. A robust methodology ensures thatresearch findings are reliable, valid, and replicable, contributing to the advancement of knowledge in diverse fields such as science, social sciences, humanities, and business.

Components of Research Methodology

1. Research Design

The research design outlines the overall strategy and structure of the study. It specifies whether the research is exploratory, descriptive, experimental, or correlational. The choice of design depends on the research questions, objectives, and nature of the phenomena under investigation.

2. Research Methods

Research methods encompass the techniques and procedures used to collect and analyze data. They can be broadly categorized into qualitative and quantitative methods:

- **Qualitative Methods:** These methods focus on understanding phenomena through in-depth exploration of meanings, experiences, and perceptions. Techniques include interviews, focus groups, and ethnographic observations.

- **Quantitative Methods:** Quantitative research involves measurement and statistical analysis of numerical data. Surveys, experiments, and statistical modeling are common quantitative methods used to test hypotheses and examine relationships between variables.

3. Data Collection

Data collection involves gathering information relevant to the research question. Depending on the methodology, data can be collected through surveys, interviews, observations, experiments, archival research, or a combination of these methods. The quality and reliability of data collection methods significantly impact the validity of research findings.

4. Data Analysis

Data analysis is the process of examining and interpreting collected data to derive meaningful conclusions. Qualitative data are analyzed through thematic analysis, content analysis, or grounded theory, while quantitative data are analyzed using statistical techniques such as descriptive statistics, inferential statistics, and regression analysis.

5. Validity and Reliability

Validity refers to the extent to which research accurately measures what it intends to measure. Reliability indicates the consistency and replicability of research findings. Researchers employ various strategies, such as triangulation (using multiple methods or data sources) and pilot testing, to enhance validity and reliability.

Significance of Research Methodology

Research methodology is critical for several reasons:

- **Ensures Rigor:** A well-defined methodology ensures that research is conducted systematically, minimizing biases and errors.
- **Facilitates Replication:** Clear methodologies enable other researchers to replicate studies, validating findings and building cumulative knowledge.

- **Ethical Considerations:** Methodologies include ethical guidelines for conducting research with integrity and respect for participants' rights and welfare.
- **Informs Decision-Making:** Research findings influenced by robust methodologies provide evidence-based insights for policy-making, practice, and further research.

Application Across Disciplines

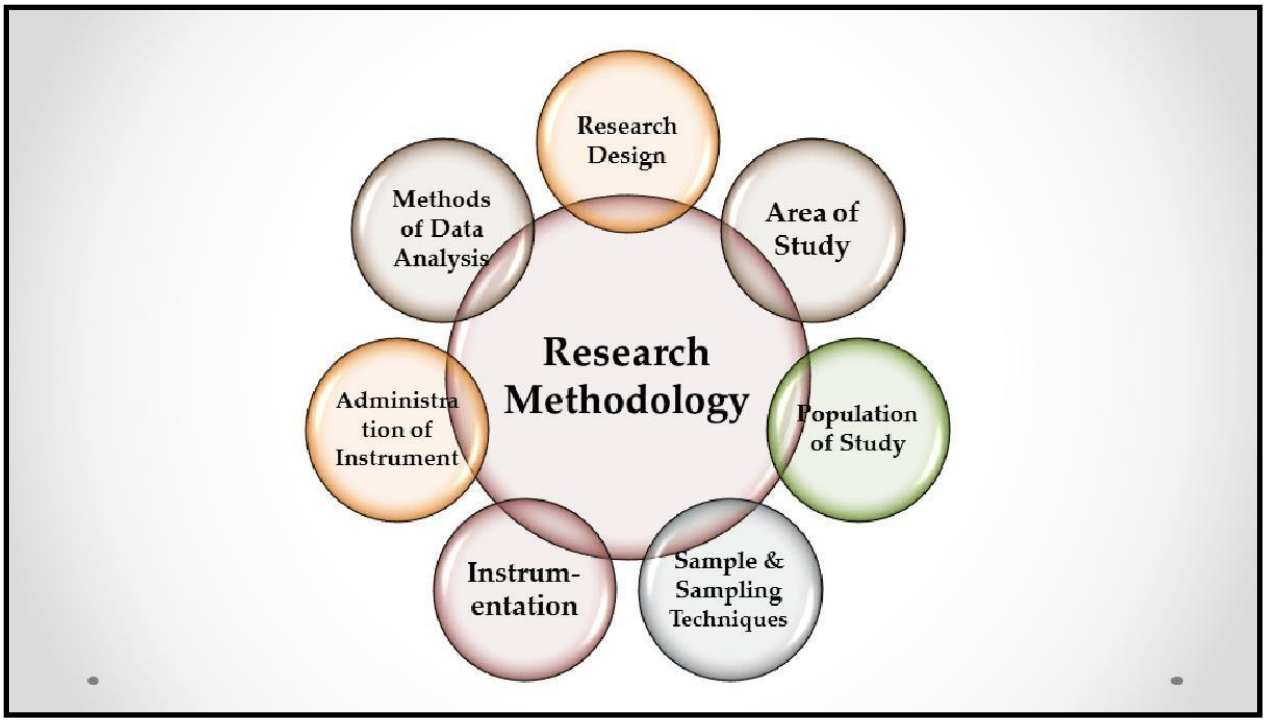
Research methodology is adaptable across diverse disciplines:

- **Natural Sciences:** Experimental designs and quantitative methods are prevalent in disciplines like biology, chemistry, and physics to test hypotheses and explore natural phenomena.
- **Social Sciences:** Qualitative methods such as interviews and case studies are used in psychology, sociology, and anthropology to understand human behavior and social dynamics.
- **Humanities:** Research in history, literature, and philosophy often employs archival research, textual analysis, and qualitative methods to interpret cultural artifacts and historical texts.
- **Business and Management:** Quantitative methods such as surveys and statistical analysis are utilized in marketing research, finance, and operations management to analyze consumer behavior and organizational performance.

Conclusion

Research methodology serves as the foundation for rigorous and systematic inquiry across academic and professional domains. By delineating research designs, methods, data collection, and analysis techniques, it ensures the reliability, validity, and ethical integrity of research endeavors. Embracing diverse methodologies allows researchers to approach complex questions from multiple perspectives, fostering innovation and advancing knowledge. As disciplines evolve and new challenges emerge, the ongoing refinement and application of research methodologies remain essential for addressing global issues and driving progress in science and society.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY



MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

LEVELS OF SIGNIFICANCE,STUDENT T-TEST

Levels of Significance:

- In biostatistics, the level of significance in a student's t-test is a critical value used to determine whether the results of the t-test are statistically significant. The level of significance, often denoted by alpha (α), is the probability of rejecting the null hypothesis when it is actually true.
- Commonly used levels of significance include 0.05 (5%) and 0.01 (1%). If the p-value obtained from the t-test is less than the chosen level of significance (e.g., $p < 0.05$), then the results are considered statistically significant. This indicates that there is strong evidence to reject the null hypothesis in favor of the alternative hypothesis.
- Choosing an appropriate level of significance is essential in hypothesis testing as it helps determine the strength of evidence required to make conclusions about the data.

Student's T-test:

Use this test to compare two small sets of quantitative data when Samples are collected independently of one another. Student's T-test is One of the most commonly used Techniques for testing a hypothesis on the basis of a difference between Sample means**.**

$$t = m-u/ s/\sqrt{n}$$

Where t student's t-test

m mean

u = theoretical dev value

8 = standard deviation1

n = variable set size.

- The student's t-test is widely used when the sample. Bize is reasonably small. (less than approximately 30).
- In these Cares the sample distribution of the mean is known to follow a distribution.
- A student's t-tests are Parametric tests based on the Student's or t-distribution.
- The student t-test Compares the mean of a data Set of Mean a new ar of a modified assay to the sample reference assay.
- The Student's i-test distribution is named in honor of william Sealy Gossel, who first determined it in 1908.
- The Conditions required to conduct a t-test include the measured values in gratio scale or intervalscale, simple random Extraction, homogeneity of variance, appropriate sample size, and normal distribution of data. AZ

- A one-sample location test of whether the mean of a Population has a value specified in a null hypothesis.
- A two-sample location test of the null hypothesie Such that the means of two Populations are Equal All such tests are called as student's t-tests.

Paired :

A Paired t-test are also known as dipendent or Correlated t-test.

The test that composed the mean ans standard deviation of 2 related groups to determine segnificance defference.

2 typer - null, "ternative.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

METHODOLOGY IN WRITING A RESEARCH PAPER FORA SCIENTIFIC JOURNAL

Contents:

- What is a research paper
- Goals in writing a research paper
- Steps in writing a research paper
- Structure of a research paper

Steps in writing a research paper:

Criteria for journal selection:

- Referring system
- Citation scores
- Circulation
- Journal type
- Time lag
- Reputation of editors
- Professional vs. commercial ownership
- Quality of production

DEPARTMENT OF PG MICROBIOLOGY

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Structure of a research paper

While academic disciplines vary on the exact format and style of journal articles in their field, most articles contain similar content and are divided in parts that typically follow the same logical flow. Following is a list of the parts commonly found in research articles.

- ✓ Title
- ✓ Abstract
- ✓ Introduction
- ✓ Literature Review
- ✓ Methods
- ✓ Results
- ✓ Discussion/Conclusion
- ✓ References

Title

The title should be specific and indicate the problem the research project addresses using keywords that will be helpful in literature reviews in the future.

- Clear and explanatory title
- Brief and concise

Abstract

- The abstract is used by readers to quickly review the overall content of the paper.
- Purpose of the study – hypothesis, overall question, objective.
- Results, including specific data – if the results are quantitative in nature, report quantitative data.
- Important conclusions or questions that follow from the experiment.

Style:

- Single paragraph, and concise
- As a summary of work done, it is always written in past tense
- Focus on summarizing results
- Correct spellings, clarity of sentences and phrases, and proper reporting of quantities

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Introduction

The introduction begins by introducing the broad overall topic and providing basic background information. It then narrows down to the specific research question relating to this topic.

- Describe the importance of the study
- Provide a rationale.
- Very briefly describe the experimental design and how it accomplished the stated objectives.

Literature Review

- Establishes context of the study by providing a brief and balanced review of pertinent published literature available on subject.
- General to specific
- Name – date citations

Method

- Study design
- Participants and their characteristics
- When and where study conducted
- Sampling method and size
- Variables measured
- Method of collecting data

Results

- Objective presentation of key findings without interpretation
- Orderly, logical sequence using text and illustrations (figures/tables)

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Tips in writing results:

- Do not reiterate each value from the table – only mention salient points
- Do not present the same data in table and figures
- Do not report raw data values when they can be summarized as means, percent, etc.
- Statistical test and values should be mentioned
- Always mention appropriate units with data
- Only present percentages to one decimal point

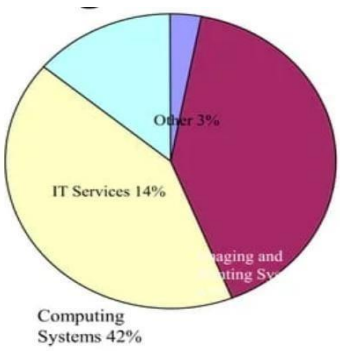
Figures and tables:

- Either place figures and tables within the text of the result.
- If you place figures and tables at the end of the report, make sure they are clearly distinguished from any attached appendix materials, such as raw data.
- Regardless of placement, each figure must be numbered consecutively and complete with caption
- Regardless of placement, each figure must be numbered consecutively and complete with heading

Tables

Grade	Number of viewing hours	Reading level
First Grade	5-10 hours	2.8
Second Grade	16-20 hours	2.6
Third Grade	11-15 hours	4.2

Figures



Pie chart of total sales

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Discussion

- Answers questions posed in introduction
- Interprets results in comparison to what is already known on the subject
- Explains new understanding of the subject based on present results
- It tells the readers how present study moved from what was known to what is new.
- Outlines limitations of the study
- Discusses ideas for future research

Conclusion

- Summary
- Recommendations

References

- Complete citations for research cited
- References are listed in alphabetical order by the first author's last name
- Citations according to style manual, e.g., APA, MLA, Chicago, etc

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Basic research

Basic research is defined as “a systematic study directed toward greater knowledge or understanding of the fundamental aspects of phenomena and of observable facts without specific applications towards processes or products in mind.”

This type of research contributes to the intellectual body of knowledge. Basic research is concerned with the generalization of a theory in a branch of knowledge;its purpose is usually to generate data that confirm or refute the initial thesis of the study.
It can also be called foundational research; many things get built on thisfoundation, and more practical applications are made.

Basic Research vs. Applied Research

Basic Research finds its counterpart and complement in applied research. They aretwo handy research methods when generating and giving a utility to the generated data. There are very marked differences, and understanding them will allow you tounderstand the path followed to create new knowledge.

The most important difference between basic research and applied research lies inthe objective of each. It seeks to expand the information and understanding of theobject of study, while applied research aims to provide a solution to the problem studied.

The relationship between these two types of research is usually very close since the methodologies used are often quite similar; the significant change is found in the initial and final point of the investigation.

Basic Research Examples

There can be many examples of basic research; here are some of them:

- A study of how stress affects labor productivity.
- Studying the best factors of pricing strategies.
- Understand the client's level of satisfaction before certain interactions withthe company providing solutions.
- The understanding of the leadership style of a particular company.

ADVANTAGES & DISADVANTAGES

Basic research is critical for expanding the pool of knowledge in any discipline. The introductory course usually does not have a strict period, and the researcher'sconcern commonly guides them. The conclusion of the fundamental course is generally applicable in a wide range of cases and plots.

At the same time, the basic study has disadvantages as well. The findings of this type of study have limited or no constructive conclusions. In another sense, fundamental studies do not resolve complex and definite business problems, but itdoes help you understand them better.

Taking actions and decisions based on the results of this type of research will increase the impact these insights may have on the problem studied if that is thepurpose.

ORAL PRESENTATION

An important aspect to any research project is the oral presentation of the experiment to other people. As with a research report, you want to tell the story ofyour experiment: why the experiment was done, how it was done, the results, interpretation of the results, and why the experiment matters.

However, a good presentation is different from a good paper. The presentation should not consist of simply reading from a paper that was previously prepared. Care should be taken to not overwhelm the listener with needless detail. Much

more detailed information can be presented and understood in a written paper thanin an oral presentation.

The style of a presentation is also important. The presenter must try to keep thelistener focused on the key information that is being conveyed.

The following are specific things that should be considered when preparing an oral presentation.

ORGANIZATION

Oral presentations should be organized to have introduction, body and conclusionsections.

Introduction

This section should be brief. It should provide enough background information so that the listener understands the general hypothesis and why the experiments were done. It should also state the specific research question that was studied.

Body

This section is the major portion of the talk. It should include research methods aswell as research results. The methods should be briefly stated, providing detail when necessary for understanding a particular result.

Conclusion

This section should also be brief. A clear, concise statement of what the resultsprove should be made. The data can be related to experiments others have performed, but this should not be overdone. Future experiments to test unanswered questions could be suggested. State why this experiment matters.

Presentation Style

The following are things that should be considered when designing a presentation.

Time

Pay attention to the time. Most research talks are short and no more than 15minutes.

Pace

Do not talk too quickly. Slow down so that the listener has time to hear you.

- If you think you are speaking too slowly, then you probably are going at theright pace.

Volume/Tone

Talk loud enough so that your listener can hear you. Use a variety of voiceinflections and pitches so that the listener stays interested.

- Nothing is more boring than a monotone presentation.
- Alterations in volume/tone gives the listener the feeling that the presenter is interested in the topic.

Eye Contact

Try to maintain eye contact with the listener; this helps them stay focused on thetalk.

- DO NOT SIMPLY READ YOUR PAPER! Whether or not you are presenting from notes, a fully prepared script, or from memory, eye contactmust be made frequently.
- Face the audience: DO NOT READ OFF OF YOUR SLIDES! You aretalking to the people so look at them.

Poise

The presentation should be made in a formal, professional manner.

- Dress appropriately.
- Maintain good, erect posture
- Refrain from informal speech patterns and actions.
- Minimize unnecessary movements such as excessive walking, hand motions,etc.
- Keep your hands out of your pockets

Visual Aids

In general, all research presentations need some sort of visual aid. This is mostoften done using PowerPoint.

- Graphs, tables, photographs etc. of data help the listener sort through the material.
- Complex methods can be presented clearly through visuals.
- A list of conclusion statements helps the listener focus on the final statement.
- A clearly stated research question when visually presented helps.
- Be sure that the visuals are not too complicated. Include only the information you will be discussing.
- Be sure the visual is large enough to be clearly seen by the listener.
- Point to the visuals during the presentation
- Leave the visual up long enough so that the listener can assimilate it.

Present Information Clearly

The information in a presentation should be organized logically and clearly in away that the listener can understand and follow.

- Use of visuals helps here.
- Details should be included when they are important in reaching a particular conclusion. They should be omitted when they get in the way of seeing a particular point.
- Remember: it is not what you say that is important, it is what the listenerhears, understands, and takes with him/her that is important.

Subject Knowledge

The presenter should demonstrate that he/she understands the subject beingpresented. This is done by:

- presenting accurate information,
- by responding to controversies in an appropriate way,
- by answering reasonable questions from the audience.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

CHI SQUARE

- The chi-square test is a statistical tool used to check if two categorical variables are related or independent.
- It helps us understand if the observed data differs significantly from the expected data.
- By comparing the two datasets, we can draw conclusions about whether the variables have a meaningful association.
- The purpose of this test is to determine if a difference between observed data and expected data is due to chance, or if it is due to a relationship between the variables you are studying.
- **Formula**
- The chi-squared test is done to check if there is any difference between the observed value and expected value. The formula for chi-square can be written as;

- **Properties**
- The following are the important properties of the chi-square test:
- Two times the number of degrees of freedom is equal to the variance.
- The number of degree of freedom is equal to the mean distribution
- The chi-square distribution curve approaches the normal distribution when the degree of freedom increases.
- **Chi-Square Distribution**
- When we consider, the null speculation is true, the sampling distribution of the test statistic is called as **chi-squared distribution**.
- The chi-squared test helps to determine whether there is a notable difference between the normal frequencies and the observed frequencies in one or more classes or categories. It gives the probability of independent variables.

- P stands for probability here. To calculate the p-value, the chi-square test is used in statistics. The different values of p indicates the different hypothesis interpretation, are given below:
- P≤ 0.05; Hypothesis rejected
- P>.05; Hypothesis Accepted
- **Question:**
- A survey on cars had conducted in 2011 and determined that 60% of car owners have only one car, 28% have two cars, and 12% have three or more. Supposing that you have decided to conduct your own survey and have collected the data below, determine whether your data supports the results of the study.
- Use a significance level of 0.05. Also, given that, out of 129 car owners, 73 had one car and 38 had two cars.
- **Solution:**
- Let us state the null and alternative hypotheses.
- H_0: The proportion of car owners with one, two or three cars is 0.60, 0.28 and 0.12 respectively.
- H_1: The proportion of car owners with one, two or three cars does not match the proposed model.
- A Chi-Square goodness of fit test is appropriate because we are examining the distribution of a single categorical variable.

One car	73	0.60x129=77.4	_4.4	19.36	0.2501
Two cars	38	0.28x129=36.1	1.9	3.61	0.1
Three or more cars	18	0.12x129=15.5	2.5	6.25	0.4032
Total	129				0.7533

Therefore, $\chi^2 = \sum(O_i - E_i)^2/E_i = 0.7533$

Let's compare it to the chi-square value for the significance level 0.05.

The degrees for freedom = 3 – 1 = 2

Using the table, the critical value for a 0.05 significance level with df = 2 is 5.99.

That means that 95 times out of 100, a survey that agrees with a sample will have a χ^2 value of 5.99 or less.

The Chi-square statistic is only 0.7533, so we will accept the null hypothesis.

BAYES THEOREM:

- Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability.
- Bayes theorem is also known as the formula for the probability of "causes".
- For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

- Bayes Theorem Derivation
- Bayes Theorem can be derived for events and random variables separately using the definition of conditional probability and density.
- From the definition of conditional probability, Bayes theorem can be derived for events as given below:
- $P(A|B) = P(A \cap B)/ P(B)$, where $P(B) \neq 0$
- $P(B|A) = P(B \cap A)/ P(A)$, where $P(A) \neq 0$
- Here, the joint probability $P(A \cap B)$ of both events A and B being true such that,
- $P(B \cap A) = P(A \cap B)$
- $P(A \cap B) = P(A \mid B)\, P(B) = P(B \mid A)\, P(A)$
- $P(A|B) = [P(B|A)\, P(A)]/ P(B)$, where $P(B) \neq 0$
- Similarly, from the definition of conditional density, Bayes theorem can be derived for two continuous random variables namely X and Y.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Experimental Record And Budgetary Report

The Final Report covers the duration of the grant and is likely to require information about key achievements against the program and project objectives,any issues encountered and how they were addressed, and important metrics onoutputs and outcomes.

Here are the top tips to follow while writing a grant report:

1. Know the reporting requirements.
2. Start the report by thanking the funder.
3. Provide the right details in the report.
4. Demonstrate accountability.
5. Invite questions and feedback.

Determine the eligible expense categories and maximum amount allowed by the sponsor. Adjust scope of the project to make sure proposed activities fit within theallowance. Categorize these costs (e.g., salaries, supplies, equipment…) per year, in some cases by quarter. Ensure that project scope and budget match.

BASIC COMPONENTS OF A RESEARCH BUDGET

A research budget contains both direct costs and indirect costs (overhead), but the level of detail varies from sponsor to sponsor. The first step in developing a budgetis to carefully read the guidelines of the funding opportunity being pursued.

There is no magic formula available for developing a budget but there are somebasic steps to follow in order to develop an accurate budget:

1. Define project tasks, timelines and milestones and determine the actualresources and costs required to complete these. Consider whether contingencies are needed (and confirm they are eligible expenses).
2. Determine the eligible expense categories and maximum amount allowed bythe sponsor. Adjust scope of the project to make sure proposed activities fit within the allowance.
3. Categorize these costs (e.g., salaries, supplies, equipment…) per year, insome cases by quarter.
4. Ensure that project scope and budget match. Include indirect costs ofresearch as permitted by sponsor and the University policy.

TWO MODELS OF BUDGET DEVELOPMENT

The examples below developed by the University of British Columbia demonstratetwo ways to include indirect costs in your budget.

- **Price model:** Indirect cost is built into each budget line item.
- **Cost model:** Indirect cost of research is presented as a separate line item.

Unless the sponsor specifies in writing that they require the indirect costs of research to be presented as a separate line item (Cost Model), the indirect costshould be built into each budget line item (Price Model). Indirect costs are normally included in the price of goods and services worldwide.

For example, you are developing a budget for a funding opportunity with an indirect cost rate of 25%. Your direct costs are $201,000 broken down by expensecategories shown in the second column of the table below. The third and fourth

colums present the two ways you can include the 25% overhead in your budgetusing the Price Model or the Cost Model

OTHER FACTORS AFFECTING YOUR BUDGET

In-kind and cash contributions, like other costs to the sponsored project, must beeligible and must be treated in a consistent and uniform manner in proposal preparation and in financial reporting.

CASH CONTRIBUTIONS

Cash contributions are actual cash transactions that can be documented in theaccounting system. Examples of cash contributions include:

- allocation of compensated faculty and staff time to projects, or
- the purchasing of equipment by the university or other eligible sponsor forthe benefit of the project.

IN-KIND CONTRIBUTIONS

In-kind contributions are both non-monetary or cash equivalent resources that can be given a cash value, such as goods and/or services in support of a research project or proposal. It is challenging to report on in-kind contribution, please makesure the numbers you use are well supported, consistent and easy to quantitate.

Examples of an in-kind contribution may include:

- Access to unique database or information

- Professional, analytical, and other donated services
- Employee salaries including benefits for time allocated to the project
- Study materials, technologies, or components
- Patents and licenses for use
- Travel
- Use of facilities (e.g., lab or meeting spaces)
- Partner organization time spent participating in the project
- Eligible infrastructure items

The budget justification is a categorical description of the proposed costs. Generally, it explains staffing and supply/service consumption patterns, the methods used to estimate/calculate (including escalation or inflation factors) andother details such as lists of items that make up the total costs for a category

Using integrated technology

Given the complexities of a grant proposal budget, traditional enterprise resource planning (ERP) platforms cannot support the level of precision and collaboration necessary for adequate budgeting. But by integrating advanced technology, such as Cayuse's Fund Manager and Sponsored Projects, you canupgrade these legacy systems for optimized budgeting and proposal development.
Fund Manager allows for granular accounting and is accessible to all stakeholders through a customizable dashboard. Not only does it manage internal expenses and faculty salary splits, but it also supplements what-if planning and average monthly projections on a grant-by-grant basis.
With Sponsored Projects, administrators can track proposals, awards, contracts, and agreements via a user-specific dashboard. Its modern, user-friendly interface provides instant access, full visibility, and a collaboration workspace, relieving administrative burdens.

Boost your budgeting accuracy

Regardless of how appealing or innovative your research project might be, without justifiable numbers to back it up, it may be tossed aside. Don't take this risk. Instead, leverage integrated technology, such as Cayuse's Fund Manager and Sponsored Projects, to develop more accurate proposal budgetsand maximize your chances to win funding.

Correlation

Correlation definition:

Correlation is the analysis of relationship Between two or more variable.Twovariable are said to be correlated if rate of Change in one variable corresponding to change In other variable.

Or

In biostatistics, correlation measures the strength and direction of the relationship between two continuous variables. It's an essential tool for understanding how variables interact and can help in identifying potentialassociations or causal relationships.

Or

Correlation measures how two variables move in relation to each other. For example, if one variable increases when the other increases, they are positivelycorrelated. If one increases when the other decreases, they are negatively correlated.

Types of correlation:

- On the basis of degree of correlation, they are
    a) Positive correlation
    b) Negative correlation
- On the basis of no.of variables, they are
    a) Simple correlation
    b) Partial correlation
- Multiple correlation On the basis of linearity, they are
    a) Linear correlation
    b) Non linear correlation

1.Positive Correlation:-

If corresponding to increase in value of one Variable, there is increase in valueof other variable

Is said to be positive correlation between two Variables.Ex. Age & Weight.

2.Negative Correlation:-

If corresponding to increase in value of one Variable, there is decrease invalue of other variable

Is said to be negative correlation between two Variables.Ex. Price & Demand.

3.Simple correlation:-

When only two variables are involved the correlation Is known as simplecorrelation.

Ex. Height & weight , demand & supply.

4. Partial correlation:-

It is you to measure the degree of association between two random variables.Ex. Demand , supply & income where income is constant.

5.Multiple correlation:-

When more than two variables are involved the Correlation is known asmultiple correlation.

Ex. Rainfall, production of rice & price of rice

6. Linear Correlation:-

If ratio of change between two variables is constant, is Said to be linearcorrelation.

Ex. X 2 4 6 8

Y 3 6 9 12

7.Non-Linear Correlation:-

If ratio of change between two variables is not constant, Is said to be Non-linear correlation.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Ex. X 2 4 6 9 : Y 3 2 8 11

Methods of studying correlation:-

Two types of methods

1. Graphic method : it's of one type
   (a) Scattered diagram

2. Mathematical method : it's of three types
   (a) Karl Pearson coefficient
   (b) Rank correlation coefficient
   (C) Kendall's Tau
   (D) Point- Biserial coefficient

Methods of Studying Correlation:

1. Graphical Method:

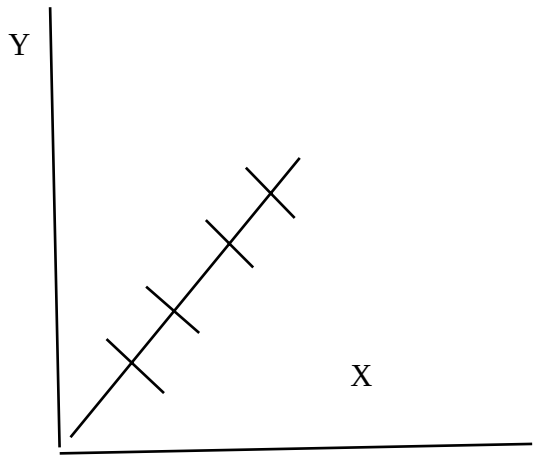
Scatter Plot: A graphical representation where each point represents a pair ofvalues. The pattern of the points indicates the type of correlation.

Or

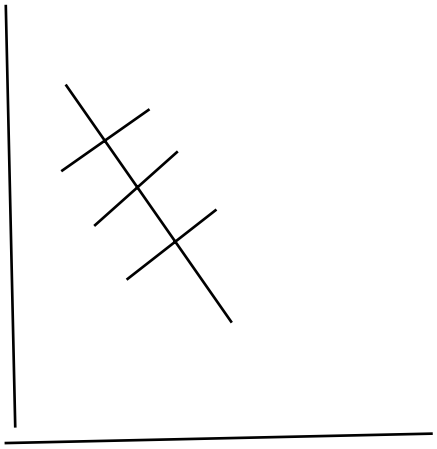
It is a graphical method to determine degree or Direction of the correlation. The pair values (x,y) of two variables X & Y are plotted On graph paper , thediagram is obtained by plotting This points is called scatter diagram.

1) Perfect Positive correlation:-

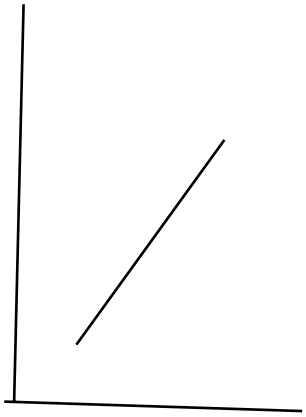
If all points of scatter Diagram lie on straight line With positive slope , thenthe Correlation is perfect positive correlation.



2) Perfect Negative Correlation:- If all points of scatter Diagram lie on straightline With negative slope , then the Correlation is Perfect Negative correlation.



3) Positive Correlation:- If all points of scatter Diagram lie above or below onStraight line with positive slope, Then the correlation is Positive correlation.



1) Negative Correlation:- If all points of scatter Diagram lie above or below onStraight line with negative slope, Then the correlation is Negative Correlation

2) No/Zero Correlation:- If all points of scattered Diagram does not follow anyPattern along straight line then It is No or Zero Correlation

2.Mathematical Methods:

- Pearson's Correlation Coefficient: Measures the strength and direction of the linear relationship between two variables. Values range from -1 to +1.

- Spearman's Rank Correlation Coefficient (ρ): Measures the strength and direction of the monotonic relationship between two ranked variables.

- Kendall's Tau: Measures the strength of the relationship between two variables using the ranks of the data.

- Point-Biserial Correlation: Used when one variable is continuous and the other is dichotomous (binary).

Importance of Correlation in Biostatistics:

- Understanding Relationships: Correlation helps in identifying and quantifying the relationship between biological variables, such as the relationship between a risk factor and a disease outcome.

- Prediction: High correlation between variables can be used to predict the value of one variable based on the other. For example, predicting blood pressure based on age.

- Hypothesis Testing: Correlation is used to test hypotheses about relationships between variables, helping in making inferences about population parameters.

- Data Reduction: By understanding correlations, researchers can reduce the dimensionality of data, retaining only variables that provide unique information.
- Study Design: Knowledge of correlations informs the design of experiments and studies, such as determining sample sizes and choosing appropriate statistical tests.

- Public Health Policy: Correlation studies can guide public health interventions by highlighting significant associations between health behaviors and outcomes.

- It is used direction of relationship.

- Measure the degree of relationship.

- Understanding economical behaviour.

- Research work.

- It is used in reducing the range of uncertainty in matter of prediction.

- In the field of science this methods are used for making conclusion.

- In the field of nature and business.

- It is used to identify relationships, inform predictive modeling, select variables, study epidemiological links, explore clinical trial outcomes, reveal genetic associations, monitor health trends, ensure quality control, and guide public health policy.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Problem Identification and Hypothesis Development

In the realm of scientific inquiry and problem-solving, the process of identifying a problem and formulating hypotheses is fundamental. This essay explores these concepts, delving into their significance, processes involved, and their interdependence in driving research and innovation.

Identifying the Problem:

Identifying a problem is akin to recognizing a gap in knowledge or a discrepancy between current understanding and observed phenomena. This initial step is crucial as it sets the stage for focused investigation and directs scientific inquiry towards meaningful goals. Problems can arisefrom various sources:

1. **Observations and Anomalies:** Scientists often notice patterns or inconsistencies duringexperiments or observations that defy existing explanations.
2. **Literature Review:** Reviewing existing research can reveal gaps in knowledge or unresolved questions that prompt further investigation.
3. **Practical Issues:** Real-world challenges, such as environmental concerns, technologicallimitations, or societal needs, can drive problem identification.

For example, a biologist might observe declining populations of a species in a specific habitat, prompting investigation into potential causes and solutions. Similarly, a computer scientist might identify inefficiencies in a particular algorithm, leading to efforts to develop more efficient alternatives.

Formulating Hypotheses:

Once a problem is identified, hypotheses are formulated as tentative explanations or predictions that can be tested through experimentation or observation. Hypotheses serve as the foundation ofscientific research, guiding the collection and interpretation of data. They are:

1. **Testable:** Hypotheses must be precise and specific, allowing for empirical testing. This often involves defining variables and outlining the expected outcomes.
2. **Falsifiable:** A good hypothesis must be capable of being proven false through experimentation or observation. This ensures rigor and reliability in scientific inquiry.
3. **Grounded in Theory:** Hypotheses should be based on existing knowledge or theories in the field. They represent educated guesses that aim to extend or challenge current understanding.

Continuing the earlier example, a hypothesis in the context of declining species populations might propose that habitat loss due to human development is the primary cause. This hypothesis could then be tested through field studies, statistical analysis, or controlled experiments.

Interdependence and Iterative Process:

Problem identification and hypothesis formulation are deeply interconnected and often iterative processes in scientific research:

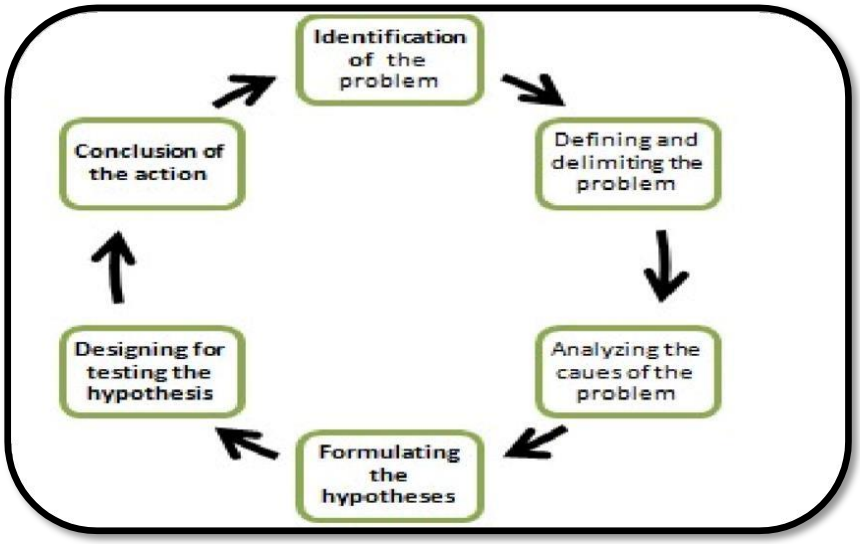
1. **Feedback Loop:** As hypotheses are tested, new data and insights emerge, which may refine or redefine the original problem statement.
2. **Iterative Refinement:** Researchers continually refine their hypotheses based on experimental outcomes and evolving understanding, leading to deeper insights and new avenues of inquiry.
3. **Creative Exploration:** Sometimes, hypotheses can lead to unexpected discoveries or paradigm shifts in scientific thinking, challenging established norms and fostering innovation.

For instance, the study of quantum mechanics emerged from attempts to resolve anomalies in classical physics, leading to a fundamentally new understanding of the universe.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Conclusion:

In conclusion, problem identification and hypothesis development are foundational to scientific inquiry, driving the advancement of knowledge and innovation across disciplines. By carefully defining problems and formulating testable hypotheses, researchers not only address existing gaps in understanding but also pave the way for transformative discoveries. This iterative processof inquiry ensures that scientific knowledge evolves, offering solutions to practical challenges and expanding the boundaries of human understanding. Thus, problem identification and hypothesis development stand as essential pillars of the scientific method, guiding researchers towards deeper insights and meaningful contributions to society.



MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

SAMPLING METHODS:

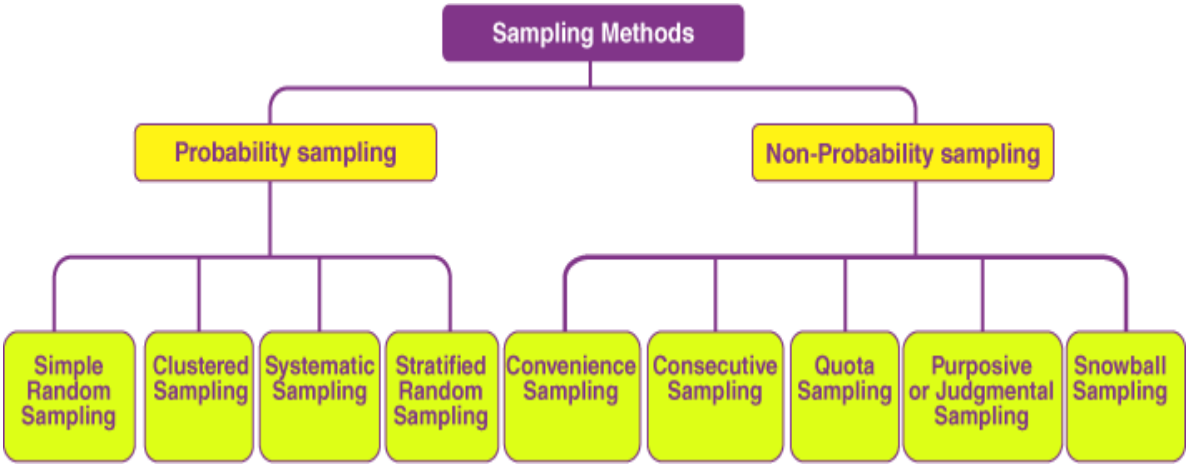
- In Statistics, the sampling method or sampling technique is the process of studying the population by gathering information and analyzing that data. It is the basis of the data where the sample space is enormous.

Types Of Sampling Methods:

- In Statistics, there are different sampling techniques available to get relevant results from the population. The two different types of sampling methods are:

  1. Random Sampling

  2. Non Random Sampling



Random Sampling or probability sampling:

Definition:

- Random sampling is a method of choosing a sample of observations from a population to make assumptions about the population. It is also called probability sampling**.** The counterpart of this sampling is Non-probability sampling or Non-random sampling. The primary types of this sampling are simple random sampling, stratified sampling, cluster sampling, and multistage sampling. In the sampling methods**,** samples which are not arbitrary are typically called convenience samples.

Types Of Random Sampling:

- The random sampling method uses some manner of a random choice. In this method, all the suitable individuals have the possibility of choosing the sample from the whole sample space. It is a time consuming and expensive method. The advantage of using probability sampling is that it ensures the sample that should represent the population. There are four major types of this sampling method, they are;

1. Simple Random Sampling
2. Systematic Sampling
3. Stratified Sampling
4. Clustered Sampling

Simple Random Sampling:

- In this sampling method, each item in the population has an equal and likely possibility of getting selected in the sample (for example, each member in a group is marked with a specific number). Since the selection of item completely depends on the possibility, therefore this method is called "Method of chance Selection**"**. Also, the sample size is large, and the item is selected randomly. Thus it is known as "Representative Sampling".

Systematic Random Sampling:

- In this method, the items are chosen from the destination population by choosing the random selecting point and picking the other methods after a fixed sample period. It is equal to the ratio of the total population size and the required population size.

Stratified Random Sampling:

- In this sampling method, a population is divided into subgroups to obtain a simple random sample from each group and complete the sampling process (for example, number of girls in a class of 50 strength). These small groups are called strata. The small group is created based on a few features in the population. After dividing the population into smaller groups, the researcher randomly selects the sample.

Clustered Sampling:

- Cluster sampling is similar to stratified sampling, besides the population is divided into a large number of subgroups (for example, hundreds of thousands of strata or subgroups). After that, some of these subgroups are chosen at random and simple random samples are then gathered within these subgroups. These subgroups are known as clusters. It is basically utilised to lessen the cost of data compilation.

Random Sampling Formula:

- If P is the probability, n is the sample size, and N is the population. Then;
- The chance of getting a sample selected only once is given by;

$$\mathbf{P = 1 – (N\text{-}1/N).(N\text{-}2/N\text{-}1)…..(N\text{-}n/N\text{-}(n\text{-}1))}$$

Cancelling = 1-(N-n/n)

$$\mathbf{P = n/N}$$

- The chance of getting a sample selected more than once is given by;

$$\mathbf{P = 1\text{-}(1\text{-}(1/N))^{n}}$$

Advantages of Random Sampling:

- It helps to reduce the bias involved in the sample, compared to other methods of sampling and it is considered as a fair method of sampling.
- This method does not require any technical knowledge, as it is a fundamental method of collecting the data.
- The data collected through this method is well informed.
- As the population size is large in the simple random sampling method, researchers can create the sample size that they want.
- It is easy to pick the smaller sample size from the existing larger population.

Example:

Suppose a firm has 1000 employees in which 100 of them have to be selected for onsite work. All their names will be put in a basket to pull 100 names out of those. Now, each employee has an equal chance of getting selected, so we can also easily calculate the probability (*P*) of a given employee being selected since we know the sample size (*n*) and the population size(*N*).

Therefore, the chance of selection of an employee only once is;

$P = n/N = 100/1000 = 10\%$

And the chance of selection of an employee more than once is;

$P = 1-(1-(1/N))^n$

$P = 1 – (999/1000)^{100}$

$P = 0.952$

$P \approx 9.5\%$

Non- Random Sampling or Non-probability Sampling:

- The Non - Random Sampling method is a technique in which the researcher selects the sample based on subjective judgment rather than the random selection. In this method, not all the members of the population have a chance to participate in the study.

Types Of Non Random Sampling:

- Non – Random Sampling methods are further classified into different types, such as convenience sampling, consecutive sampling, quota sampling, judgmental sampling, snowball sampling.

Convenience Sampling:

- In a convenience sampling method, the samples are selected from the population directly because they are conveniently available for the researcher. The samples are easy to select, and the researcher did not choose the sample that outlines the entire population.

Example:

- In researching customer support services in a particular region, we ask your few customers to complete a survey on the products after the purchase. This is a convenient way to collect data. Still, as we only surveyed customers taking the same product. At the same time, the sample is not representative of all the customers in that area.

Consecutive Sampling:

- Consecutive sampling is similar to convenience sampling with a slight variation. The researcher picks a single person or a group of people for sampling. Then the researcher researches for a period of time to analyze the result and move to another group if needed.

Quota Sampling:

- In the quota sampling method, the researcher forms a sample that involves the individuals to represent the population based on specific traits or qualities. The researcher chooses the sample subsets that bring the useful collection of data that generalizes the entire population.

Purposive or Judgement Sampling:

- In purposive sampling, the samples are selected only based on the researcher's knowledge. As their knowledge is instrumental in creating the samples, there are the chances of obtaining highly accurate answers with a minimum marginal error. It is also known as judgmental sampling or authoritative sampling.

Snowball Sampling:

- Snowball sampling is also known as a chain-referral sampling technique. In this method, the samples have traits that are difficult to find. So, each identified member of a population is asked to find the other sampling units. Those sampling units also belong to the same targeted population.

Advantages Of Non – Random Sampling:

- **Cost Savings**: Non-random sampling is cheaper because researchers can select participants more easily, avoiding the expenses of reaching a random sample.
- **Time Efficiency**: It saves time since researchers can quickly find and include participants without following strict random selection rules.
- **Convenience**: Researchers can easily pick participants who are accessible, which is helpful when the target group is small or hard to reach randomly.
- **Targeted Selection**: It allows researchers to focus on specific groups or traits of interest, making it ideal for studying unique or specialized populations

- **Flexibility**: Researchers have more freedom to choose sample sizes and types that fit their study goals, adapting as needed during the research process.
- **Ethical Considerations**: Non-random sampling can sometimes be more ethical, respecting participants' rights and preferences more effectively than random methods.

While these benefits make non-random sampling attractive, it's essential to be aware of potential biases that could affect the study's reliability and validity. Balancing these advantages with careful planning and analysis is crucial for ensuring credible research outcomes.

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

PROBABILITY

Definition:

Probability means possibility. It is a branch of mathematics that deals with the occurrence of a random event. The value is expressed from zero to one. Probability has been introduced in Maths to predict how likely events are to happen. The meaning of probability is basically the extent to which something is likely to happen. This is the basic probability theory, which is also used in the probability distribution, where you will learn the possibility of outcomes for a random experiment. To find the probability of a single event to occur, first, we should know the total number of possible outcomes.

Formula:

The probability equation defines the likelihood of the happening of an event. It is the ratio of favorable outcomes to the total favorable outcomes. The probability formula can be expressed as,

P(A) = Number of favorable outcomes to A / Total number of possible outcomes

i.e., P(A) = n(A)/n(S)

where,

- P(A) is the probability of an event 'B'.
- n(A) is the number of favorable outcomes of an event 'B'.
- n(S) is the total number of events occurring in a sample space.

Calculating Probability:

In an experiment, the probability of an event is the possibility of that event occurring. The probability of any event is a value between (and including) "0" and "1". Follow the steps below for calculating probability of an event A:

- **Step 1:** Find the sample space of the experiment and count the elements. Denote it by n(S).
- **Step 2:** Find the number of favorable outcomes and denote it by n(A).
- **Step 3:** To find probability, divide n(A) by n(S). i.e., P(A) = n(A)/n(S).

Example 1: Find the probability of getting a number less than 5 when a dice is rolled by using the probability formula.

Solution:

To find:

Probability of getting a number less than 5

Given: Sample space, S = {1,2,3,4,5,6}

Therefore, n(S) = 6

Let A be the event of getting a number less than 5. Then A = {1,2,3,4}

So, n(A) = 4

Using the probability equation,

P(A) = (n(A))/(n(s))

p(A) = 4/6

m = 2/3

Answer: The probability of getting a number less than 5 is 2/3.

Example 2: What is the probability of getting a sum of 9 when two dice are thrown?

Solution:

There is a total of 36 possibilities when we throw two dice.

To get the desired outcome i.e., 9, we can have the following favorable outcomes.

(4,5),(5,4),(6,3)(3,6). There are 4 favorable outcomes.

Probability of an event P(E) = (Number of favorable outcomes) ÷ (Total outcomes in a sample space)

Probability of getting number 9 = 4 ÷ 36 = 1/9

Answer: Therefore the probability of getting a sum of 9 is 1/9.

Example 3: Find the probability of 'getting 3 on rolling a die'.

Solution:

Sample Space = S = {1, 2, 3, 4, 5, 6}

Total number of outcomes = n(S) = 6

Let A be the event of getting 3.

Number of favourable outcomes = n(A) = 1

i.e. A = {3}

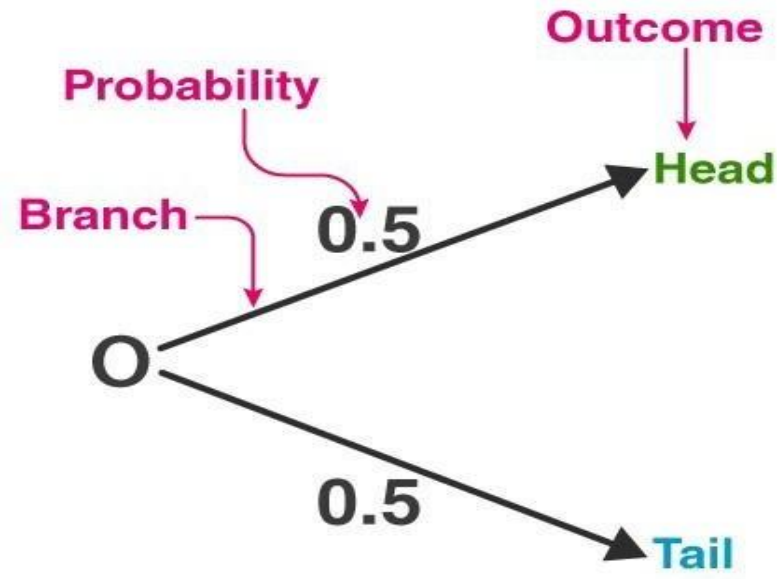
Probability, P(A) = n(A)/n(S) = 1/6

Hence, P(getting 3 on rolling a die) = 1/6

Probability Tree Diagram:

A tree diagram in probability is a visual representation that helps in finding the possible outcomes or the probability of any event occurring or not occurring. The tree diagram for the toss of a coin given below helps in understanding the possible outcomes when a coin is tossed. Each branch of the tree is associated with the respective probability (just like how 0.5 is written on each brack in the figure below). Remember that the sum of probabilities of all branches that start from the same point is always 1 (here, 0.5 + 0.5 = 1).

Tree Diagram for the Toss of a coin:



Types of Probability:

There can be different perspectives or types of probabilities based on the nature of the outcome or the approach followed while finding probability of an event happening. The four types of probabilities are,

- Classical Probability
- Empirical Probability
- Subjective Probability
- Axiomatic Probability

Classical Probability:

Classical probability, often referred to as the "priori" or "theoretical probability", states that in an experiment where there are B equally likely outcomes, and event X has exactly A of these outcomes, then the probability of X is A/B, or P(X) = A/B. For example, when a fair die is rolled, there are six possible outcomes that are equally likely. That means, there is a 1/6 probability of rolling each number on the die.

Empirical Probability:

The empirical probability or the experimental perspective evaluates probability through thought experiments. For example, if a weighted die is rolled, such that we don't know which side has the weight, then we can get an idea for the probability of each outcome by rolling the die number of times and calculating the proportion of times the die gives that outcome and thus find the probability of that outcome.

Subjective Probability:

Subjective probability considers an individual's own belief of an event occurring. For example, the probability of a particular team winning a football match on a fan's opinion is more dependent upon their own belief and feeling and not on a formal mathematical calculation.

Axiomatic Probability:

In axiomatic probability, a set of rules or axioms by Kolmogorov are applied to all the types. The chances of occurrence or non-occurrence of any event can be quantified by the applications of these axioms, given as,

- The smallest possible probability is zero, and the largest is one.
- An event that is certain has a probability equal to one.
- Any two mutually exclusive events cannot occur simultaneously, while the union of events says only one of them can occur.

Probability of an event:

Assume an event E can occur in r ways out of a sum of n probable or possible equally likely ways**.** Then the probability of happening of the event or its success is expressed as;

P(E) = r/n

The probability that the event will not occur or known as its failure is expressed as:

P(E') = (n-r)/n = 1-(r/n)

E' represents that the event will not occur.

Therefore, now we can say;

P(E) + P(E') = 1

This means that the total of all the probabilities in any random test or experiment is equal to 1.

Bayes' Theorem on Conditional Probability:

Bayes' theorem describes the probability of an event based on the condition of occurrence of other events. It is also called conditional probability. It helps in calculating the probability of happening of one event based on the condition of happening of another event.

For example, let us assume that there are three bags with each bag containing some blue, green, and yellow balls. What is the probability of picking a yellow ball from the third bag? Since there are blue and green colored balls also, we can arrive at the probability based on these conditions also. Such a probability is called conditional probability.

The formula for Bayes' theorem is

$$P(A|B)=P(B|A)\cdot P(A)/P(B)$$

where, P(A|B) denotes how often event A happens on a condition that B happens.
where, P(B|A) denotes how often event B happens on a condition that A happens.
P(A) the likelihood of occurrence of event A.
P(B)the likelihood of occurrence of event B.

Law of Total Probability:

If there are n number of events in an experiment, then the sum of the probabilities of those n events is always equal to 1.

$$P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) = 1$$

MBY-404: BIOSTATISTICS AND RESEARCH METHODOLOGY

Applications of probability:

Probability has a wide variety of applications in real life. Some of the common applications which we see in our everyday life while checking the results of the following events:

- Choosing a card from the deck of cards
- Flipping a coin
- Throwing a dice in the air
- Pulling a red ball out of a bucket of red and white balls
- Winning a lucky draw

Other Major Applications of Probability:

- It is used for risk assessment and modelling in various industries
- Weather forecasting or prediction of weather changes
- Probability of a team winning in a sport based on players and strength of team
- In the share market, chances of getting the hike of share prices