# Masters of Public Health

# Biostatistics I

## Louvenser Minthor
### (August 2022)

Introduction

Robert R. Sokal and F. James Rohlf define Statistic as the scientific study of numerical data based on natural phenomena, and Biostatistics is the application of the statistical method and reasoning to the biological, medical and public health field. Hence, biostatistics must meet the commonly established criteria of validity of scientific evidence. In this course we will have to discuss on general biostatistics and some of it principles.

For video Part 01, Overview of Biostatistics… answer these questions.

1. What is the Scientific method for Biostatistics?
Ans: the scientific method is the process that we use to answer questions in order to support or reject hypothesis that explain The biological, medical and public health problem

2. What is the role of Biostatistics in science?
Ans: the role of biostatistics is to rise questions in order to generate hypothesis, secondly to design study to observe, generate information and find evidence to answer these questions

3. What is Descriptive Statistics?
Ans: it is a set of technic that help us describe, organize, summarize the data. also help create Hypothesis and use the proper inferential method

4. Explain Variation with Statistical Reasoning.
Ans: We do not have 100% of certainty about a phenomenon , there will always be some variation , this is why we use probability to determine the likelihood of an event to occurring by quantifying the variability and the uncertainty

5. What are some types (sources) of Variation?
Ans: There are multiple sources of variations :
- Natural variation ( across individual naturally )
- Measurement  variation, error within the tools
- Bias which is the difference between the average value of the measurement and the real value
- Variance : variation among measurement we see about their average

6. What is the Biostatistics paradigm?
Ans: biostatistics attempt to simply explain the the relationship between natural phenomena, the relationship between variables, their association , their connection

7. What are the types of variables? Define each one.
Ans: there are several types of variables:
- random variable: values obtain as result of chance
- Response variable: Alison called dependent variable are variable that affected or can be caused by other variables
- Explanatory variables: independent variable are those which provoke the response

8. What are measurement scales?
Ans: measurement scales refer to the way that the variables and defined and categorized. Each of them has their own properties and their own use. There are:
- Nominal
- Ordinal
- Interval
- Ratio

9. What are Exploratory Data Methods?
Ans: the explanatory data methods helps researchers to identify and analyze configurations of relations between two or more variables, most of them with a high accuracy, as there is a possibility of testing statistic significance by calculating the confidence level associated with validation of relation concerned across the entire population and not only the surveyed sample.

For video Part 02, Probability & Probability Distributions… answer these questions.

1. What is the difference between a population and a sample?
Ans: A population is the entire body of an observation while a sample is a subset of the population

2. When do you know that your data refers to a sample? When do you know that your data refers to a population?
Ans: we will know that the data refers to a population as it will refer to the entire body of the observation and beneath the them of parameter as oppose, the sample is a set of the population and its descriptive measure will be call statistics that is a estimate of a population parameter.

3. What is statistical inference?
Ans: statistical inference is the process of using data to make conclusion or conjecture of what is likely to be true about a population or a sample

4. Why is probability the basis of statistical inference?
Ans: because to make a statement or an inference about something that we don't know yet, we need the probability tools measure the observation , to generate the data in order to make that statement

5. How do you use statistical inference?
Ans: we use statistical inference to surmise what is true or likely to be true about the observation

6. Create some data using a method similar to coin tossing. One idea is to write various colors on pieces of paper and take them out of a hat. Remember to put the color taken out back into the hat before taking out the next color.
Ans:
Every piece of paper has one color in each side:
H0: blue / black
H1: red / red
H2: red /  blue

The probability of picking a red color:
-   Pr ( red given H0 : blue / black ) = 0 ➔ Inference: not possible
-   Pr ( red given H1 : red / red ) = 1 ➔ Inference: practically 100% possible
-   Pr ( red given H2 : red / blue ) = ½ ➔ Inference: possible

7. Define probability.
Ans: probability is the measure of the uncertainty associated with occurrence of an event, the likelihood of occurrence of the event

8. Define probability distribution. What is the purpose of a probability distribution?
Ans: the probability distribution gives the description of all possible results and the probability of each of them to occur And they are been used with the purpose of describing variable of interest


9. What is a z value? How do you interpret a z value?
Ans: Z value or standard normal distribution is the number of standard deviations from the mean value of the reference population. As interpretation: when values are within Z+- 1.96 there are more likely to occur, about 95% chance to appear

10. When would you use a probability distribution?
Ans: we use a probability distribution to calculate confidence intervals for parameters and to calculate critical regions for hypothesis tests.

11. How do you calculate a probability?
Ans: we can calculate the probability by dividing the favorable number of outcomes by the total number of possible outcomes

12. Calculate the probability for a normal distribution.
Ans: $z = \frac{X-m}{D}$

For video Part 03, Sampling distribution & Estimation… answer these questions.

1. Define a Sampling Distribution. Why is it considered Theoretical?
Ans: - A sampling distribution is the theoretical Distribution of all possible values which can be assumed by some sample statistic computed from samples of the
same size randomly drawn from the same population
    - it is theoretical because we don't built this distribution on data


2. Define a distribution of sampling means. Why are sampling means different?
Ans: - Is the kind of distribution where we can take many samples, all of the same size, and calculate the mean of each of those, we could put those means together to form a distribution

3. What are the most frequently used sampling distributions? Define each one.
Ans: - the sample mean : dimpling from a population of continuous values
The average taking from the simple that we are collected

- the sample proportion : if we are sampling from a population of binary outcomes

- the difference between two sample means of samples from two different populations
- the difference between two sample proportion samples from two different populations

4. Define Estimation for Biostatistics.
Ans: - is the process of guessing the unknown value, it's what derives from the sample that we've chosen

5. What is a Point Estimation?
Ans: - is the single value that we might get from our sample

6. What is Interval Estimation? How would uncertainty and variability affect the interval estimation?
Ans: - It's a point estimate surrounded by an interval that expresses the uncertainty or variability associated with the estimate
    - It's influenced by the sample size…

7. What is a Confidence Interval Estimation?
Ans: - An interval of values computed from sample data that is likely to cover the true parameter of interest.

8. Why is a 95% confidence interval commonly used?
Ans - because we can be confident it has 95% chance to contain the true population parameter

9. What is the difference between Z and t distributions? When is each one used? How does

the sample size affect whether you use a Z or t distribution?

Ans: - The difference between t distribution and Z distribution is that the t distribution is a family of distribution and it is determined by the sample size minus 1 ( n-1 ),

- The t distribution is used when the population standard deviation is unknown  and when the sample is small ( less than 30 )

- The simple size affect whether we use a Z or t distribution as the t distribution become fairly like a Z distribution as the simple size increases

10. Define Degrees of Freedom? Why is it used?

Ans: - The degrees of freedom is a part of inferential statistical analyses which estimate or make inferences about population parameters based on sample data that are free to vary.

- The degrees of freedom are important for finding critical cutoff value for inferential statistical tests

11. How does sample size affect the Estimation? (refer to discussion after 46:15 minute point in video.)

Ans: To have more precision and a fairly narrow confidence interval the simple size would become bigger

12. How do you calculate the needed sample size (n) in order to estimate a desired range for a true population mean to within a 95% confidence interval? (Look at figure at 47:30 minute point of video.) What are the 3 steps in the calculation?

Ans: - we calculate the needed sample size by: sample mean + or – Z time sigma over the square root of n

- The 3 steps are:
    1. Specify the confidence interval
    2. Make assumption about the standard deviation
    3. Solve for n

Conclusion

Biostatistics is the base of is the application of statistical methods to the solution of biological problems. Thus in this course we review some basics of the biostatistic methods. Started by an overview of biostatistics, afterwards probability and distribution and finally sampling distribution and estimation.

Références

Robert r. Sokal and F. James Rohlf, ( 2009 ), introduction to Biostatistics 2$^{nd}$ ed, Mineola, New York , Dover Publication.inc.
Getu Degu, Fasil Tessema, (2005), Biostatistics, Ethiopia Public Health Training Initiative, Tha Carter Center, The Ethiopia Ministry of Health, The Ethiopia Ministry of Education.
Lloyd D. Fisher, Gerald Van Belle, (2004), Biostatistics: a methodology for the health sciences, New Jersey, John Wiley & sons, inc.
Degrees of Freedom, 2015 The University of Texas at Austin,
https://sites.utexas.edu/sos/degreesfreedom/

Video links:
Part 1… https://www.youtube.com/watch?v=H1hPz9LTjno
Part 2… https://www.youtube.com/watch?v=723D-MCiMyc
Part 3… https://www.youtube.com/watch?v=kjyQwPEWQlQ