

# Statistics course notes



**ASSIGN  
BUSTER**

• Defining What Statistics Really Is 1. 1 Nature of Statistics The term “ Statistics” came from the Latin word ‘ status’ which could be translated as ‘ state’. The usage of this term only became popular during the 18 th century where they defined Statistics as “ the science of dealing with data about the condition of a state or community”. The practice of statistics could be traced back even from the early biblical times where they gather figures related to governance of the state for they realized the importance of these figures in governing the people.

Even until today, worldwide, governments have intensified their data gathering and even widen the scope of their numerical figures due to the rise of more cost-efficient methods for collecting data. Some of the most popular figures that are being released by almost all countries are Gross National Product (GNP), Birth rates, Mortality Rates, Unemployment Rate, Literacy Rates and Foreign Currency Exchange Rates. Also, the use of Statistics is not limited to government use only. Right now, almost all business sectors and fields of study use statistics.

Statistics serves as the guiding principle in their decision making and helps them come up with sound actions as supported by the analysis done in their available information. Indicated below are some of the uses of Statistics in various fields: Medicine: Medical Researchers use statistics in testing the feasibility or even the efficacy of newly developed drugs. Statistics is also used to understand the spread of the disease and study their prevention, diagnosis, prognosis and treatment (Epidemiology).

JDEUSTAQUIO 1 Economics: Statistics aids Economists analyze international and local markets by estimating some Key Performance Indicators (KPI) such

as unemployment rate, GNP/GDP, amount of exports and imports. It is also used to forecast economic fluctuations and trends. Market Research: derives statistics by conducting surveys and coming up decisions from these statistics through feasibility studies or for testing the marketability of a new product.

Manufacturing: use statistics to assure the quality of their products through the use of sampling and testing some of their outputs Accounting/Auditing: uses sampling techniques in statistics to examine and check their financial books. Education: Educators use statistical methods to determine the validity and reliability of their testing procedures and evaluating the performance of teachers and students. 1. 2 Basic Concepts We normally hear the word “statistics” when people are talking about basketball or the vital statistics of beauty contestants.

In this context the word “statistics” is used in the plural form which simply means a numerical figure. But the field of Statistics is not only limited to these simple figures and archiving them. In the context of this course, the definition of “Statistics” is mainly about the study of the theory and applications of the scientific methods dealing all about the data and making sound decisions on this. Statistics is the branch of science that deals with the collection, presentation, organization, analysis and interpretation of data. Sometimes, gathering the entire collection of elements is very tedious, expensive or even timeconsuming.

Because of this data gatherers sometimes resort to collecting just a portion of the entire collection of elements. The term coined for the entire collection of elements is called Population while the subset of the population is referred

as the Sample. JDEUSTAQUIO 2 Population is the collection of all elements under consideration in a statistical inquiry while the sample is a subset of a population. THINK: Could you say that the entire population is also a sample? The specification of the population of interest depends upon the scope of the study.

Let's say that if we wish to know the average expenditure of all households in Metro Manila, then the population of interest is the collection of all households in Metro Manila. If there is a need to delimit the scope of the study due to some constraints, we could redefine the population of interest. We could delimit the scope of the study to only specific city in Metro Manila. With this the study would only include the collection of all households in \_\_\_\_\_ City. The elements of the population is not only limited to individuals, it can be objects, animals, geographical areas, in other words, almost anything.

Some examples of possible populations are: the set of laborers in a certain manufacturing plant, the set of foreigners residing on Boracay for a certain day, set of Ford Fiesta produced in the entire Philippines on a month. In any studies involving the use of Statistics, there would be at least one attribute of the element in the population which we would be studying. This attribute or characteristic is what we call variable. Just like in the field of Mathematics, we normally denote a variable with a single capital letter i. e. A, X, Z.

The variable is a characteristic or attribute of the elements in a collection that can assume different values for the different elements. While an observation is a realized value of the variable, and the collection of these observations is called the data. Example: The Department of Health is

interested in determining the percentage of children below 12 years old infected by the Hepatitis B virus in Metro Manila in 2006. Population: Set of all children below 12 years old in Metro Manila in 2006 Variable of Interest: whether or not the child has ever been infected by the Hepatitis B virus.

Possible Observations: Infected, Never Infected Regardless of whether every element of the data on the population or sample is used, it is often still difficult to convey meaning to these observations is not summarized. This is the JDEUSTAQUIO 3 reason why it is important to condense these observations to a single figure to completely describe the entire data. This condensed value is what we call summary measure. The parameter is a summary measure describing a specific characteristic of a population while a statistic is a summary measure describing a specific characteristic of the sample. . 3 Fields of Statistics There are two major fields in Statistics. The first one is (i) Applied Statistics, this deals mainly with the procedures and techniques used in the collection, presentation, organization, analysis and interpretation of data. On the other hand, the second one is (ii) Mathematical Statistics, which is concerned with the development of the mathematical foundations of the methods used in Applied Statistics. In this course, we would mostly deal with the basics of Applied Statistics. This field could also be sub-divided into two areas of interest.

These two are Descriptive and Inferential Statistics. Both are definitive of their names. Descriptive Statistics includes all the techniques used in organizing, summarizing, and presenting the data on hand, while Inferential Statistics includes all the techniques used in analyzing the sample data that will lead to generalizations about a population from which the sample came

from. To clarify, we may use descriptive statistics for population data or sample data. If we are dealing with population data, then the results of the study are applicable only to the defined population.

In the same manner, if we use descriptive statistics to sample data, then the conclusions are applicable only to the selected sample. JDEUSTAQUIO 4 1. 4 Statistical Inquiry Statistical Inquiry is a designed research that provides information needed to solve a research problem. Oftentimes, researchers can now find an appropriate statistical technique that will help them answer their research problems. This is because of the wide array of applications of the various statistical techniques used in a statistical inquiry. Below is the diagram depicting the entire process of statistical inquiry.

Step 1: • Identify the Problem • Plan the Study • Collect the Data • Explore the Data • Analyze Data and Interpret the Results • Present the Results Step 2 Step 3 Step 4 Step 5 Step 6 JDEUSTAQUIO 5 • Theory without data is just an Opinion 2. 1 Measurement The data used for statistical analysis should always be accurate, complete, and up-to-date because the information that we would get is only as good as the data that we have. Good quality data comes at a cost but if we have the assurance of obtaining essential information that answers our research problem then it is all worth it.

Measurement is the process of determining the value or label of the variable based on what has been observed. Naturally, our interpretation of the values in our data will depend on the measurement system or the rule that we used to assign the values to the different categories of the variable. In particular, it will depend on the relationship among the values used in the system. The general classification used to describe the types of relationship among these

values or categories is what is known as “ levels of measurement”. The four levels of measurement are nominal, ordinal, interval and ratio level.

It is necessary to know the level of measurement used to measure a variable because this will help in the interpretation of the values of the variables and choosing the suitable statistical technique to use in the analysis. Ratio level of measurement has all of the following properties : a) the numbers in the system are used to classify a person/object into distinct, nonoverlapping, and exhaustive categories; b) the system arranges the categories according to magnitude; c) the system has a fixed unit of measurement representing a standard size throughout the scale; and d) the system has an absolute zero.

JDEUSTAQUIO 6 Some examples of variables with ratio level of measurement are: 1. Distance traveled by a car (in km) 2. Height of a flag pole (in metres) 3. Weight of a whole dressed chicken (in kilograms) Now we will discuss each of the properties that is required for a measuring scale to have in order for it to be considered as having a ratio level of measurement: a) The numbers in the system are used to classify a person/object into distinct nonoverlapping, and exhaustive categories. This first condition requires that we use categories that would place the observations logically into one and only one category.

This means that two objects assigned the same value must belong in the same category and be placed in a different category if the characteristics of interest is really different. b) The system arranges the categories according to magnitude. This second property requires that the measurement system must arrange the categories according to either ascending or descending order. c) The system has a fixed unit of measurement representing a

standard size throughout the scale. The third property requires the scale to use a unit of measure that depicts a fixed and determinate quantity.

This means that a one-unit difference must have the same interpretation wherever it appears in the scale. d) The system has an absolute zero. The fourth property requires the measurement system to have an absolute zero or the true zero point. This means that the scale considers the value, “ 0” (zero) as the complete absence of the characteristic itself. One example of this is any monetary measurement where zero means that there is absolutely no money. Interval Level of Measurement satisfies only the first three conditions of the ratio level of measurement.

The only difference of the interval level of measurement to the ratio level of measurement is the absence of the absolute zero value. This means that the interval level of measurement considers “ 0” (zero) as a value like any other numbers and not as the absence of the characteristic of interest. The most common example of this is measuring temperature in Celsius or Fahrenheit where the value “ zero” does not mean that there is no temperature. Ordinal Level of Measurement satisfies only the first two conditions of the ratio level of measurement.

The ordinal level of measurement only uses a scale that ranks or orders the observed values in either ascending or descending order. The interval or simply the difference of the scale from one point to another does not need to be equal all throughout the scale. For example the ranking of the student in class according to their grades could be tagged as 1st, 2nd, 3rd, 4th and so on. The difference of the grade between the 1st student and the 2nd placed



student does not need to be of the same gap between the 4<sup>th</sup> placer and the 5<sup>th</sup> placer.

Nominal Level of Measurement satisfies only the first property of the ratio level of measurement. The nominal level of measurement is the weakest level of measurement among the four. This is because its only aim is to classify the values into separate categories without regards to the ordering of these categories in ascending or descending manner. Most often, this level of measurement uses non-quantifiable categories like the different religions, zip code or the student number.

## 2.2 Collecting Data

### 2.2.1 Data Collection Methods

The most commonly used methods for collecting data are: i. Use of Documented Data, ii. ) Surveys, iii. ) Experiments, and iv. ) Observation.

**Use of Documented Data** It is not necessary to use original data in conducting studies; sometimes it would make things easier if the researcher uses the data that is already available if there is such one suitable for the study. The only dilemma with using documented data is its reliability and veracity. Therefore, the researcher must look closely on the source of this data to have a measure on the reliability of the data that would be used.

Also, these documented data can be categorized in to two, the primary data and the secondary data. Primary Data are data documented by the primary source, meaning, the data collectors themselves documented the data. Secondary Data are data documented by a secondary source, meaning, an individual/agency, other than the data collectors, documented the data.

**Surveys** Another common method of collecting data is the survey. The people who answer the questions in a survey are called the respondents.

This method is much more expensive than collecting data using documented stuff.

Another problem of using surveys is that reliability of the data depends mainly on the survey process itself, either from the respondent, the survey design, questionnaire or if it is a personal interview there might be a problem with the interviewer if he/she lacks training. The Survey is a method of collecting data on the variable/s of interest by asking people questions. When data came from asking all the people in the population, then it is called census. On the other hand, when the data came from asking a sample of people selected from a well-defined population, then it is called a sample survey.

**Experiments** If the researcher is interested in something that involves cause-and-effect relationship, conducting the experiment is most likely the suitable way of collecting data. The most common experiment that is normally conducted during the primary level is the mongo seed experiment. The aim of this experiment is to see the relationship of the growth of the mongo in relation with sunlight exposure, amount of water and the type of soil. The Experiment is a method of collecting data where there is direct human intervention on the conditions that may affect the values of the variable of interest.

**Observation Method** The Observation Method is a method of collecting data on the phenomenon of interest by recording the observations made about the phenomenon as it actually happens. JDEUSTAQUIO 9 The observation method is useful in studying the reactions and behavior of individuals or groups of persons/objects in a given situation or environment as it happens,

For example, a researcher may use the observation method to study the behavior patterns of an indigenous tribe which is difficult to be gathered using the other methods.

## 2. 2. 2 The Questionnaire

The questionnaire is an instrument for measuring which is used in various data collection methods (commonly used in surveys). The questionnaire may either be self-administered or interview-based which are both explanatory of their names.

### 2. 2. 2. 1 Type of Questions ?

A Closed-ended question is a type of question that includes a list of response categories from which the respondent will select his/her answer. ? An Open-ended question is a type of question that does not include response categories.

#### Comparison of Open-Ended and Closed-Ended Questions

##### Open-Ended ?

Respondent can freely answer ? Can Elicit feeling and emotions of the respondent ? Can reveal new ideas and views that the researcher might not have considered ? Good for complex issues ? Good for questions whose possible responses are unknown ? Allow respondents to clarify answers ? Get detailed answers ? Shows how respondent think ? ? ? ? Closed-Ended

Facilitates tabulation of responses Easy to code and analyze Saves time and money High response rate since it is simple and quick to answer ? Response categories make questions easy to understand ?

Can repeat the study and easily make comparisons JDEUSTAQUIO

Advantages 10 Disadvantages ? Difficult to tabulate and code ? High refusal rate because it requires more ? ? ? ? ? Increases respondent to burden when time and effort on the respondent Respondents need to be articulate Responses can be inappropriate or vague May threaten respondent Responses have different levels of detail there are too many or too limited

response categories ? Bias responses against categories excluded in the choices ? Difficult to detect if the respondent misinterpreted the question 2.

2. . 2 Response Categories for Close-ended Questions 1. Two-way Question – provides only two alternative answers from which the respondent can choose Example: Have you ever traveled outside the country by any means of transportation? Yes No 2. Multiple-choice Question – provides more than two alternatives from which the respondent can only choose one. Example: What is your marital status? Never Married Divorced/Separated Married Widowed 3. Checklist Question – provides more than two alternatives from which the respondent can choose as many responses that apply to him/her.

Example: What kind/s of novel do you like to read? Comedy Romance Fantasy Sci-Fi Horror Non-fiction Mystery Others, please specify \_\_\_\_\_

JDEUSTAQUIO 11 4. Ranking Question – provides categories that respondents have to either arrange from highest to lowest or vice versa depending upon a particular criterion. Example: Below is a list of considerations in choosing and buying a new laptop. Put number (1) beside the quality that you prioritize the most, (2) for the second priority and so on. Prize Brand Quality Durability Style Novelty Warranty [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] . Rating Scale Question – provides a graded scale showing all possible directions and intensity of attitude of a respondent on a particular question or statement. Example: How satisfied are you on the teaching method of your instructor in this course? 1 Very Dissatisfied 2 Dissatisfied 3 Neutral 4 Satisfied 5 Very Satisfied 6. Matrix Question – a type of question which places various questions together to save space in the questionnaire. It is like having any of

the five earlier types of questions and squeezing more than one question in a form of a table.

Example: For each statement, please indicate with a checkmark whether you agree or disagree with it

Statements

Statistics is a very difficult subject

Only few people could understand Statistics

I would rather sleep than study Statistics at home

Agree Disagree

JDEUSTAQUIO 12 2. 2. 2. 3

Pitfalls to Avoid in Wording Questions

1. Avoid Vague Questions – State all question clearly. All respondents must have the same interpretation to a question. If not, their answers will not be comparable, making it difficult to analyze their responses. Example: How often do you watch a movie in a movie theatre?

Very Often Often Not too often Never

Problem: The word “ often” is vague. Instead, you may ask how many times did he/she watched a movie last month.

2. Avoid Biased Question – A biased question influences the respondents to choose a particular response over the other possible responses. Whether the bias is caused accidentally or intentionally, the data would become useless because it still failed to reveal the truth. Example: There are many different types of sport like badminton, basketball, billiards, bowling and tennis. Which type of sport d you enjoy watching?

Problem: The sports mentioned in the first sentence will be in the top of the minds of the respondents. It is likely for the respondents to choose from among these sports. This will result in a bias against the sports not mentioned in the list.

3. Avoid Confidential and Sensitive Questions – These questions usually offend the pride or jeopardize the prestige of the respondent. Example: Do you bring home office supplies? If yes, how often

do you bring home office supplies? Problem: The question may sound offensive to the pride of the respondent. 4.

Avoid Questions that are difficult to answer – Do not ask questions that are too difficult for the respondent to answer truthfully. Such questions would only encourage respondents to guess their answers, if not totally refuse to answer the question. Example: If you are the president of the nation, what are you going to do to attain economic recovery? JDEUSTAQUIO 13 5. Avoid Questions that are confusing or perplexing to answer – Sometimes a poorly written question can confuse the respondent on how to answer the question Example: Did you eat out and watch a movie last weekend?

Problem: This is a double-barreled question, where you combine two or more question in to a single question. You should opt to separate this question into two to avoid confusion. 6. Keep the Questions short and simple – Long and complicated question can be difficult to understand. The respondent may lose interest in the question because of its length or might have problem comprehending very long statement needed to understand the question. 2. 3 Sampling and Sampling Techniques 2. 3. 1 Basic Concepts As we have discussed on the previous Chapter 1, sample is the subset of a population.

Some people think that if we are basing our analysis on samples, why don't we just guess our analysis entirely without any data? This question could be partially answered by a quote from Sir Charles Babbage, the Father of the Computer who said that, " Errors using inadequate data are much less than those using no data at all". So now, before we can talk about the different sampling selection procedures, we need to familiarize ourselves first with some terms. The target population is the population we want to study The

sampled population is the population from where we actually select the sample

It is good if the target and the sampled population have the same collection of elements. The problem is that often times in life, expectations do not jive well with reality. One example where the target and the sampled population would be different from each other is the case where the target population is the collection of all the residents of Metro Manila. If we would be using a telephone directory to select our sample, this collection would be very different from the target population since this would exclude all the residents that have no landline.

JDEUSTAQUIO 14 The sampling frame or frame is a list or map showing all the sampling units in the population. In any statistical inquiry, whether the data will come from a census or from a sample, it is important that we are conscious of all the possible errors that we introduce (hopefully not intentionally) in the results of the study. In order for us to do this and reduce these errors, we need to understand the possible sources of errors, namely, the sampling errors and the nonsampling errors.

Sampling error is the error attributed to the variation present among the computed values of the statistic from the different possible samples consisting of  $n$  elements. Nonsampling errors is the error from other sources apart from sampling fluctuations Note that the ONLY TIME that the sampling error would not be present is if we have conducted a census. However, census results are NOT ERROR-FREE. Census and samples can both have nonsampling errors (simply the errors not brought solely by sampling). Total

Error Nonsampling Error Error in the implementation of the sampling design  
Measurement Error Sampling Error

Selection Error Instrument Error Frame Error Population Specification Error  
Response Error Processing Error Interviewer Bias Surrogate Information Error  
Diagram of the Various Sources of Error JDEUSTAQUIO 15 2. 3. 2 Methods of  
Probability Sampling Probability Sampling is a method of selecting a sample  
wherein each element in the population has a known, nonzero chance of  
being included in the sample; otherwise, it is a nonprobability sampling  
method. ? A nonzero chance of inclusion means that the sampling procedure  
must give all the elements of the sample population an opportunity of being  
a part of the sample.

All of the elements that belong in the sampled population must be included  
in the selection process. ? Another requirement of probability sampling is  
that we should be able to determine the chance that an element will be  
included in the selected sample. Take note that the probability of each  
element in the sampled population need not be equal to each other. 2. 3. 2.  
1 Simple Random Sampling Simple Random Sampling (SRS) is a probability  
sampling method wherein all possible subsets consisting of  $n$  elements  
selected from the  $N$  elements of the population have the same chances of  
selection.

In simple random sampling without replacement (SRSWOR), all the  $n$   
elements in the sample must be distinct from each other. In simple random  
sampling with replacement (SRSWR), the  $n$  elements in the sample need not  
be distinct, that is, an element can be selected more than once as a part of  
the sample. The most apparent example of SRSWOR that we could see every



day on mass media is the National lottery where the numbers that would be drawn must be distinct and every number should have an equal chance of being selected in the draw. JDEUSTAQUIO 16 Visual representation of Simple Random Sampling without Replacement. 2. 3. 2. 2 Stratified Sampling

Stratified sampling is a probability sampling method where we divide the population into nonoverlapping subpopulations or strata, and then select one sample from each stratum. The sample consists of all the samples in the different strata. Stratified sampling, in general, simply requires the division of the population into nonoverlapping strata, wherein each element of the population needs to belong to exactly one stratum. Then each sample would be selected from the strata using any probability sampling method. If simple random sampling used for each sample in the strata then this sampling is called stratified random sampling.

JDEUSTAQUIO 17 Visually, it might look something like the image below. With our population, we can easily separate the individuals by color. Once we have the strata determined, we need to decide how many individuals to select from each stratum. The most common practice is that the number selected should be proportional. In our case,  $\frac{1}{4}$  of the individuals in the population are blue, so  $\frac{1}{4}$  of the sample should be blue as well. Working things out, we can see that a stratified (by color) random sample of 4 should have 1 blue, 1 green and 2 red. JDEUSTAQUIO 18 2. 3. 2. 3 Systematic Sampling

Systematic sampling is a probability sampling method wherein the selection of the first element is at random and the selection of the other elements in the sample is systematic by taking every  $k$ th element from the random start,

where  $k$  is the sampling interval To select a sample using systematic sampling, we need to perform the following steps: 1. Decide on a method of assigning a unique serial number, from 1 to  $N$ , to each one of the elements in the population. 2. Choose  $n$  = sample size so that it is a divisor of  $N$  = population size. Compute for the sampling interval  $k = N/n$ . 3.

Select a number from 1 to  $k$ , using a randomization mechanism. Denote the selected number by  $r$ . The element in the population assigned to this number is the first element of the sample. 4. The other elements of the sample are those assigned to the numbers  $r + k$ ,  $r + 2k$ ,  $r + 3k$ , and so on, until you get a sample size of  $n$ . 5. In case that  $k = N/n$  is not a whole number; the first element would still be  $r$  but would be a randomly chosen number from 1 to  $N$  instead  $k$  as used on the previous step. By visual explanation, so to use systematic sampling, we need to first order our individuals, then select every  $k$ th.

In our example, we want to use 3 for  $k$ ? Can you see why? Think what would happen if we used 2 or 4. JDEUSTAQUIO 19 For our starting point, we pick a random number between 1 and  $k$ . For our visual, let's suppose that we pick 2. The individuals sampled would then be 2, 5, 8, and 11. 2. 3. 2. 4 Cluster Sampling Cluster sampling is a probability sampling method wherein we divide the population into nonoverlapping groups or clusters consisting of one or more elements, and then select a sample of clusters. The sample will consist of all the elements in the selected clusters.

To select a sample using cluster sampling, we need to perform the following steps: 1. Divide the population into nonoverlapping clusters. 2. Number the clusters in the population from 1 to  $N$ . 3. Select  $n$  distinct numbers from 1 to

N using a randomization mechanism. The selected clusters are the clusters associated with the selected numbers 4. The sample will consist of all the elements in the selected clusters. Cluster sampling is often confused with stratified sampling, because they both involve "groups". In reality, they're very different. In stratified sampling, we split the population up into groups (strata) based on some characteristic.

In essence, we use cluster sampling when our population is already broken up into groups (clusters), and each cluster represents the population. That way, we just select a certain number of clusters. JDEUSTAQUIO 20 With our visual, let's suppose the 12 individuals are paired up just as they were sitting in the original population. Since we want a random sample of size four, we just select two of the clusters. We would number the clusters 1-6 and use technology to randomly select two random numbers. It might look something like this: JDEUSTAQUIO 21 2. 3. 2. 5 Multistage Sampling

Multistage sampling is a probability sampling method where there is a hierarchical configuration of sampling units and we select a sample of these units in stages. Unlike all the other previously presented sample selection procedures where the process of sampling takes place in a single phase, we accomplish the selection of the elements in the sample under multistage sampling after several stages of sampling. We first partition the population into non-overlapping primary stage units (PSUs) and select a sample of PSUs. We then subdivide the selected PSUs into non-overlapping second-stage units (SSUs) and select a sample of SSUs.

We continue the process until we identify the elements in the sample at the last stage of sampling. For example, consider a light-bulb example using

two-stage sampling procedure. Let's suppose that the bulbs come off the assembly line in boxes that each contains 20 packages of four bulbs each. One strategy would be to do the sample in two stages: Stage 1: A quality control engineer removes every 200th box coming off the line. (The plant produces 5, 000 boxes daily. (This is systematic sampling. ) Stage 2: From each box, the engineer then samples three packages to inspect. (This is an example of cluster sampling. 2. 3. 3 Methods of Nonprobability Sampling All sampling methods that do not satisfy the requirements of probability sampling are considered as nonprobability sampling selection procedures. These methods do not make use of randomization mechanism in identifying the sampling units included in the sample. It allows the researcher to choose the units in the sample subjectively. And since the sample selection is subjective, there is really no way to assess the reliability of the results without so much assumptions (remember assumptions are very prone to mistakes). JDEUSTAQUIO 22

Despite this drawback of nonprobability sampling, these methods are still more commonly used since it is less costly and easier to administer. Here are some of the most basic nonprobability sampling selection procedures: 2. 3. 3. 1 Haphazard or Convenience Sampling In haphazard or convenience sampling, the sample consists of elements that are most accessible or easier to contact. This usually includes friends, acquaintances, volunteers, and subject who are available and willing to participate at the time of the study. The most common example that we could see on the television is the text polls about a certain issue.

This type of sampling the opinion of the people doesn't involve randomization mechanism in the selection of the units in the sample. This is sometimes referred to as the nonprobability counterpart of simple random sampling.

### 2.3.3.2 Judgement or Purposive Sampling

The elements are carefully selected to provide a "representative" sample. Studies have demonstrated that selection bias can arise even with expert choice but nevertheless the method may be appropriate for very small samples when the expert has a good deal of information about the population-elements. The two common features of the method are: a. sampling units often consist of relatively large groups; and, b. ) sampling units are chosen so that they will provide accurate estimates for important control variables for which results are known for the whole population and its hoped that it will give "good" estimates for other variables that are highly correlated with the control variables. This sampling method may be considered as the nonprobability counterpart of Cluster sampling.

### 2.3.3.3 Quota Sampling

This is considered as the nonprobability counterpart of stratified sampling. In this method, interviewers are assigned quotas of respondents of different types to interview.

The quotas are sometimes chosen to be in proportion to the estimated population figures for various types, often based on past census data. The researcher also chooses the groups or strata in the study but the selection of the sampling units within the stratum does not make use of a probability sampling method.

### 2.4 Presentation of Data

After data collection, we organize and analyze the data, and then we present the

results of our analysis in some form that will allow us to reveal and highlight the important information that we were able to extract.

Unless we do this, we will only get lost in huge mound of numbers and labels that we have collected. Our grade school teachers already taught us this various kinds of presenting the data so why do we need to study this again? We may be familiar with the line chart and the bar chart but we need to learn or review the basic principles of constructing a good table and a good graph. With good data presentation, we can discover, and even explore possible relationships. Poor data presentation will only mislead, deceive, and misinform.

It is therefore essential that we remember to put a more conscious effort to use these different methods of presentation properly in order to maximize data description and analysis.

#### 2. 4. 1 Textual Presentation

Textual Presentation of data incorporates important figures in a paragraph of text. In textual presentation, it aims to direct the readers' attention to some data that need particular emphasis as well as to some important comparisons and to supplement with a narrative account from a table or a chart. It could also show the summary measures like minimum, maximum, totals and percentages.

We do not need to put all figures in a textual presentation; we just have to select the most important ones that we want to focus on. Example: The Philippine Stock Exchange composite index lost 7. 19 points to 2, 099. 12 after trading between 2, 095. 30 and 2, 108. 47. Volume was 1. 29 billion shares worth 903. 15 million pesos (16. 7milliondollars). The broader all share index gained 5. 21 points to 1, 221. 34. (From: Free mandated March

17, 2005) When the data become voluminous, the textual presentation is strongly not advised because the presentation becomes almost incomprehensible.

JDEUSTAQUIO 24 2. 4. 2 Tabular Presentation Tabular Presentation of data arranges figures in a systematic manner in rows and columns. Tabular presentation is the most common method of data presentation. It can be used for various purposes such as description, comparison, and even showing relationships between two or more variables of interest. We will discuss three types of presenting in tabular form, namely; Leader Work, Text Tabulation and Formal Statistical table which is categorized according to their format and layout. Leader Work

Leader work has the simplest layout among the three types of tables. It contains no table title or column headings and has no table borders. This table needs an introductory or descriptive statement so that the reader can understand the given figures. The Population in the Philippines for the Census Years 1975 to 2000 is as follows

Year	Population
1975	42,070,660
1980	48,098,460
1990	60,703,206
1995	68,616,536
2000	76,498,735

National Statistics Office The 1990 and 1995 figures include the household population, homeless population, and Filipinos in Philippines embassies and mission abroad.

In addition, the census comprise institutional population found living quarters such as penal institutions, orphanages, hospitals, military camps, etc. As you can see, the above table would not be clear without the introductory statement. Likewise, both have no table numbers that we can use to refer to these figures. Thus, we use the leader work when there are only one or two

columns of figures that we can incorporate as part of the textual presentation for a more organized presentation. Text Tabulation The format of text tabulation is a little bit more complex than leader work.

It already has column headings and table borders so that it is easier to understand than leader work. However it still does not have table title and table number. Thus, it also requires an introductory statement so that the readers can comprehend the given figures. Similar to leader work, we can place additional explanatory statement in the footnote. JDEUSTAQUIO 25 The Population in the Philippines for the Census Years 1975 to 2000 is as follows:

Year	1975	1980	1990	1995	2000
a. No. of Filipinos (in thousands)	42,070,664	48,098,466	60,703,216	68,616,546	76,498,404

4 National Statistics Office

The 1990 and 1995 figures include the household population, homeless population, and Filipinos in Philippines embassies and mission abroad. In addition, the census comprise institutional population found living quarters such as penal institutions, orphanages, hospitals, military camps, etc. Formal Statistical Table The formal statistical table is the most complete type of table since it has all the different and essential parts of a table like table number, table title, head note, box head, stub head, column headings, and so on.

It could be a stand-alone table since it does not need any accompanying texts and it could be easily understood on its own. Heading consists of the table number, title and head note. It is located on top of the table of figures.

- i. Table number is the number that identifies the position of the table in a sequence.
- ii. Table title states in telegraphic form of the subject, data classification, and place and period covered by the figures in the table.
- iii.



Head note appears below the title but above the top cross rule of the table and provides additional information about the table.

Box head consists of pner heads and columns heads. i. Spanner head is a caption or label describing two or more column heads. ii. Column head is a label that describes the figures in a column. iii. Panel is a set of column heads under the same pner head. Stub consists of row captions, center head, and stub head. It is located at the left side of the table. i. Row caption is a label that describes the figures in a row. ii. Center head is a label describing a set of row captions. iii. Stub head is a caption or label that describes all of the center heads and row captions.

It is located on the first row. iv. Block is a set of row captions under the same center head. JDEUSTAQUIO 26 Table number Stub head Title Head note

Table 10. 9 Employed Persons by Major Industry Group January 2008 - October 2010 (in thousands) Panel Heading Spanner head Column head Industry Group Oct Total Agriculture Center head Agriculture, Hunting and Forestry Fishing 36, 488 12, 265 10, 769 1, 496 5, 375 197 3, 058 163 1, 957 18, 550 2010 Jul Apr 36, 237 12, 244 10, 760 1, 484 5, 409 194 3, 003 141 2, 071 18, 585 35, 413 11, 512 10, 073 1, 439 5, 487 212 3, 063 137 2, 075 18, 414 Jan 6, 001 11, 806 10, 351 1, 455 5, 322 193 3, 009 157 1, 963 18, 872 Oct 35, 478 12, 072 10, 563 1, 509 5, 154 169 2, 937 160 1, 888 18, 250 2009 Jul Apr 35, 508 11, 940 10, 476 1, 464 5, 273 177 2, 947 145 2, 004 18, 294 34, 997 12, 313 10, 841 1, 472 5, 088 166 2, 841 130 1, 951 17, 595 Jan 34, 262 11, 846 10, 446 1, 400 4, 856 152 2, 849 134 1, 721 17, 560 Oct 34, 533 12, 320 10, 860 1, 460 5, 078 176 2, 897 123 1, 882 17, 135 2008 Jul Apr 34, 593 12, 103 10, 695 1, 408 5, 130 154 2, 960 146 1, 870 17, 360 33,

535 11, 904 10, 450 1, 454 5, 000 151 2, 883 123 1, 843 16, 630 Jan 33, 693  
 11, 792 10, 409 1, 383 4, 981 152 2, 963 126 1, 740 16, 919

Industry Mining and Quarrying Manufacturing Electricity, Gas and Water  
 Construction Services Wholesale & Retail Trade, Repair of Motor Vehicles,  
 Motorcycles & Personal & Household Goods Hotels and Restaurants  
 Transport, Storage and Communication Financial Intermediation Real Estate,  
 Renting and Business Activities Public Administration & Defense, Compulsory  
 Social Security Education Health and Social Work Other Community, Social &  
 Personal Service Activities Private Households with Employed Persons Extra-  
 Territorial Organizations & Bodies 7, 158 7, 030 6, 885 7, 064 6, 901 , 725 6,  
 681 6, 635 6, 528 6, 599 6, 322 6, 333 1, 119 2, 711 412 1, 239 1, 037 2,  
 704 420 1, 166 991 2, 741 383 1, 061 1, 104 2, 735 384 1, 119 1, 012 2, 735  
 375 1, 100 1, 064 2, 694 376 1, 090 976 2, 628 389 1, 023 988 2, 660 337 1,  
 044 941 2, 587 373 985 984 2, 525 369 969 924 2, 575 366 953 964 2, 674  
 364 904 BLOCK 1, 771 1, 165 465 855 1, 954 1 1, 835 1, 238 457 866 1, 831  
 1 1, 959 1, 156 447 984 1, 804 3 1, 823 1, 146 432 949 2, 114 2 1, 771 1,  
 168 412 868 1, 908 0 1, 772 1, 157 428 876 2, 110 2 1, 794 1, 068 408 907  
 1, 718 3 1, 659 1, 157 435 857 1, 785 3 1, 690 1, 096 406 796 1, 733 \* 1,  
 741 1, 076 386 847 1, 863 1 , 661 1, 028 384 843 1, 572 2 1, 612 1, 083 390  
 846 1, 747 2

Notes: 1. Data were taken from the results of the quarterly rounds of the Labor Force Survey (LFS) using past week as reference period.  
 2. Details may not add up to totals due to rounding. 3. The definition of unemployment was revised starting the April 2005 round of the LFS. As such, LFPRs, employment rates and unemployment rates are not comparable with those of previous survey rounds. Also starting with January 2007, estimates

were based on 2000 Census-based projections. 4. Data are as of January 2012. / - preliminary source note Source: National Statistics Office (NSO).  
 footnote JDEUSTAQUIO 27 2. 4. 3 Graphical Presentation Tabular  
 Presentation of data portrays numerical figures or relationships among variables in pictorial form. The graph or statistical chart is a very powerful tool in presenting data. It is an important medium of communication because we can create a pictorial representation of the numerical figures found in tables without showing too many figures. We construct graphs not only for presentation purposes but also as an initial step in analysis.

The graph, as a tool for analysis, can exhibit possible associations among the variables and can facilitate the comparison of different groups. It can also reveal trends over time. The different types of statistical charts are line chart, vertical bar chart, horizontal bar chart, pictograph, pie chart, and statistical map. It is important to know when and how to use these different charts. The selection of the correct type of chart depends upon the specific objective, the characteristic of the users, the kind of data, and the type of device and material on hand. Line Chart The line chart is useful for presenting historical data. This chart is effective in showing the movement of a series over time. As shown in the figures below, the movement can be increasing, decreasing, stationary, or could be fluctuating. Title at Top Scale figures for y-axis 20 No. of Accidents involving Company B during their Years of Service No. of Accidents Scale label for y-axis 15 10 Grid lines 5 0 1 2 3 4 5 6 7 8 9 10 Footnote Source Note Years of Service Scale label for x-axis

Scale figures for x-axis JDEUSTAQUIO 28 NEVER use line charts/graphs that are too stretched either horizontally or vertically, for it may mislead the

person looking at the graph and interpret it as something that it is not really representing. JDEUSTAQUIO 29 Types of Line Chart Simple Line Chart – This has only one curve and is appropriate for one series of time data. Multiple Line Chart – This type of line chart shows two or more curves. We use this if we wish to compare the trends in two or more data series.

Although the use of Multiple Line Chart is now commonly used, it should be taken notice the number of series that you include in a graph, if there are a lot of series in a single chart, it might become too confusing to see. Number of Daily Responses (Example of Single Line Chart) JDEUSTAQUIO 30 Column Chart We use the column charts to compare amounts in a time series data. The emphasis in a column chart is on the differences in magnitude rather than the movement of a series. ? We can also use the column chart to graph the frequency distribution of a quantitative variable.

We call this chart a frequency histogram. ? For time series data, we arrange the columns on the horizontal axis in chronological order, starting with the earliest date. Title at Top Grid lines Scale label for y-axis Scale figures for x-axis Scale figures for y-axis The proportions of the columns must be just right. Columns must not be too wide or too narrow. The space between the bars must also be just right. Usually, the space between bars is around one-fourth of the width of the column. It is also advisable to use scale figures that are multiples of 5.

If the observed values are so small, we can use multiples of 1 or 2. JDEUSTAQUIO 31 Horizontal Bar Chart Its use is appropriate when we wish to show the distribution of categorical data. We use the horizontal bar chart so we can compare the magnitudes for the different categories of a qualitative

variable. We place the categories of the qualitative variable on the y-axis. This will be more practical than placing the categories on the x-axis because there is more space for text labels on the y-axis. Just like the column charts, the bars should not be too wide, too narrow, too long and nor too short. Arranging the bars according to length usually facilitates comparisons. It may be decreasing or ascending order. ? If there are “ Others” category, we always place this as the first or the last category. ? If the categorical variables have a natural ordering, such as a rating scale, then we should retain the order of the categories in the scale instead of arranging the bars according to length. ? We should always choose appropriate colors or patterns for the bars. We should avoid selecting wavy and weird patterns since this will only produce an optical illusion.

JDEUSTAQUIO 32 Pie Chart It is a circle divided into several sections. Each section indicates the proportion of each component or category. This is useful for data sorted in to categories for a specific period. The purpose is to show the component parts with respect to the total in terms of the percentage distribution. The components of the pie chart should be arranged according to magnitude. If there’s an ‘ Others’ category, we put it in the last section. We use different colors, shading, or patterns to distinguish one section of the pie to the other sections.

We plot the biggest slice at 12 o’clock. If we want to emphasize a particular sector of the pie chart, we may explode that slice by detaching it from the rest of the sectors. The pie chart is applicable for qualitative rather than quantitative data. However, if the variable has too many categories (more than 6), we should use the horizontal bar chart rather than the pie chart.

JDEUSTAQUIO 33 Pictograph o It is like a horizontal bar chart but instead of using bars, we use symbols or pictures to represent the magnitude. o The purpose of this chart is to get the attention of the reader. The pictograph provides an overall picture of the data without presenting the exact figures. o Usually, we can only show approximate figures in a pictograph since we have to round off figures to whole numbers. It still allows the comparison of different categories even if we just present only the approximate values. o The choice for the symbol or picture should be apt for the type of data. It should be self-explanatory, interesting, and simple. Statistical Maps ? ? ? ? ? This type of chart shows statistical data in geographical areas. This could also be called as crosshatched maps or shaded maps.

Geographic areas may be barangays, cities, districts, provinces, and countries. The figures in the map can be ratios, rates, percentages, and indices. We do not use the absolute values and frequencies in statistical maps. JDEUSTAQUIO 34 Types of Statistical Maps ? Shaded Map – map that makes use of shading patterns. The shading pattern indicates the degree of magnitude. It usually runs gradually from dark to light (Darker shading of the map usually means larger magnitude). ? Dot map – chart that gives either the location or the number of establishments in a certain geographical area.

The example below is a dot map of the number of people with Hispanic descent in the US. JDEUSTAQUIO 35 2. 5 Organization of Data The first step in data analysis is organizing the collected data. In its organized form, important features of the data become clear and apparent. The two common forms of organized data are the array and the frequency distribution 2. 5. 1 Raw Data and Array Raw Data are data in their original form. The actual data that we

collect from surveys, observation, and experimentation are what we call raw data. Raw data have not yet been organized or processed in any manner.

Example: Raw Data of the Final Grades of 100 Selected Students who took Stat 101 79 62 74 79 81 65 79 94 75 52 73 85 78 82 83 79 73 81 88 81 74 60 92 86 86 60 90 64 57 63 88 63 87 69 77 53 76 52 72 89 66 56 57 92 82 66 70 72 73 63 88 77 60 97 70 92 67 92 50 65 72 74 79 51 86 55 67 66 79 95 60 93 66 99 89 94 97 78 55 79 77 92 93 92 50 65 79 62 56 77 53 72 57 62 80 79 76 82 74 76 Array is an ordered arrangement of data according to magnitude. We also refer to the array as sorted data or ordered data Arranging the observations manually according to magnitude is very tedious especially if we are dealing with voluminous data.

Thus, it is more convenient to use computer programs to sort the data. The array is not a summarized data set. It is simply an ordered set of observations. We consider both the raw data and array as ungrouped data.

JDEUSTAQUIO 36 Example: Array of the Final Grades of 100 Selected Students who took Stat 101 50 50 51 52 52 53 53 55 55 56 56 57 57 57 60 60 60 62 62 62 63 63 63 64 65 65 65 66 66 66 66 67 67 69 70 70 72 72 72 72 73 73 73 74 74 74 74 75 76 76 76 77 77 77 77 78 78 79 79 79 79 79 79 79 79 80 81 81 81 82 82 82 83 85 86 86 86 87 88 88 88 89 89 90 92 92 92 92 92 93 93 94 94 95 97 97 99

2. 5. Frequency Distribution (FDT) The frequency distribution (FDT) is a way of summarizing data by showing the number of observations that belong in the different categories or classes. We also refer to this as grouped data. The frequency distribution is another way of organizing the data. It is a summarized form of the raw data or array wherein we do not see the actual observed values anymore. The two general

forms of frequency distribution are single-value grouping and grouping by class intervals: 1. Single-value grouping – is a frequency distribution where the classes are the distinct values of the variable.

This is applicable for data with only a few unique values. 2. Grouping by Class Intervals – is a frequency distribution where the classes are the intervals. Example: Suppose we have data on the number of children of 50 married women using any modern contraceptive method. 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5

JDEUSTAQUIO 37 Since there are only 6 unique values in the data set, then we use single-value grouping, Distribution of Married Women Using Any Modern Method of Contraceptive by Number of Children

No. of Children	0	1	2	3	4	5
Number of Married Women	7	8	11	14	8	2

Concepts related to Frequency Distribution 1. Class Interval – is the range of values that belong in the class or category. 2. Class Frequency – is the number of observations that belong in a class interval. 3. Class Limits – are the end numbers used to define the class interval. The lower class limit (LCL) is the lower end number while the upper class limit (UCL) is the upper end number. 4. Open Class Interval – is a class interval with no lower class limit or no upper class limit. 5. Class Boundaries – are the true class limits.

If the observations are rounded figures, then we identify the class boundaries based on the standard rules of rounding as follows: the lower class boundary (LCB) is halfway between the lower class limit of the class and the upper class limit of the preceding class while the upper class boundary (UCB) is halfway between the upper class limit of the class and the



lower class limit of the next class. 6. Class size – is the size of the class interval. It is the difference between the upper class boundaries of the class and the preceding class; or the difference between the lower class boundaries of the next class and the class. . Class Mark - is the midpoint of a class interval. It is the average of the lower class limit and the upper class limit or the average of the lower class boundary and upper class boundary of a class interval. JDEUSTAQUIO 38

After learning the concepts that we need to construct a frequency distribution table, we can now list down the steps in constructing a frequency distribution table.

- Determine the adequate number of classes denoted by K

Step 1: • We can use the Sturges's rule to approximate the number of classes which is given by  $K = 1 + 3.322(\log n)$

Determine the range,  $R = \text{highest observed value} - \text{smallest observed value}$

Step 2: value • Compute for the pre-class size  $C' = R/K$

Step 3: • Determine the class size, C, by rounding-off  $C'$  to a convenient number

Step 4: number • Choose the lower class limit of the first class. Make sure that the smallest observation will belong in the first class.

Step 5: • List the class intervals. Determine the lower class limits of the succeeding classes by adding the class size to the lower class limit of the previous class. The last class should include the largest observation.

Step 6: • Tally all the observed values in each class interval

Step 7: Sum the frequency column and check against the total number of observations

Step 8: After constructing the basic frequency distribution table, we could now add some other components to it that would help us in the analysis of the data.

- o o Relative Frequency – is the class frequency divided by the total number of observations

Relative Frequency Distribution Percentage (RFP) – is relative frequency multiplied by 100. JDEUSTAQUIO 39

The relative frequency and RFP show the proportion and percentage of

<https://assignbuster.com/statistics-course-notes/>

observations falling in each class. The RFP allows us to compare two or more data sets with different totals.

The sum of the RFP column is one hundred percent (100%). Another component that could be added to the FDT is the cumulative frequency distribution which is comprised of two components. o o The less than cumulative frequency distribution (CFD) shows the number of observations with values higher than or equal to the lower class boundary. Example: Using the data of the Grades of 100 Students who took Stat 101, we would construct the frequency distribution table with the extra components; RF, RFP CFD. First, we will compute for K using the Sturges' rule,  $K = 1 + (3.322 \cdot \log n) = 1 + (3.322 \cdot \log 100) = 1 + (3.322 \cdot 2) = 7.644 \approx 8$  ? Secondly, we compute for the range,  $R = \text{max. value} - \text{min. value} = 99 - 50 = 49$  Third, compute for  $C'$  and eventually  $C = R / K = 49 / 8 = 6.125 \approx 7$  Now we can create the FDT for the data set, Class Limits LCL 50 57 64 71 78 85 92 99 UCL 56 63 70 77 84 91 98 105 Class Boundaries LCB 49.5 56.5 63.5 70.5 77.5 84.5 91.5 98.5 UCB - 56.5 - 63.5 - 70.5 - 77.5 - 84.5 - 91.5 - 98.5 - 105.5 Frequency f 11 13 13 19 19 11 13 1 n= 100 Class Mark x 53 60 67 74 81 88 95 102 RF f/n 0.11 0.13 0.13 0.19 0.19 0.11 0.13 0.01 RFP % 11 13 13 19 19 11 13 1 ; CFD 11 24 37 56 75 86 99 100 CFD ; CFD 100 89 76 63 44 25 14 1

JDEUSTAQUIO 40 Graphical Presentation of the Frequency Distribution We can effectively interpret the frequency distribution when displayed pictorially since more people understand and comprehend the data in graphic form. In this section we would discuss the various method of presenting the frequency distribution in graphical form. 1. Frequency Histogram The

frequency histogram shows the overall picture of the distribution of the observed values in the dataset. It displays the class boundaries on the horizontal axis and the class frequencies on the vertical axis. The frequency histogram shows the shape of the distribution.

The area under the frequency histogram corresponds to the total number of observations. The tallest vertical bar shows the frequency of the class interval with the largest class frequency.

2. Relative Frequency/ Relative Frequency Percentage Histogram The RF or RFP histogram displays the class boundaries on the horizontal axis and the relative frequencies or RFPs of the class intervals on the vertical axis. It represents the relative frequency of each class by a vertical bar whose height is equal to the relative frequency of the class. The shape of the relative frequency histogram and frequency histogram are the same.

JDEUSTAQUIO 41 3. Frequency Polygon For the frequency polygon, plot the class frequencies at the midpoint of the classes and connect the plotted points by means of straight lines. Since it is a polygon we need to close the ends of the graph. To close the polygon, add an additional class mark on both ends of the graph wherein both ends have the frequency of 0. The advantage of the frequency polygon over the frequency histogram is that it allows the construction of two or more frequency distributions on the same plot area. This facilitates the comparison of the different frequency distributions.

The frequency polygon also exhibits the shape of the data distribution.

JDEUSTAQUIO 42 4. Ogives The ogive is the plot of the cumulative frequency distribution. This graphical representation is used when we need to

determine the number of observations below or above a particular class boundary. The less than ogive is the plot of the less than cumulative frequencies against the upper class boundaries. On the other hand, the greater than ogive is the plot of the greater than cumulative frequencies against the lower class boundaries. Connect the successive points by straight lines.

If we superimpose the less than and greater than ogives, the point of intersection gives us the value of the median. The median divides the ordered observations into two equal parts. JDEUSTAQUIO 43 • Summary Measures Part 1 3. 1 Measures of Central Tendency The average is the popular term that is used to refer to a measure of central tendency. Most are already accustomed to thinking in terms of an average as a way of representing the collection of observations by a single value. For instance, we often use the average score to represent the scores in the exam of all students in a class.

We can say that if the average score is high, then we conclude that the class performed well. The average could also be used to compare the performance of two groups based on the average of both groups and comparing which one has the higher average. The most common measure of central tendency is the arithmetic mean. The two other measures of central tendency that we will present in this section are the median and the mode. All of these measures aim to give information about the ‘center’ of the data or distribution. 3. 1 . 1 Summation Notation

The summation notation provides a compact way of writing the formulas for some of the summary measures that would be discussed in this section. The

capital Greek letter “ sigma”, $\sum$  is the mathematical symbol that represents the process of summation. The symbol,  $\sum$  is equal to  $X_1 + X_2 + X_3 + \dots + X_n$  where  $X_i$  = value of the variable for the  $i$ th observation  $i$  = index of the summation (the letter below  $\sum$  ).  $1$  = lower limit of the summation (the number below  $\sum$  ).  $n$  = upper limit of the summation (the letter above  $\sum$  ). We read  $\sum$  as “ summation of  $X$  sub  $i$ , where  $i$  is from  $1$  to  $n$ ”.

JDEUSTAQUIO 44 S