PSYCHOLOGICAL STATISTICS - EMBAT

LESSON 1 PART 1: Statistics and Behavioral Sciences

[Explain the importance of statistics in psychology.]

STATISTICS

- ➢ The term statistics refers to a set of mathematical procedures for organizing, summarizing, and interpreting information.
- organizing information: e.g. by gender, section
  - through population
- summarize: straight summarization
- interpret info: you need to interpret the information provided

---

TWO GENERAL PURPOSE:

TO DESCRIBE

- ➢ Statistics are used to organize and summarize the information so that the researcher can see what happened in the study and can communicate the results to others.
  - seeing what happened in the study
  - mean, median, mode

TO MAKE INFERENCES

- ➢ Statistics help the researcher to answer the questions that initiated the research by determining exactly what general conclusions are justified based on the specific results that were obtained.
  - to make a conclusion

DESCRIPTIVE STATISTICS

- ➢ These are statistical procedures that take raw scores and organize, summarize, or simplify them in a form that is more manageable.
  - ***to describe***

INFERENTIAL STATISTICS

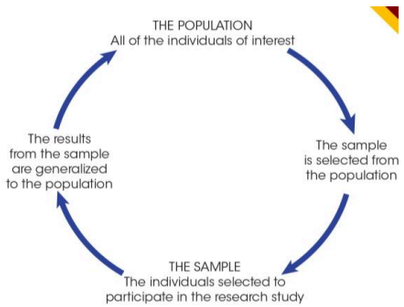
- ➢ These consist of techniques that allow us to study samples and then make generalizations about the populations from which they were selected.
  - ***to make inferences***

POPULATION

- ➢ A set of all the individuals of interest in a particular study.

SAMPLE

- ➢ A set of individuals selected from a population, usually intended to represent the population in a research study.



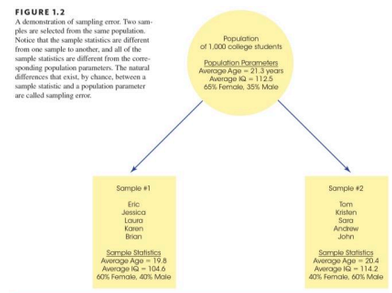
[need to set a certain qualifications or characteristics]

POPULATION PARAMETER

- ➢ a value, usually a numerical value, that describes a population. A parameter is usually derived from measurements of the individuals in the population.
  - lahat kukunin para masabi ang result

SAMPLE STATISTICS

- ➢ a value, usually a numerical value, that describes a sample. A statistic is usually derived from measurements of the individuals in the sample.
  - average



SAMPLING ERROR

- ➢ the naturally occurring discrepancy, or error, that exists between a sample statistic and the corresponding population parameter.
  - not totally a mistake
  - naturally occurs

MARGIN OF ERROR

- ➢ tells you how many percentage points your results will differ from the real population value.

- There always will be some “margin of error” when sample statistics are used to represent population parameters.
    - “gaano kalayo ang discrepancy”
    - gaano kalayo/kababa/kalapit ang error

LESSON 1 PART 2: **Observations, Measurement, and Variables**

PSYCHOLOGY IS AN EVIDENCE-BASED FIELD!

[Contemporary psychology is based on observations rather than intuition, through measurement (e.g., numerical or categorical).]

VARIABLE

- A variable is a characteristic or condition that changes or has different values for different individuals.

DATA

- are measurements or observations. A data set is a collection of measurements or observations.
- A datum (singular) is a single measurement or observation and is commonly called a score or raw score.

DISCRETE AND CONTINUOUS VARIABLES

DISCRETE

- It consists of separate, indivisible categories. No values can exist between two neighboring categories.
- Examples are: birth order, occupation, college programs, gender.
- In measuring discrete variables, we can use either *NOMINAL* or *ORDINAL* scales to determine *QUALITATIVE* differences.

CONTINUOUS

- There are an infinite number of possible values that fall between any two observed values; it is divisible into an infinite number of fractional parts.
- Examples are: Height, weight, distance, time, intelligence quotient.
- In measuring discrete variables, we can use either *INTERVAL* or *RATIO* scales to determine *QUANTITATIVE* differences.
    - decimal points & fractions

When measuring continuous variable:

- It should be very rare to obtain identical measurements for two different individuals.
- If there is identical scores or observations:

1. The variable is not really continuous.
2. The measurement procedure is very *crude* - the continuous variable is divided into widely separated discrete numbers.

- Each measurement category is actually an interval that must be defined by boundaries for us to meaningfully differentiate scores from one another.

REAL LIMITS

- Real limits are the boundaries of intervals for scores that are represented on a continuous number line.
- The real limit separating two adjacent scores is located exactly halfway between the scores.
- Each score has two real limits. The upper real limit is at the top of the interval, and the lower real limit is at the bottom.

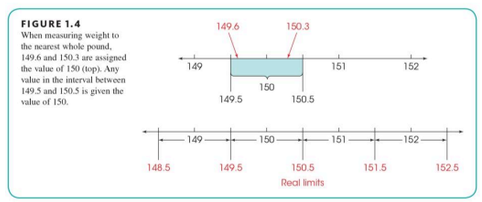


FIGURE 1.4
When measuring weight to the nearest whole pound, 149.6 and 150.3 are assigned the value of 150 (top). Any value in the interval between 149.5 and 150.5 is given the value of 150.

Two people who both claim to weigh 150 pounds are probably not exactly the same weight. One person may actually weigh 149.6 and the other 150.3, but they are both assigned a weight of 150 pounds.

    - 80 — 80.5 (upper limit), 79.5 (lower limit)

SCALES OF MEASUREMENT

- Measurement involves assigning individuals or events to categories (either qualitatively or quantitatively). The categories used to measure a variable make up a scale of measurement, and the relationships between the categories determine different types of scales.
    - thematic analysis & quantitative

- ➢ The distinctions among the scales are important because:

1. They identify the limitations of certain types of measurement.
2. Certain statistical procedures are appropriate for scores that have been measured on some scales but not on others.

NOMINAL SCALE

- ➢ It consists of a set of categories that have different names.
- ➢ Measurements on a nominal scale label and categorize observations, but do not make any quantitative distinctions between observations.
- ➢ It allows us to determine whether two individuals are different, but they do not identify either the direction or the size of the difference.

GENDER

- ➢ You can place people as either heterosexual or homosexual. However, you cannot make an assumption that one group is better than the other.

MARITAL STATUS

- ➢ People can be single, married, or divorced. We can tell the qualitative difference between these categories, but we cannot say that being married is better than being single.

COLLEGE PROGRAM

- ➢ We can categorize people based on their college program. Although we can say the qualitative difference between nursing and psychology, there is no quantitative value that differentiates them from one other.
    - [no ranking]

ORDINAL SCALE

- ➢ It consists of a set of categories that are organized in an ordered sequence.
- ➢ Measurements on an ordinal scale rank observations in terms of size or magnitude.
- ➢ There is direction of difference but it do not allow you to determine the size of the difference between two individuals.
    - organized in order
    - ranked in terms of size and magnitude
    - direction but does not determine the difference

OLYMPICS

- ➢ Athletes can won either gold, silver, and bronze medal, indicating their final placement. However it does not tell how much the gold medalist is better than the silver or bronze medalist.

EDUCATION LEVEL

- ➢ Educational attainment can be ranked such as junior HS, senior HS, bachelor, masters, and doctorate degree. We can categorize people according to this but we cannot tell how much a bachelor student is better than senior HS student.

LIKERT SCALE

- ➢ We can categorize people based on their college program. Although we can say the qualitative difference between nursing and psychology, there is no quantitative value that differentiates them from one other.

INTERVAL SCALE

- ➢ It consists of ordered categories that are all intervals of exactly the same size.
- ➢ Equal differences between numbers on a scale reflect equal differences in magnitude.
- ➢ However, the zero point on an interval scale is arbitrary and does not indicate a zero amount of the variable being measured.

INTELLIGENCE

- ➢ Values obtained in intelligence scales has real limits. Each real limit has an equal interval from one another.
- ➢ No zero value because there is no person who has “zero intelligence.”

TEMPERATURE

- ➢ It has arbitrary zero and not an absolute zero since the value “0” is only used to differentiate the categorical values between hotness and coldness. There is always temperature, either cold or hot.

pH LEVEL

- The difference between a pH of 4 and 5 is the same as the difference between 7 and 8. However, the pH scale does not have a true zero point; a pH of 0 does not represent the absence of acidity but rather an extreme level of acidity.

RATIO SCALE

- It is an interval scale with the additional feature of an absolute zero point.
- A zero point is not arbitrary but rather is a meaningful value representing none (a complete absence) of the variable being measured.
- It possible to compare measurements in terms of ratios (e.g., 10 gallons is twice as much as 5 gallons).

BRAIN ACTIVITY

- "0 brain activity" means there is no brain activity and the person is probably dead.
- This is the same with other medical-related measurement (e.g., 0 heartbeat means no heartbeat).

REACTION TIME

- "0 milliseconds reaction time" means the person did not react at all.
- There is also equal intervals between milliseconds and there is an absolute zero measurement.

VELOCITY

- A speed of "0 kph" means the object is not object. Between each kilometer, there is always 1000 meters.
- It is also possible to make ratio judgements: 10 kph is twice as speed as 5 kph.

SCALES OF MEASUREMENT

NOMINAL

- Categorizes and labels variables.

ORDINAL

- Nominal + ranked the variables.

INTERVAL

- Nominal + ordinal + has equal intervals

RATIO

- Nominal + ordinal + interval + absolute zero

LESSON 1 PART 3: **Statistical Notation**

[Identify and define observations, measurements, and variables.]

X

- data for a particular variable (an observed score)
- A set of scores can be presented in a column that is headed by X.

Y

- data for a particular variable (used when there are two sets of data)
- Each pair X, Y represents the observations made of a single participant.

N

- number of observations in the population
  - n = number of observations in the sample
  - n = 7, for both height and weight

The sigma sign " " is used to stand for summation. Many of the computations required statistics involve adding a set of scores.

- X means to add all the scores for variable X. It is read as "the sum of all X values."
  - E.g., The following set of quiz scores, 10, 6, 7, 4: X = 27 and n = 4

Review of order of mathematical operations:

1. Any calculation contained within parentheses.
2. Squaring (or raising to other exponents).
3. Multiplying and/or dividing is done third. A series of multiplication and/or division operations should be done in order from left to right.
4. Summation using the notation is done next.
5. Finally, any other addition and/or subtraction is done.

LESSON 1 PART 4: Frequency, Distribution & Tables

[Construct and interpret frequency distribution tables.]

FREQUENCY DISTRIBUTION

- ➢ A frequency distribution is an organized tabulation of the number of individuals located in each category on the scale of measurement.

PROPORTION

- ➢ Proportion refers to a part or share of a whole, expressed as a fraction or ratio of the part to the whole. It represents the relative size or amount of one component within a larger entity or group.

PERCENTAGE

- ➢ Percentage is a proportion or ratio expressed as a fraction of 100. It represents the relative size or amount of one component in relation to the whole, with each percentage point equivalent to one-hundredth of the total.

PERCENTILE

- ➢ A percentile is a measure used in statistics to indicate the value below which a given percentage of observations in a group fall. For example, the 70th percentile of a dataset indicates that 70% of the observations are below this value, while 30% are above it.

PERCENTILE RANK

- ➢ Percentile rank is the percentage of values in a dataset that are below a particular observation. It represents the relative position of a value within the dataset. For instance, if a student's score on a test is at the 75th percentile, it means that their score is higher than 75% of the scores in the dataset, and their percentile rank is 75.

GETTING THE PERCENTILE:

1. To determine percentiles or percentile ranks, the first step is to find the cumulative frequency – the number of individuals who are located at or below each point in the distribution.
2. We must convert these frequencies into percentages.(cumulative percentage).
3. The resulting values are called cumulative percentages because they show the percentage of individuals who are accumulated as you move up the scale.

LESSON 1 PART 5: Frequency, Distribution, and Graphs

[Create and analyze frequency distribution graphs.]

HISTOGRAM

- ➢ to construct a histogram, you first list the numerical scores or class intervals (the categories of measurement) along the X-axis. Then you draw a bar above each X value.
- ➢ Each bar in a histogram extends to the midpoint between adjacent categories. As a result, adjacent bars touch and there are no spaces or gaps between bars.

POLYGONS

- ➢ to construct a polygon, you begin by listing the numerical scores (the categories of measurement) along the X-axis. Then:
- ➢ A dot is centered above each score so that the vertical position of the dot corresponds to the frequency for the category.
- ➢ A continuous line is drawn from dot to dot to connect the series of dots.

LESSON 2 PART 1: ***Grouped Frequency Distribution***

[Create and analyze distribution graphs]

Why do we need to group the data?

- ➢ Grouping data into classes is important for simplifying large datasets and facilitating easier analysis.
- ➢ It helps in organizing raw data into meaningful categories or intervals, making it easier to identify patterns, trends, and outliers.
- ➢ Grouping also reduces the complexity of data presentation, making it more understandable to stakeholders.
- ➢ Additionally, grouping data into classes helps in summarizing and interpreting data effectively, allowing for clearer communication of findings.

CLASS NUMBER

- ➢ Class number refers to the sequential numbering or labeling of the different groups or intervals into which the data is divided.
- ➢ Each class represents a distinct range of values within the dataset.

CLASS INTERVAL

- ➢ Class interval refers to the range of values covered by each group or category in a frequency distribution.
- ➢ It divides the entire range of data into non-overlapping intervals.

CLASS WIDTH

- ➢ Class width is the difference between the upper and lower limits of a class interval.
- ➢ It determines the range of values (class interval) included in each group.

LESSON 2 PART 2 : ***Measures of Central Tendency***

[Understand the concepts of central tendency and compute them.]

01

- ➢ Central tendency is a statistical measure to determine a single score that defines the center of a distribution.

02

- ➢ The goal of central tendency is to find the single score that is most typical or most representative of the entire group.

ARITHMETIC MEAN

- ➢ The mean for a distribution is the sum of the scores divided by the number of scores.It is computed by adding all the scores in the distribution and dividing by the number of scores.

MEDIAN

- ➢ If the scores in a distribution are listed in order from smallest to largest, the median is the midpoint of the list.More specifically, the median is the point on the measurement scale below which 50% of the scores in the distribution are located.

MODE

- ➢ In a frequency distribution, the mode is the score or category that has the greatest frequency.In simple terms, mode is the most frequently occurring value in a given data set.

Comparing the Mean, Median, and Mode

- ➢ Usually the mean, median and mode will give different values of the central tendency when applied to the same set of scores.
- ➢ It is only when a distribution is perfectly symmetrical and the distribution peaks in the middle that they coincide completely.
- ➢ Regard big differences between the mean, median and mode as a sign that your distribution of scores is rather asymmetrical or lopsided.
- ➢ Distributions of scores do not have to be perfectly symmetrical for statistical analysis, but symmetry tends to make some calculations a little more accurate.

LESSON 2 PART 3 : Measures of Variability

[Understand the concepts of variability and compute them.]

01

- Variability provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.

02

- To say that things are variable means that they are not all the same. In statistics, if the scores in a distribution are all the same, then there is no variability.

03

- Variability describes the distribution of scores. Specifically, it tells whether the scores are clustered close together or spread out over a large distance.

04

- Variability measures how well an individual score (or group of scores) represents the entire distribution.

RANGE

- Range are easily understood indicators of the spread of scores on a variable. However, they only involve the extremes of your scores.

VARIANCE

- Variance is a statistical formula indicating spread which involves all of the scores in its calculation.

STANDARD DEVIATION

- It uses the mean of the distribution as a reference point and measures variability by considering the distance between each score and the mean.

RANGE AND INTERQUARTILE RANGE

RANGE

- it is the distance covered by the scores in a distribution,from the smallest to the largest score.

INTERQUARTILE RANGE

- one problem with range is that it can be heavily influenced by extreme cases (or outliers). For this reason, the interquartile range might be preferred as this basically ignores the extreme quarters of the distribution.
- Quartiles divide the distribution into four equal quarters such that each quarter corresponds to 25% of the distribution.
- To calculate the interquartile range, we split the age distribution into quarters (quartiles) and take the range of the middle two quarters (or middle 50%), ignoring the extreme quarters.
- The interquartile range is the range between the boundaries cutting off this middle 50% of scores from the 25% below and the 25% above.

AVERAGE/MEAN DEVIATION (AD)

- Useful as the range is, a lot of information is ignored when calculating the range. It merely is based on the two extreme scores at each end of the distribution.
- Other measures of spread or variability involve the extent to which every score differs from the mean score.

AVERAGE/MEAN DEVIATIONS

- it tells us the average deviations of all the scores away from the mean. It calculated by getting the specific distance between each score and the mean without regard whether the score is above or below the mean.
- It is rarely used since deletion of algebraic sign renders it as useless in any operations.

VARIANCE

- is much more useful and and has widespread and extensive applications.
- Variance is calculated in an almost identical way to average deviation but for one thing. When we draw up a table to calculate the variance, we square each deviation from the mean (instead of getting their absolute value) before summing the total of these squared deviations.

VARIANCE ($\sigma 2$)

- ➢ it can be calculated by summing the squared deviations and dividing the resulting value by the number of observation.

STANDARD DEVIATION

- ➢ it uses the mean of the distribution as a reference point and measures variability by considering the distance between each score and the mean.
- ➢ In simple terms, the standard deviation provides a measure of the standard, or average, distance from the mean, and describes whether the scores are clustered closely around the mean or are widely scattered.
- ➢ It is equal to the square root of variance.

RELATIONSHIP OF $\sigma 2$ AND σ

- ➢ The rationale for taking the square root of the variance to obtain the standard deviation lies in the need to express the standard deviation in the same units as the original data.
- ➢ The variance, which is the average of the squared differences between each data point and the mean, is expressed in squared units (e.g., square centimeters, square dollars).
- ➢ Taking the square root of the variance cancels out the squared units, resulting in a measure that is in the same units as the original data.
- ➢ This makes the standard deviation a more interpretable measure of dispersion or variability in the data, as it provides information about the spread of the data points around the mean in a more meaningful way.

LESSON 2 PART 4 : Sampling Error And Margin Of Error

[Understand and apply the concept of margin of error in data interpretation.]

SAMPLING ERROR

- ➢ Sampling error is the naturally occurring discrepancy, or error, that exists between a sample statistic and the corresponding population parameter.

MARGIN OF ERROR

- ➢ tells you how many percentage points your results will differ from the real population value.There always will be some “margin of error” when sample statistics are used to represent population parameters.

Ways to Calculate MoE:

1. Find the critical value of the confidence interval.
2. Find the Standard Error of Mean (SEM) using the standard deviation.
3. Multiply the critical value from Step 1 by the standard deviation or standard error from Step 2.