UNIT ONE

Part 2

Sham and Bilal

VARIABLES

1. A variable is an object, characteristic, or property that can have different values.
2. A quantitative variable can be measured in some way. Ex. Number of siblings.
3. A qualitative variable is characterized by its inability to be measured but it can be sorted into categories. Ex. Where are you from.

and each category is given a certain number.

*A research is mainly a relation between two variables.

TYPES OF VARIABLES

- Ideas which leads to make a research
*Mostly it's to answer a specific question which it is NOT answered.

Independent variable—the presumed cause (of a dependent variable)

Dependent variable—the presumed effect (of an independent variable)

Example: Smoking (IV) → Lung cancer (DV)

IV
Does makeup has a relation with developing skin cancer?!
DV

LEVELS OF MEASUREMENT

- **Nominal**
- **Ordinal**

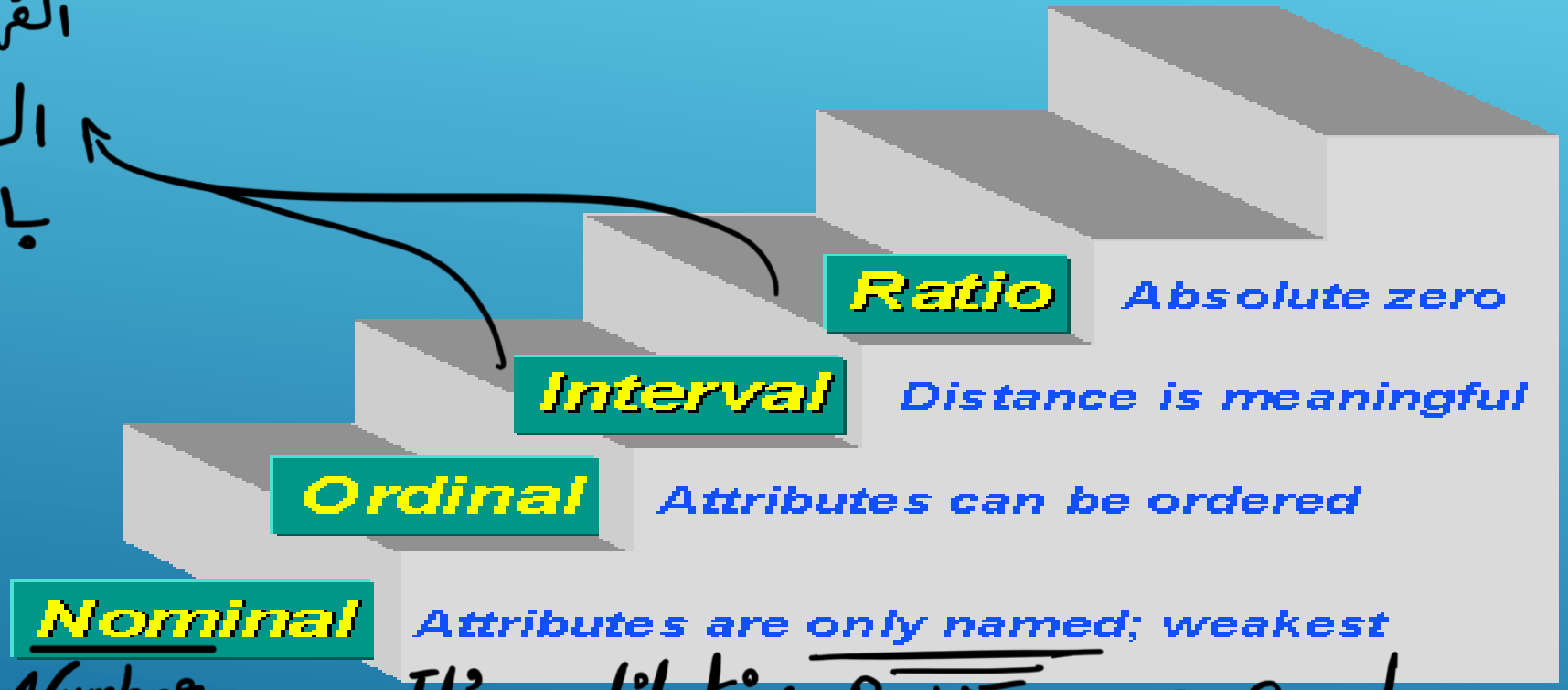
} Discrete/Categorical
*No fraction, it's yes or no / Male or Female.

- **Interval**
- **Ratio**

} Continuous
* Accepts fraction, as Glucose levels / Temprature.

LEVELS OF MEASUREMENT

الفرق الوحيد أنه :
الطولة معنى
بال Interval.



الأعلى

الأدنى

Number

It's qualitative BUT we represent it as numbers. * Without order.

NOMINAL LEVEL OF MEASUREMENT

- Categories that are distinct from each other such as gender, religion, marital status.
- They are symbols that have no quantitative value.
- Lowest level of measurement.
- Many characteristics can be measured on a nominal scale: race, marital status, and blood type.
- Dichotomous. Only two categories. Ex. Male/Female.
- Appropriate statistics: mode, frequency
- We cannot use an average. It would be meaningless here.

ORDINAL LEVEL OF MEASUREMENT

- The exact differences between the ranks cannot be specified such as it indicates order rather than exact quantity.
- Involves using numbers to designate ordering on an attribute.
- Example: anxiety level: mild, moderate, severe. Statistics used involve frequency distributions and percentages.
- Appropriate statistics: same as those for nominal data, plus the median; but not the mean.

1 2 3

3 > 2 > 1

No. has a mathmatical meaning

INTERVAL LEVEL OF MEASUREMENT

- They are real numbers and the difference between the ranks can be specified.
- Equal intervals, but no "true" zero. Doesn't mean absence.
- Involves assigning numbers that indicate both the ordering on an attribute, and the distance between score values on the attribute
- They are actual numbers on a scale of measurement.
- Example: body temperature on the Celsius thermometer as in 36.2, 37.2 etc. means there is a difference of 1.0 degree in body temperature.
- Appropriate statistics
  - same as for nominal
  - same as for ordinal plus,
  - the mean

RATIO LEVEL OF MEASUREMENT

- Is the highest level of data where data can categorized, ranked, difference between ranks can be specified and a true or natural zero point can be identified.
- A zero point means that there is a total absence of the quantity being measured.
- All scales, whether they measure weight in kilograms or pounds, start at 0. The 0 means something and is not arbitrary (SUBJECTIVE).
- Example: total amount of money

علامتك بالامتحان صفر ن

WHAT TYPE OF DATS TO COLLECT?

We collect data by asking quesntions

بأي طريقة كانت

- The goal of the researcher is to use the highest level of measurement possible.
  - Example: Two ways of asking about Smoking behavior. Which is better, A or B?
    - (A) Do you smoke? ▫Yes ▫No
    - (B) How many cigarettes did you smoke in the last 3 days (72 hours)?

(A) Is nominal, so the best we can get from this data are frequencies. (B) is ratio, so we can compute: mean, median, mode, frequencies.

PARAMETER AND STATISTIC

- Parameter is a descriptive measure computed from the data of the population.
  - The population mean, μ, and the population standard deviation, σ, are two examples of population parameters.
  - If you want to determine the population parameters, you have to take a census of the entire population.
  - Taking a census is very costly.
  - Parameters are numerical descriptive measures corresponding to populations.
  - Since the population is not actually observed, the parameters are considered unknown constants.
- Statistic is a descriptive measure computed from the data of the sample.
  - For example, the sample mean, $\bar{x}$, and the standard deviation, s, are statistics.
  - They are used to estimate the population parameters.

STATISTICS

- It is a branch of applied mathematics that deals with collecting, organizing, & interpreting data using well-defined procedures in order to make decisions.
- The term parameter is used when describing the characteristics of the population. The term statistics is used to describe the characteristics of the sample.
- Types of Statistics:
  - Descriptive Statistics. It involves organizing, summarizing & displaying data to make them more understandable. To make sense of your sample characteristics.
  - Inferential Statistics. It reports the degree of confidence of the sample statistic that predicts the value of the population parameter Conclusion regarding the Population

معلومات معقدة أكثر.

DESCRIPTIVE STATISTICS

1. Measures of Location
   - A. Measures of Central Tendency:
     - Mean
     - Median
     - Mode
   - B. Measures of noncentral Tendency-Quantiles:
     - Quartiles.
     - Quintiles.
     - Percentiles.
2. Measure of Dispersion (Variability):
   - Range
   - Interquartile range
   - Variance
   - Standard Deviation
   - Coefficient of variation
3. Measures of Shape:
   - Mean > Median-positive or right Skewness
   - Mean = Median- symmetric or zero Skewness
   - Mean < Median-Negative of left Skewness

STATISTICAL INFERENCE

- Is the procedure used to reach a conclusion about a population based on the information derived from a sample that has been drawn from that population.

INFERENTIAL STATISTICS

- Inferential statistics are used to test hypotheses (prediction) about relationship between variables in the population. A relationship is a bond or association between variables.
- It consists of a set of statistical techniques that provide prediction about population characteristics based on information in a sample from population. An important aspect of statistical inference involves reporting the likely accuracy, or of confidence of the sample statistic that predicts the value of the population parameter.

At the end you can prove or deny your hypotheses.

INFERENTIAL STATISTICS

- Bivariate Parametric Tests:
  - One Sample t test (t)
  - Two Sample t test (t)
  - Analysis of Variance/ANOVA (F).
  - Pearson's Product Moment Correlations (r).
- Nonparametric statistical tests: Nominal Data:
  - Chi-Square Goodness-of-Fit Test
  - Chi-Square Test of Independence
- Nonparametric statistical tests: Ordinal Data:
  - Mann Whitney U Test (U
  - Kruskal Wallis Test (H)

RESEARCH HYPOTHESIS

- A tentative prediction or explanation of the relationship between two or more variables
- It's a translation of research question into a precise prediction of the expected outcomes
- In some way it's a proposal for solution/s
- In qualitative research, there is NO hypothesis

RESEARCH HYPOTHESIS

- States a prediction
- Must always involve at least two variables
- Must suggest a predicted relationship between the independent variable and the dependent variable
- Must contain terms that indicate a relationship (e.g., more than, different from, associated with)

HYPOTHESES CRITERIA

- Written in a declarative form.
- Written in present tense.
- Contain the population NOT samples. (imp.)
- Contain variables.
- Reflects problem statement or purpose statement.
- Empirically testable. (وقابلية)

$H_A : P_1 \neq P_2$

$H_o : P_1 = P_2$

1: Make-up using
2: Skin cancer.

HYPOTHESIS TESTING

* دائمًا نطرح ونتأمل لرفض الـ H_0.

- A hypothesis is made about the value of a parameter, but the only facts available to estimate the true parameter are those provided by the sample. If the statistic differs (and of course it will) from the hypothesis stated about the parameter, a decision must be made as to whether or not this difference is *significant*. If it is, the hypothesis is rejected. If not, it cannot be rejected.
- H_0: The null hypothesis. This contains the hypothesized parameter value which will be compared with the sample value.
- H_1: The alternative hypothesis. This will be "accepted" only if H_0 is rejected.

  Technically speaking, we never accept H_0 What we actually say is that we do not have the evidence to reject it.

TWO TYPES OF ERRORS: ALPHA AND BETA

*1-α is the level of confindance(how true is our results)).

*1-β is Power.

Acceptable margine of error

5% is acceptable. I

II

20% is acceptable.

- Two types of errors may occur: α (alpha) and β (beta). The α error is often referred to as a Type I error and β error as a Type II error.
  - You are guilty of an alpha error if you reject H_0 when it re
  - You c ou "accept" H_0 when it is false

		STATE OF NATURE	
		H_0 Is True	H_0 Is False
DECISION	Do Not Reject H_0	GOOD	β Error (Type II Error)
	Reject H_0	α Error (Type I Error)	GOOD

*Making the α smaller increases the difficulty in sampling.

TYPES OF ERRORS

If You……	When the Null Hypothesis is…	Then You Have…….
Reject the null hypothesis	True (there really are no difference)	Made a Type I Error F+
Reject the null hypothesis	False (there really are difference)	☻
Accept the null hypothesis	False (there really are difference)	Made Type II Error F−
Accept the null hypothesis	True (there really are no difference)	☻

5% → More dangerous.

20% → Less dangerous.

STEPS IN HYPOTHESIS TESTING

1. Formulate H_0 and H_1. H_0 is the null hypothesis, a hypothesis about the value of a parameter, and H_1 is an alternative hypothesis.

   e.g., H_0: μ=12.7 years; H_1: $\mu \neq$12.7 years

2. Specify the level of significance (α) to be used. This level of significance tells you the probability of rejecting H_0 when it is, in fact, true. (Normally, significance level of 0.05 or 0.01 are used)
3. Select the test statistic: e.g., Z, t, F, etc.
4. Establish the critical value or values of the test statistic needed to reject H_0. DRAW A PICTURE!
5. Determine the actual value (computed value) of the test statistic.
6. Make a decision: **Reject H_0** or **Do Not Reject H_0**.

More favorable/stronger BUT

PARAMETRIC ASSUMPTIONS

- The observations must be independent.
- Dependent variable should be continuous (I/R)
- The observations must be drawn from normally distributed populations
- These populations must have the same variances. Equal variance (homogeneity of variance)
- The groups should be randomly drawn from normally distributed and independent populations

e.g. Male X Female

Pharmacist X Physician

Manager X Staff

NO OVERLAP

PARAMETRIC ASSUMPTIONS

- The independent variable is categorical with two or more levels.
- Distribution for the **two or more independent** variables is normal.

ADVANTAGES OF PARAMETRIC TECHNIQUES

- They are more powerful and more flexible than nonparametric techniques.
- They not only allow the researcher to study the effect of many independent variables on the dependent variable, but they also make possible the study of their interaction.

*We compare means; $\mu_1 \stackrel{?}{=} \mu_2$

NONPARAMETRIC METHODS

*No assumptions BUT less stronger/flexibale.

- Nonparametric methods are often the only way to analyze nominal or ordinal data and draw statistical conclusions.
- Nonparametric methods require no assumptions about the population probability distributions.
- Nonparametric methods are often called distribution-free methods.
- Nonparametric methods can be used with small samples

NONPARAMETRIC METHODS

- In general, for a statistical method to be classified as nonparametric, it must satisfy at least one of the following conditions.
  - The method can be used with nominal data.
  - The method can be used with ordinal data.
  - The method can be used with interval or ratio data when no assumption can be made about the population probability distribution (in small samples).

NON PARAMETRIC TESTS

- Do not make as many assumptions about the distribution of the data as the parametric (such as *t* test)
  - Do not require data to be Normal
  - Good for data with outliers
- Non-parametric tests based on ranks of the data
  - Work well for ordinal data (data that have a defined order, but for which averages may not make sense).

NONPARAMETRIC METHODS

- There is at least one nonparametric test equivalent to each parametric test
- These tests fall into several categories
  1. Tests of differences between groups (independent samples)
  2. Tests of differences between variables (dependent samples)
  3. Tests of relationships between variables

*:. We compare proportions; $P_1 \stackrel{?}{=} P_2$

TABLE OF STATISTICAL TESTS

Level of Measurement (D.V.)	Sample Characteristics (I.V.)					Correlation
	1 Sample	2 Sample		K Sample (i.e., >2)		
		Independent (2 groups independent from each other)	Dependent	Independent	Dependent	
Categorical or Nominal	X^2	X^2	Macnamar's X^2	X^2	Cochran's Q	
Rank or Ordinal		Mann Whitney U	Wilcoxin Matched Pairs	Kruskal Wallis H	Friendman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1 way ANOVA between groups	1 way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

SUMMARY: PARAMETRIC VS. NONPARAMETRIC STATISTICS

- Parametric Statistics are statistical techniques based on assumptions about the population from which the sample data are collected.
  - Assumption that data being analyzed are randomly selected from a normally distributed population.
  - Requires quantitative measurement that yield interval or ratio level data.
- Nonparametric Statistics are based on fewer assumptions about the population and the parameters.
  - Sometimes called "distribution-free" statistics.
  - A variety of nonparametric statistics are available for use with nominal or ordinal data.