

BRIDGE INTERNATIONAL COLLEGE

- Introduction To Statistics
- *COURSE COD: STAT 2171 /2051*
- *CREDIT HOURS: 3*

Presentation outlines

- **Unit 3 Measures Of Central Tendency**
- Technical Note: Summation Notation
- 3.1 Definition And Purpose
- 3.2 The Mean
- 3.2.1 The Arithmetic Mean: For Ungrouped And Grouped Data
- 3.2.2 The Weighted Mean And The Grand Mean
- 3.3. Median, Mode And Quartiles For Ungrouped Data
- **Unit 4 Measures Of Dispersion**
- 4.1 Definition ,
- **Measures Of Relative Variation**
- **Coefficient of Variation (CV)**
- 4.2. The Range, Variance And Standard Deviation.
- **Unit 5 Simple Linear Regression and Correlation**
- **5.1 Introduction**
- **5.2 Correlation Analysis**
- **5.3 Correlation coefficient**

UNIT THREE

- **Measures of Central Tendency**
- Measures of central tendency are the values that describe a data set by identifying the central position of the data. There are 3 main measures of central tendency - Mean, Median and Mode.
- Mean- Sum of all observations divided by the total number of observations.
- Median- The middle or central value in an ordered set.
- Mode- The most frequently occurring value in a data set.

Continue----

- **EXAMPLE**
- **Mean:** The mean salary of the given salary can be used as one of the measures of central tendency, i.e., $\bar{x} = (105 + 95 + 105 + 105 + 100)/5 = \$102$.
- **Mode:** If we use the most frequently occurring value to represent the above data, i.e., \$105, the measure of central tendency would be mode.
- **Median:** If we use the central value, i.e., \$105 for the ordered set of salaries, given as, \$95, \$100, \$105, \$105, \$105, then the measure of central tendency here would be median.

Cont,...

- **We can use the following table for reference to check the best measure of central tendency suitable for a particular type of variable:**

Type of Variable	Best Measure of Central Tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean

Cont,...

- **Example:** The weights of 8 boys in kilograms: 45, 39, 53, 45, 43, 48, 50, 45. Find the median.
- **Solution:**
- Arranging the given data set in ascending order: 39, 43, 45, 45, 45, 48, 50, 53
Total number of observations = 8
For even number of observation, Median = $[(n/2)\text{th term} + ((n/2) + 1)\text{th term}]/2$
 $\Rightarrow \text{Median} = (4\text{th term} + 5\text{th term})/2 = (45 + 45)/2 = 45$
- For grouped data, the sum of all values is obtained by multiplying the frequency or percentage of occurrence by the value of the variable.
- There are three main methods of calculating the mean of grouped data, they are - ***direct method, assumed mean method, and step deviation method.***

- **Arithmetic Mean of Ungrouped Data**
- For ungrouped data, we can easily find the arithmetic mean by adding all the given values in a data set and dividing it by a number of values.
- **What is the Mean Formula for Ungrouped Data?**
- The mean formula to find the mean for an ungrouped set of data can be given as, $\text{Mean} = (\text{Sum of Observations}) \div (\text{Total Numbers of Observations})$

Cont....

Grand mean

- The grand mean or pooled mean is ***the average of the means of several subsamples***, as long as the subsamples have the same number of data points.
- **What does the grand mean mean?**
- **a numerical average (mean) of a group of averages.** For example, if the average test score for one classroom is 75 and the average score for another classroom is 73, the grand mean of the two classrooms is $(75 + 73)/2 = 74$.

- The grand mean is **the mean of the means you have already recorded**. It's not achieved by dividing the total number of sets, but rather the total group sets within specific data. Determine the mean of each group or set's samples. Remember the formula for mean (sum of data divided by the number of data).
- **Quartiles** are the values of individual arranged **data** which divide whole **data** into four equal parts. They are 3 in numbers namely Q1, Q2 and Q3. Here Q1 is first **quartile**, **Q2 is the 2nd quartile and Q3 is the 3rd quartile**.
- .

What is grand mean vs. group mean?

- ❖ Group mean centering will produce intercepts weighted by the proportion of 1 to 0 values for each group, whereas grand-mean centering will produce intercepts weighted by the proportion of 1 to 0 in the entire sample.

Cont, ...

- **Unit 4 Measures Of Dispersion.**

Definition

- There is another name for the measure of variability that is the measure of dispersion. It is used to portray variability in a sample or population. Also, there are three popular measures of variability.
- The first one is a **range**, which in simple language means maximum value minus minimum value.
- The second is **variance**, which clearly illustrates how much a spontaneous variable is different from the anticipated value.
- The third one is **Standard Deviation (S.D.)**, which is a measure of the dispersion of a set of data from its mean.
- In statistics, the measures of dispersion help to **interpret the variability (spread) of data** i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable.

Cont,...

- This single number is called a 'measure of dispersion'. Dispersion: “**The variability (spread) that exists between the values of a data** is called. Dispersion” OR “The extent to which the observations are spread around an average is called dispersion or scatter”
- Examples of positional measures of dispersion are: Standard Scores, Quartiles, Five Number Summary and Outliers. Computational measures are the measures of spread of data about the mean. These include Standard Deviation, Range, \$ Mean Deviation .

Measures Of Relative Variation

- Relative measures of dispersions are a ratio or percentage of a measure of absolute dispersion to an appropriate measure of central tendency and are thus pure numbers independent of the units of measurement.
- For comparing the variability of two distributions (even if they are measured in the same unit), we compute the relative measure of dispersion instead of absolute measures of dispersion.

Coefficient of Variation (CV)

- Is defined as the ratio of standard deviation to the mean usually expressed as percent. That is, $CV = \frac{s}{\bar{x}} \times 100\%$
- Example** : An analysis of the monthly wages paid (in birr) to workers in two firms A ten times observations and B eleven times observations belonging to the same industry gives the following results :

Value	Firm A	Firm B
Mean wage	52.5	47.5
Median wage	50.5	45.5
Variance wage	100	121

Cont,...

- In which firm A or B is there greater variability in individual wage?
- **Solution:** calculate coefficient of variation for both firms.
- $CV_A = \frac{s_A}{\bar{w}_A} \times 100\% = \frac{3.81}{19.9} \times 100\% = 19.05\%$
- $CV_B = \frac{s_B}{\bar{w}_B} \times 100\% = \frac{5.37}{23.16} \times 100\% = 23.16\%$
- Since $CV_A < CV_B$ in firm B there is greater variability in individual wages. Appropriate as it may be influenced by some extreme values.

- **Unit 5 Simple Linear Regression and Correlation**
- **5.1 Introduction**
- Regression analysis, in the general sense, means the estimation of or prediction of the unknown values of one variable from known values of the other variables. In Regression analysis there are two types of variables. The variable whose value is influenced or to be predicted is called dependent (regressed or explained) variable and the variable which influences the values or is used for prediction, is called independent variable (repressor or predictor or explanatory).

Cont,...

- If the regression curve is a straight line, we say that there is linear relationship between the variables under study, non-linear elsewhere. When only two variables are involved, the functional relationship known as simple regression.
- If the relationship between the two variables is a straight line, it is known as simple **linear** regression: otherwise, it is called as simple **non-linear regression**.
- When there are more than two variables and one of them dependent up on the other, the functional relationship between the variables is known as **multiple**

Cont,...

Example:

- **Regression analysis is performed if one wants to know relationship between**
- A, Income – consumption
- B, Sales of ice-cream – with temperature of the day
- C, Industrial production and consumption of electricity
- D, The yield of crops, amount of rainfall, type of fertilizer, humidity... etc.

CONTINUE.....

Simple linear regression

- Under simple linear regression of Y on X , we have one independent variable which is influential usually denoted by X and one dependent variable influenced denoted by Y .
- **For example**, in real world variables that may be related linearly are, production/yield (**Y**) and amount of rain fall (**X**) monthly income (**Y**) and level of education (X), where an increase in one variable is associated with an increase in the other variable. Similar examples can also be given on the negative relation between two variables; the increase in one is accompanied by a decrease in the

CONTI,...

- The simple linear regression of Y on X in the population is given by
- $Y = \alpha + \beta X + \epsilon$
- Where α = y intercept, β = regression coefficient and ϵ = error term.

Correlation Analysis

- Correlation analysis is concerned with measuring the strength (degree) of the relationship between two or more variables. It is used if we are interested in knowing the extent of interdependence between two or more variables.

CONTI....

Correlation Coefficient

- The correlation coefficient is a statistical measure of the strength of a linear relationship between two variables. Its values can range from -1 to 1.
- A Correlation coefficient of -1 describes a perfect negative, or inverse, correlation, with values in one series rising as those in the other decline, and vice versa.
- A coefficient of 1 shows a perfect positive correlation, or a direct relationship. A correlation coefficient of 0 means there is no linear relationship.
- Correlation coefficients are measures of association between two or more variables. Correlation is a measure of association that tests whether a relationship exists between two variables. It indicates

- - Correlation coefficients are used to assess the strength of associations between data variables.
 - Values are range from -1 for a perfectly inverse, or negative, relationship to 1 for a perfectly positive correlation .
 - Values at ,or close to zero indicate no linear relationship or a very weak correlation.

Cont,...



Cont,...



Continue



continue ...



Cont,...



cont,...



Cont,...



Cont,...

-

Cont,...







YOU!

THANK