CHAPTER 2

INTRODUCTION TO FREQUENCY DISTRIBUTION

YAKUP ARI,Ph.D.(C)

math.stat.yeditepe@gmail.com

FREQUENCY DISTRIBUTION

- The most common procedure for organizing and simplyfing a set of data is to place them in a frequency distribution.

- Frequency distribution refers to **an organized tabulation of the number of individuals located in each category on the scale of measurement.**

FREQUENCY DISTRIBUTION

- It enables the researcher to see whether the scores are high or low, whether they are concentrated in one area or spread out across the entire set.

- Moreover, it allows the researcher to see the location of any individual score relative all of the other scores in the set.

FREQUENCY DISTRIBUTION

- A frequency distribution can be structured either a **graph** or a **table.**

- It has **two** elements:
  - The set of categories that make up the original measurement scale. They are listed in a column from highest to lowest. We use an X as the column heading.

  - A record of frequency or number of individuals in each category. They are listed in a column. We use an f as the column heading.

FREQUENCY DISTRIBUTION

- There are other measures that describe the distribution of scores. The two most common are **percentages** and **proportions.**

- Proportion measures **the fraction of the total group that is associated with each score.**
  - They generally appear as decimals
  - A column headed with a p can be added to frequency distribution.

FREQUENCY DISTRIBUTION

- There are other measures that describe the distribution of scores. The two most common are **percentages** and **proportions.**

- Percentages are computed by the multiplication of the proportion by 100.
  - A column headed with a **%** can be added to frequency distribution.

GROUPED FREQUENCY DISTRIBUTION TABLES

- When a set of data covers a wide range of values, it is unreasonable to list all the individuals scores in a frequency distribution table.
- In this case, we group the scores into **intervals** in order to obtain a relatively simple and organized picture of data.
- The result is a **grouped frequency distribution table**.
  - The groups are class intervals.

GROUPED FREQUENCY DISTRIBUTION TABLES

- There are some rules that we should take into consideration in the construction of a grouped frequency distribution table:
1) It should have about 10 class intervals.
2) The width of each interval should be a relatively simple number. For instance, 2,5,10, or 20 would be a good choice.
3) The bottom score in each class interval should be a multiple of the width.
4) All intervals should be the same width.

FREQUENCY DISTRIBUTION GRAPHS

- All graphs have two perpendicular lines called axes.
  - The horizontal line is called the **X-axis.**
  - The vertical line is called the **Y-axis.**

- **The measurement sca**le is listed along the X-axis
- **The frequencies** are listed on the Y-axis.

- Two axes intersect should have a value of **zero**

GRAPHS FOR INTERVAL OR RATIO SCALE

- There are **two** options for constructing a graph for an interval or a ratio scale.
  - **Histograms**
    - Height of the bar corresponds to the **frequency** for that category
    - The width of the bar extends to the **real limits** of the category
    - Since we extend the bars to real limits for each category, there are no spaces or gaps between bars.
  - **Polygons**

GRAPHS FOR NOMINAL OR ORDINAL SCALE

- When we use **nominal** or **ordinal** scale, the frequency distribution can be displayed in **a bar graph**.

- A bar graph is same as histogram however **there are spaces left between adjacent bars.**

  - For nominal data, the space between bars indicates that the scale consists of seperate, distinct categories.

  - For ordinal data, seperate bars are used because you cannot assume that the categories are all the same size.

  - X-axis⇨the categories of measurement

  - Y-axis⇨frequency for the category

THE SHAPE OF A FREQUENCY DISTRIBUTION

- There are three features that completely describe any distribution
  - **Shape**
  - **Central Tendecy**
    - Measures where the center of distribution is located
  - **Variability**
    - Measures whether the scores are spread over a wide range or clustered together

THE SHAPE OF A FREQUENCY DISTRIBUTION

- Nearly all distributions can be classified in **two** categories:
  - **Symmetrical**
    - When we draw a vertical line through the middle, one side of the distribution is a mirror image of the other.
  - **Skewed**
    - The scores tend to pile up toward one end of the scale and taper off gradually at the other end
      - Positively Skewed
      - Negatively Skewed

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- Frequency Distribution Tables also are used to describe the position of an individual within the set.

- Individual scores are raw scores and they do not provide much information. In other words, they do not tell the relative position of the score in the distribution.

- We transform raw score scores into a more meaningful form.

- One way of transformation is changing raw scores into percentiles

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- **The percentile rank** refers to a particular score is defined as the percentage of individuals in the distribution with scores **at or below** the particular value. Notice that **percentile rank refers to a percentage**

- When a score is identified by its percentile rank, the score is called a **percentile**. Notice that **percentile refers to a score.**

- Both the percentile rank and percentile describes your exact position within the distribution.

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- CUMULATIVE FREQUENCY AND CUMULATIVE PERCENTAGE
  - The first step is to **find the number of indviduals who are located at or below each point in the distribution**. That is to say, we find **cumulative frequencies** that refers to **the accumulation of individuals as you move up the scale**.
  - We add a column headed by *cf*
  - For each row, the cumulative frequency value is obtained by adding up the frequencies in and below that category

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- CUMULATIVE FREQUENCY AND CUMULATIVE PERCENTAGE
  - To find the percentiles, we must convert these frequencies into percentages. The resultin values are called **cumulative percentages.**
  - They indicate the percentage of individuals who are accumulated as you move up the scale.
  - We added a new column headed by c%.
  - It gives the percentage of individuals with at or below each X value.
  - **Notice that X values are generally measurements of a continuous variable therefore we use the real limits of each score.**
  - **Each cumulative percentage value is associated with upper real**

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- **Interpolation**

  - For some value, we can determine the percentiles and percentile ranks directly from frequency distribution table.

  - However, there are many values that do not appear directly in the table and it is imposible to determine these values precisely.

  - We obtain **estimates** of these values by using *interpolation* procedure.

PERCENTILES, PERCENTILE RANKS, AND INTERPOLATION

- ## General Process of Interpolation
  - Find the width of the interval on both scales
  - Locate the position of the intermediate value in the interval. This position corresponds to a fraction of the whole interval:
    - **Fraction=Distance from the top of the interval/interval width**
  - Use this fraction to determine the distance from the top of the interval on the other scale.
    - **Distance=fraction x width**
  - Use the distance from the top to determine the position on the other scale

STEM AND LEAF DISPLAYS

- Tukey (1977) developed a technique which is known as stem and leaf displays in order to organize data.

- In this technique, each score is separated into two parts:
  - **The first digit (digits) is called the stem**
  - **The last digit (or digits) is called the leaf**

- In order to construct a stem and leaf display,
  - First, we list all stems in a column.
  - The next step is to go through the data, one score at a time, and write the leaf for each score beside its stem.
  - This process continues for the entire set.

STEM AND LEAF DISPLAYS

- It has several advantages compared to frequency distribution table
  - It is easy to construct.
  - It allows you to identify every individual score in the data.
  - It provides both a listing of the scores and a picture of the distribution.