Descriptive Statistics

Common Measures

1. Mean (Average)

a. **Definition**: The mean is the sum of all the values divided by the number of values.

b. **Use**: The mean provides a measure of the central tendency of the scores, representing average values (mean age of respondents, mean number of household members, mean income, etc.)

c. E**xample:**

Consider the data set: 5, 7, 3, 8, 10

In the baseline/endline dataset, these values could correspond to the number of years of formal education reported by each respondent.

- **Calculation**:

$$\text{Mean} = \frac{5 + 7 + 3 + 8 + 10}{5} = \frac{33}{5} = 6.6$$

- **Calculation in Excel**:

Formula: **=AVERAGE(range)**

Example: If your data is in cells A1 to A10, you would use **=AVERAGE(A1:A10)**.

2. Median

a. **Definition**: The median is the middle value in a list of numbers sorted in ascending order. If the list has an even number of values, the median is the average of the two middle numbers.

b. **Use**: The median indicates the middle score when the data is ordered, giving a sense of the central value that is less affected by outliers[1].

c. E**xample:**

Consider the data set: 5, 7, 3, 8, 10

---

[1] Outliers are data points that significantly differ from the other observations in a dataset. They are unusually high or low values that lie outside the overall pattern, potentially indicating variability, measurement errors, or unique occurrences. Detecting and analyzing outliers is crucial as they can significantly impact statistical measures and the conclusions drawn from the data.

Again, these values could correspond to the number of years of formal education reported by each respondent in the baseline/endline dataset.

- **Calculation:**

    Ordered data set: 3, 5, **7**, 8, 10. Therefore, median=7

- **Calculation in Excel:**

    Formula: **=MEDIAN(range)**

    Example: If your data is in cells A1 to A10, you would use **=MEDIAN(A1:A10).**

3. Mode

a. **Definition**: The mode is the value that appears most frequently in a data set. A set may have one mode, more than one mode, or no mode at all.

b. **Use**: The mode identifies the most frequently occurring score. For instance, in the baseline/endline dataset, we can use it to show the most common income level of households in the sample, or the most common age level of respondents.

c. E**xample:**

Consider the data set: 5, 7, 3, 8, 5, 10.

In the baseline/endline data set, these values could correspond to the total number of household members reported by respondents.

- **Calculation:**

    Mode= 5 (since 5 appears twice)

- **Calculation in Excel:**

    Formula: **=MODE.SNGL(range)**

    Example: If your data is in cells A1 to A10, you would use **=MODE.SNGL(A1:A10)**.

4. Standard Deviation

a. **Definition**: The standard deviation measures the amount of variation or dispersion in a set of values. A low standard deviation indicates that the values

tend to be close to the mean, while a high standard deviation indicates that the values are spread out over a wider range.

A more detailed explanation can be found here.

b. **Use**: The standard deviation measures the spread of the scores around the mean. In the baseline/endline dataset, it can be used to show the variability in household income.

c. E**xample:**

Consider the data set: 5, 7, 3, 8, 10

- **Calculation:**

$$\text{Standard Deviation} = \sqrt{\frac{(5 - 6.6)^2 + (7 - 6.6)^2 + (3 - 6.6)^2 + (8 - 6.6)^2 + (10 - 6.6)}{5 - 1}}$$

$$\text{Standard Deviation} \approx 2.77$$

- **Calculation in Excel:**

Formula: **=STDEV.S(range)** (for a sample) or **=STDEV.P(range)** (for a population)

Example: If your data is in cells A1 to A10, you would use **=STDEV.S(A1:A10).**

4. Frequency Distributions

a. **Definition**: A frequency distribution is a summary of how often each value occurs in a dataset.

A more detailed explanation can be found here.

b. **Use:** Frequency distributions display how often each score occurs, helping to visualize the overall distribution and identify patterns in the data. For instance, in the baseline/endline dataset, a frequency distribution can show how many respondents fall into each age group (e.g., 20-29, 30-39, etc.), helping to identify the most common age ranges among participants.

c. **Example:**

Consider the data set: 70, 75, 80, 85, 90, 70, 80, 85, 90, 90. In the baseline/endline dataset, these values could represent the age of each respondent, as reported by them.

- **Calculation:**

DANISH
REFUGEE
COUNCIL

Funded by
European Union
Humanitarian Aid

o **Define bins** (score ranges): in this case, individual scores can be used as bins, as the range is not too large: 70, 75, 80, 85, 90.
o **Count the frequency of each score**:

> 70: 2 times
> 75: 1 time
> 80: 2 times
> 85: 2 times
> 90: 3 times

- **Calculation in Excel:**

  o **Creating a Frequency Table**:
  Step 1: Create a column with bins (ranges of values).
  Step 2: Use the FREQUENCY function.
  Example: If your data is in cells A1 to A10 and your bins are in cells B1 to B5:
  Select the range where you want the frequencies to appear (e.g., C1).
  Enter the formula **=FREQUENCY(A1:A10, B1:B5).**
  Press Ctrl+Shift+Enter to enter it as an array formula.

  o **Using Pivot Table:**
  Step 1: Select your data range.
  Step 2: Go to Insert > PivotTable.
  Step 3: Drag the data field to the Rows area and the same data field to the Values area, changing the Value Field Settings to count.