Numerical Descriptive Measures for Quantitative Data I

Dr. Tom Ilvento
FREC 408

Measures of Central Tendency

- The **central tendency** of a variable is the tendency of the data to cluster or center about certain numerical values
- The **variability** is the spread of the data
- For central tendency we will focus on the **mean**, the **mode**, and the **median**

We need some tools

- Sigma Notation (p. 131)
- This relates to the rules of summation
- We start with the Greek Sigma Σ

$$\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i = x1 + x2 + x3 + x4 + x5 \ldots xn$$

Sigma Notation

- For a variable X with the set
- {5, 7, 4, 3, 2, 5}
- $$\sum_{i=1}^{n} x_i = 5+7+4+3+2+5 = 26$$

Rules of Sigma Notation

- The sum of a constant c

$$\sum_{i=1}^{n} c = n \cdot c$$

Rules of Sigma Notation

- The sum of a constant times x

$$\sum_{i=1}^{n} (c \cdot x_i) = c \sum_{i=1}^{n} x_i$$

Rules of Sigma Notation

Let c = 5 and x be the set {2, 4, 5, 2}

$$\sum_{i=1}^{n} 5 \cdot x_i = (5\cdot 2)+(5\cdot 4)+(5\cdot 5)+(5\cdot 2) = 65$$

$$5\sum_{i=1}^{n} x_i = 5\cdot(2+4+5+2) = 5\cdot 13 = 65$$

Rules of Sigma Notation

- The sum of the addition of two variables, x and y

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Rules of Sigma Notation

Let:
x be the set {2, 4, 5, 2} and
y be the set {5, 3, 2, 1}

$$\sum_{i=1}^{n}(x_i + y_i) = (2+5)+(4+3)+(5+2)+(2+1) = 24$$

$$\sum_{i=1}^{n} x_i = (2+4+5+2) +$$

$$\sum_{i=1}^{n} y_i = 13+11 = 24$$

Rules of Sigma Notation

- Let a and c represent constants

$$\sum_{i=1}^{n}(a\cdot x_i + c) = a\sum_{i=1}^{n} x_i + n\cdot c$$

Rules of Sigma Notation

- The sum of the addition of two variables squared

$$\sum_{i=1}^{n}(x_i + y_i)^2 =$$

$$\sum_{i=1}^{n}(x_i^2 + 2x_i y_i + y_i^2) =$$

$$\sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n}(x_i \cdot y_i) + \sum_{i=1}^{n} y_i^2$$

Rules of Sigma Notation

- **Note**

$$\sum_{i=1}^{n}(x_i + y_i)^2 \neq \sum_{i=1}^{n} x^2 + \sum_{i=1}^{n} y^2$$

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

The Mean

- The arithmetic mean or mean (Def 3.1 p133) is the sum of the measurements divided by the number of measurements contained in the data set
- For a sample we use x with a bar over it
- $\bar{x}$
- For a population, we use the Greek ∝

The Mean

- Two ways to express the mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The sum of all the values, divided by the number of values

$$\bar{x} = \sum_{i=1}^{n} (x_i / n)$$

The sum of values weighted by the number of values

Properties of the Mean

- As a measure of central tendency the mean has several advantages:
- The first is that the mean uses information of all the values in a variable

Properties of the Mean

- The mean has two important mathematical properties:
  - The sum of the deviations about the mean equals zero
  - The sum of squared deviations about the mean is a minimum;

Properties of the mean

- Sum of deviations equal zero

Here's the proof!

1. $\sum_{i=1}^{n}(x-\bar{x}) = 0$
2. $\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = 0$
3. $\sum_{i=1}^{n} x_i - n \cdot \frac{\sum_{i=1}^{n} x_i}{n} = 0$ $\quad \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$

Properties of the mean

- Sum of squared deviations about the mean is a minimum – **Least Squares property**

$$\sum_{i=1}^{n} (x_i - \bar{x})^2$$

There is no other value or constant we could substitute in the equation for the mean that would result in a lower sum of squares