FOUNDATIONS OF STATISTICS – STA10003 (WEEK 3)

**A. Variability (Biến thiên)**

**1. Definition:** Variability in statistics refers to the extent to which data points in a dataset differ from each other and from the center of the distribution. It's essentially a measure of how spread out or dispersed the data is.

**2. The importance of variability:**

**- Understanding the data:** Variability shows us how spread out the numbers are, giving us a better sense of the whole picture.

**- Making decisions:** Knowing how much things can change (like stock prices or weather) is key to making good choices.

**- Statistical analysis:** To compare groups, we need to see how different the values are within each group.

**3. Key measures of variability**:

- Range (Phạm vi):

+ The **range** is the simplest way to measure how spread out a set of data is. It's calculated by finding the difference between the highest and lowest values in the dataset.

**Range = Xmax - Xmin**

**-** Variance (Phương sai):

+ Standard deviation (Độ lệch chuẩn): is the most common measure of variability in statistics. It quantifies how much each score in a dataset deviates from the mean

* **Starting Point:** The mean serves as the reference point for calculating standard deviation.
* **Deviation Score (Điểm lệch):** The deviation score (X - μ) represents the difference between an individual score (X) and the population mean (μ).
* **Standard deviation (Độ lệch chuẩn) =**

+ Variance (Phương sai) = mean of squared deviations:

* Just adding the deviation scores adds to zero
* Average of squared distance from the mean
* Definitional calculation / Computational calculation

|  |  |  |
| --- | --- | --- |
| **Scores [*X*]** | **Deviation**  **[*X* - μ]** | **Deviation Score** |
| 2 | 2-4 | -2.00 |
| 1 | 1-4 | -3.00 |
| 9 | 9-4 | 5.00 |
| 4 | 4-4 | 0.00 |

Σ = 2 + 1 + 9 + 4 = 16

Ν = 4

=> μ =

+>

**Definition Forluma: SS = Σ[*X* - μ]2**

**Computational Formula: SS = Σ**X2 -

**SS = Σ**X2 - 102 – 64 = 38

***Variance =*  = *38 / 4 = 9.5***

***Notation =* 2 [lower case sigma squared) = 9.5**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Scores [*X*]** | Deviation  [*X* - μ] | Squared Deviation |
|  | 2 | -2 | 4 |
|  | 1 | -3 | 9 |
|  | 9 | 5 | 25 |
|  | 4 | 0 | 0 |
| Total | 16 | 0 | 38 |

**B. Calculating Variance for Sample**

Variance = mean of *squared* deviations [ ]

***Definitional formula*** [process using deviation scores]

***SS = Σ(X - M)2***

***Computational formula*** [uses actual scores]

***SS = ΣX2 –***

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Number of siblings [***X***] | Deviation ***X - M*** | Deviation Score |
| 1 | 1 | 1 – 1.5 | -0.5 |
| 2 | 1 | 1 – 1.5 | -0.5 |
| 3 | 0 | 0 – 1.5 | -1.5 |
| 4 | 3 | 3 – 1.5 | 1.5 |
| 5 | 0 | 0 – 1.5 | -1.5 |
| 6 | 1 | 1 – 1.5 | -0.5 |
| 7 | 2 | 2 – 1.5 | 0.5 |
| 8 | 5 | 5 – 1.5 | 3.5 |
| 9 | 3 | 3 – 1.5 | 1.5 |
| 10 | 1 | 1 – 1.5 | -0.5 |
| 11 | 0 | 0 – 1.5 | -1.5 |
| 12 | 1 | 1 – 1.5 | -0.5 |
| Total: 18 | |  |  |

Total = ***ΣX = 18***

**n = 12**

**M =**

Variance = = 25/11 = 2.273

* Variance = 1.508

**σ = s**

**68 – 95 -99.7% Rule (Emperical Rule)**

**N (18: Mean,1: Standard deviation) Using for below**

