Self Assignment: Variance and Standard Deviation

2024

1 Introduction

This document covers variance with a focus on both population and sample definitions, explains Bessel's correction, and introduces standard deviation.

2 Detailed Formulas and Symbol Definitions

2.1 Variance

Variance measures the spread of data points from the mean by calculating the average squared deviation of each data point from the mean.

2.1.1 Population Variance (σ^2)

For a population, the variance is calculated using:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N},$$

where:

- x_i: Each individual data point.
- μ: The population mean, computed as $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.
- N: Total number of data points in the population.

2.1.2 Sample Variance (s^2) and Bessel's Correction

For a sample drawn from a larger population, the variance is calculated as:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1},$$

where:

- x_i: Each individual data point in the sample.
- $\bar{x}$: The sample mean, computed as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- n: Number of data points in the sample.
- $n-1$: Bessel's correction factor, which is used instead of n to provide an unbiased estimate of the population variance.

2.2 Standard Deviation

Standard deviation is the square root of the variance and provides a measure of dispersion in the same units as the original data.

2.2.1 Population Standard Deviation (σ)

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}.$$

2.2.2 Sample Standard Deviation (s)

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}.$$

3 Examples

Example 1: Population Variance and Standard Deviation

Problem: Calculate the population variance and standard deviation for the dataset: $2, 4, 6, 8, 10$.

Step 1: Calculate the Mean

$$\mu = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6.$$

Step 2: Calculate the Squared Deviations from the Mean

$$(2-6)^2 = 16, \quad (4-6)^2 = 4, \quad (6-6)^2 = 0, \quad (8-6)^2 = 4, \quad (10-6)^2 = 16.$$

Sum of squared deviations:

$$\sum(x_i - \mu)^2 = 16 + 4 + 0 + 4 + 16 = 40.$$

Step 3: Compute the Population Variance

$$\sigma^2 = \frac{40}{5} = 8.$$

Step 4: Compute the Population Standard Deviation

$$\sigma = \sqrt{8} \approx 2.83.$$

Example 2: Sample Variance and Standard Deviation

Problem: Calculate the sample variance and sample standard deviation for the same dataset: $2, 4, 6, 8, 10$.

Step 1: Calculate the Sample Mean

$$\bar{x} = 6 \quad \text{(as computed above)}.$$

Step 2: Calculate the Squared Deviations from the Sample Mean

$$\sum(x_i - \bar{x})^2 = 40 \quad \text{(same as in Example 1)}.$$

Step 3: Compute the Sample Variance (Using Bessel's Correction)

$$s^2 = \frac{40}{5-1} = \frac{40}{4} = 10.$$

Step 4: Compute the Sample Standard Deviation

$$s = \sqrt{10} \approx 3.16.$$

4 Illustrations

4.1 Scatter Plot with Mean and Standard Deviation

The following scatter plot illustrates a dataset along with its mean and the range corresponding to one standard deviation. Consider the dataset:

$$\{(1,2),\ (2,3),\ (3,5),\ (4,7),\ (5,8)\}.$$

For this dataset, the mean of the y-values is:

$$\mu_y = \frac{2+3+5+7+8}{5} = 5,$$

and the standard deviation is approximately 2.28. In the plot, the dashed red line represents the mean, while the dotted green lines indicate one standard deviation above and below the mean.

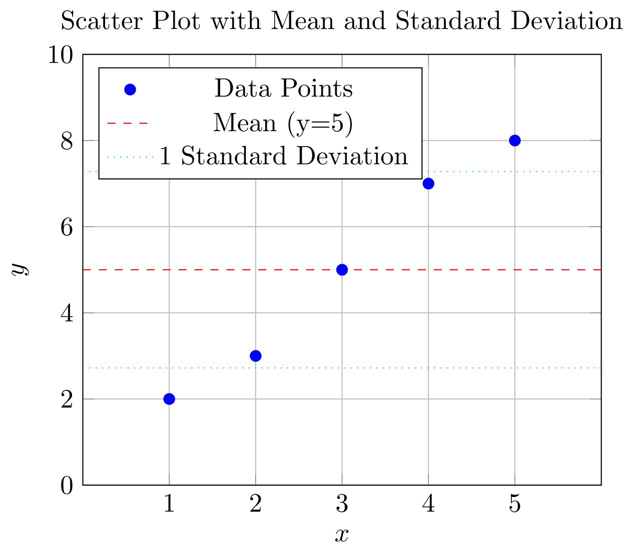

Figure 1: Scatter plot showing data points, the mean, and the range of one standard deviation.