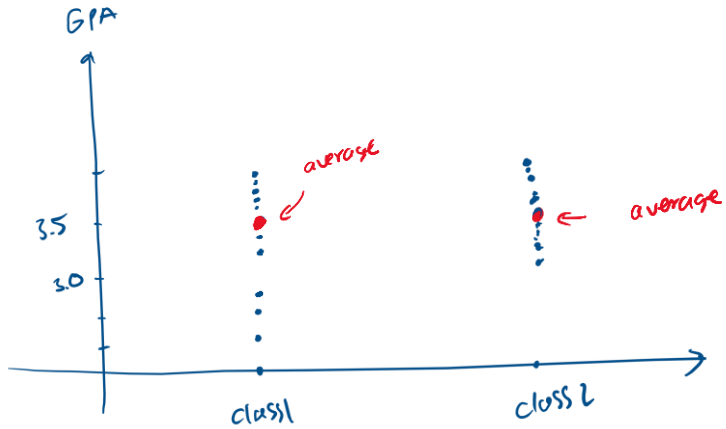
Measure of Variation:

This is to measure how the data values spread. If the measure of variation is large, then the data values more spread out.



We notice that the Class 1's GPA are more spread out than Class 2's GPA.

We will cover:

- Range
- Variance and Standard Deviation
- Chebyshev's Theorem
- The Empirical (Normal) Rule

---

1. Range:

The range of the data = maximum – minimum

Example:

Given the data: 2.0, 2.1, 4.0, 1.9, 3.6

Maximum = 4.0

Minimum = 1.9

Range = 4.0 – 1.9 = 2

---

2. Summation Notation

Example :

$$\sum_{i=1}^{100} i = 1 + 2 + 3 + 4 + \cdots + 100$$

$$\sum_{i=1}^{100} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + \cdots + 100^2$$

$$\sum_{i=1}^{17} x_i = x_1 + x_2 + x_3 + \cdots + x_{17}$$

$$\sum_{i=2}^{n} x_i \cdot y_i = x_2 y_2 + x_3 y_3 + \cdots + x_n y_n$$

Usually for a sample data:

$x_1, x_2, x_3, x_4 \cdots x_n$

$\sum_{i=1}^{n}$ can be written as $\sum$

For example: write the formular of the sample mean.

for the sample data $(x_1, x_2, \ldots, x_n)$

sample mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

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

Practice :

Write the follow using summation notation.

① $1^3 + 2^3 + 3^3 + \cdots + 56^3$

$$= \sum_{i=1}^{56} i^3$$

② $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2025}$

$$= \sum_{i=1}^{2025} \frac{1}{i}$$

③ $\frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots + \frac{1}{20^2}$

$$= \sum_{i=3}^{20} \frac{1}{i^2}$$

④ $2^{2} + 2^3 + 2^4 + 2^5 + \cdots + 2^{19}$

$$= \sum_{i=2}^{19} 2^i$$

⑤ $x_3^2 + x_4^2 + x_5^2 + \cdots + x_{100}^2$

$$= \sum_{i=3}^{100} x_i^2$$

3. Variance and Standard Deviation

* Sample Variance formula

Given the sample : $(x_1, x_2, \ldots, x_n)$

The sample variance is

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

where $\bar{x} = \frac{\sum x_i}{n}$, which is the sample mean.

* Sample Standard Deviation, S

$$S = \sqrt{\text{Sample variance}}$$

OR

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

Note : The variance can be denoted by V, Var or s^2

Other formulas for Variance and Standard Deviation.

$$V = s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} \quad ; \quad s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Example:

Find the sample variance and standard deviation for the amount of European auto sales for a sample of 6 years shown. The data are in millions of dollars.

11.2, 11.9, 12.0, 12.8, 13.4, 14.3

$n = 6$

$$\sum x = 11.2 + 11.9 + 12.0 + 12.8 + 13.4 + 14.3 = 75.6$$

$$\sum x^2 = 11.2^2 + 11.9^2 + 12.0^2 + 12.8^2 + 13.4^2 + 14.3^2 = 958.94$$

plus in the formula:

$$s^2 = \frac{6 * 958.94 - 75.6^2}{6(6-1)} = 1.276$$

$$s = \sqrt{1.276} = 1.13$$

Practice Problem:

The number of incidents in which police were needed for a sample of 10 schools in Allegheny County is 7, 37, 3, 8, 48, 11, 6, 0, 10, 3. Calculate the range, variance and standard deviation.

4. Coefficient of Variation

Data 1 : 70, 80, 90 [unit $1000]

Data 2 : 70,000 ; 80,000 ; 90,000 [unit $]

var (Data 1) = 100

var (Data 2) = 100,000,000

we notice that variance | sd is sensitive with the unit used measure in the data.

we would like to have a variation that is NOT sensitive to the unit.

$$\boxed{CV \text{ or } CVar = \frac{s}{\bar{x}} \cdot 100}$$

Let calculate CV for the 2 data sets above

For data 1 : (70, 80, 90)

$$s = \sqrt{v} = \sqrt{100} = 10$$

$$\bar{x} = \frac{70 + 80 + 90}{3} = 80$$

$$C = \frac{10}{80} \cdot 100 = 12.5$$

For data 2

$$S = \sqrt{100,000,000} = 10,000$$

$$\bar{x} = \frac{70,000 + 80,000 + 90,000}{3} = 80,000$$

$$\Rightarrow \quad CV(\text{data } 2) = \frac{10000}{80000} \cdot 100 = 12.5$$

we see that both data sets have the same CV.

Example

Sales of Automobiles

The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is \$5225, and the standard deviation is \$773. Compare the variations of the two.

Data 1 : $\bar{x} = 87, \quad s = 5$

$$\Rightarrow \quad CV = \frac{s}{\bar{x}} \cdot 100 = \frac{5}{87} \cdot 100 \approx 5.75 \ (\%)$$

Data 2 : $\bar{x} = 5225, \quad s = 773$

$$\Rightarrow \quad CV = \frac{773}{5225} \cdot 100 = 14.79 \ (\%)$$

⇒ The variation in data 2 is higher.

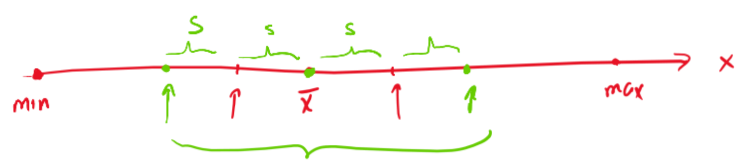
OR Data set 2 is more variable than data set 1.

Practice Problem

(Coefficient of Variation) The mean for the number of pages (variable 1) of a sample of women's fitness magazines is 132, with a variance of 23; the mean for the number of advertisements (variable 2) of a sample of women's fitness magazines is 182, with a variance of 62. Which variable is more variable?

5. Chebyshev's Theorem

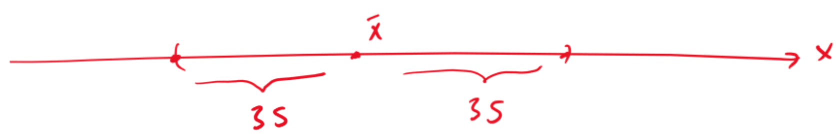
Chebyshev's theorem The proportion of values from a data set that will fall within k standard deviations of the mean will be at least $1 - 1/k^2$, where k is a number greater than 1 (k is not necessarily an integer).



$k=2$: within 2s from the mean, we would cover at least

$1 - \frac{1}{2^2} = 75\%$ of the data.

$k=3$:



within 3s from the mean, we would cover at least

$1 - \frac{1}{3^2} = 88.89\%$ of the data.

Example:

Prices of Homes

The mean price of houses in a certain neighborhood is $50,000, and the standard deviation is $10,000. Find the price range for which at least 75% of the houses will sell.

Practice Problem

1. *(Chebyshev's theorem) The average number of calories in a regularsize bagel is 240. If the standard deviation is 38 calories, find the range in which at least 75% of the data will lie.*

2. *(Chebyshev's theorem) Americans spend an average of 3 hours per day online. If the standard deviation is 32 minutes, find the range in which at least 88.89% of the data will lie.*

Example:

Travel Allowances

A survey of local companies found that the mean amount of travel allowance for executives was $0.25 per mile. The standard deviation was $0.02. Using Chebyshev's theorem, find the minimum percentage of the data values that will fall between $0.20 and $0.30.

Practice Problem

(Chebyshev's theorem) The average of the number of trials it took a sample of mice to learn to traverse a maze was 12. The standard deviation was 3. Using Chebyshev's theorem, find the minimum percentage of data values that will fall in the range of 4–20 trials.