Pharmaceutical Statistics

Lecture 5

Descriptive statistics

Measures of Dispersion

Prepared and Presented by

Dr. Muna Oqal

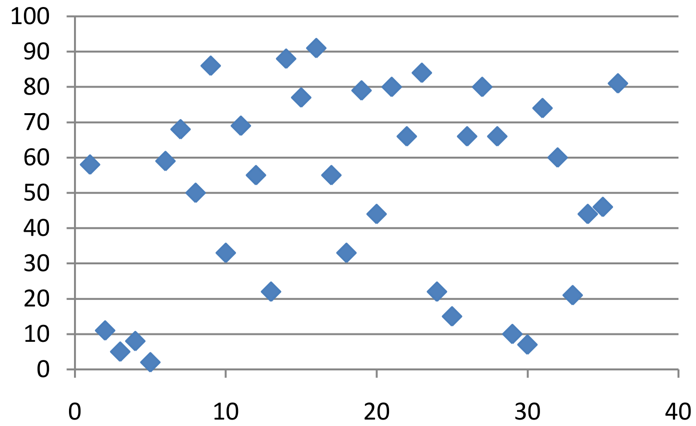
Measures of Dispersion

- Measures of dispersion characterise how spread out the distribution is, i.e., how variable the data are. It conveys information about the amount of variability present in a set of data.

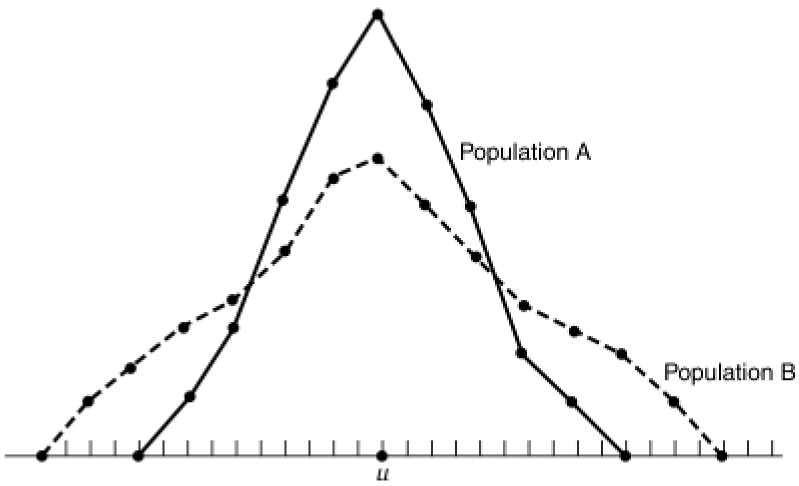
- **Note: If all the values are the same, there is no dispersion; if they are not all the same, dispersion is present in the data.** If the values **close** to each other the amount of Dispersion **small**. If the values are **widely scattered**, the Dispersion is **greater**.

- Other terms used synonymously with dispersion include variation, spread, and scatter.

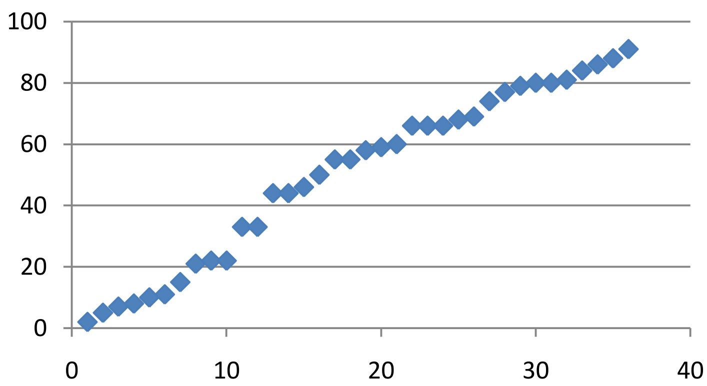
More variability

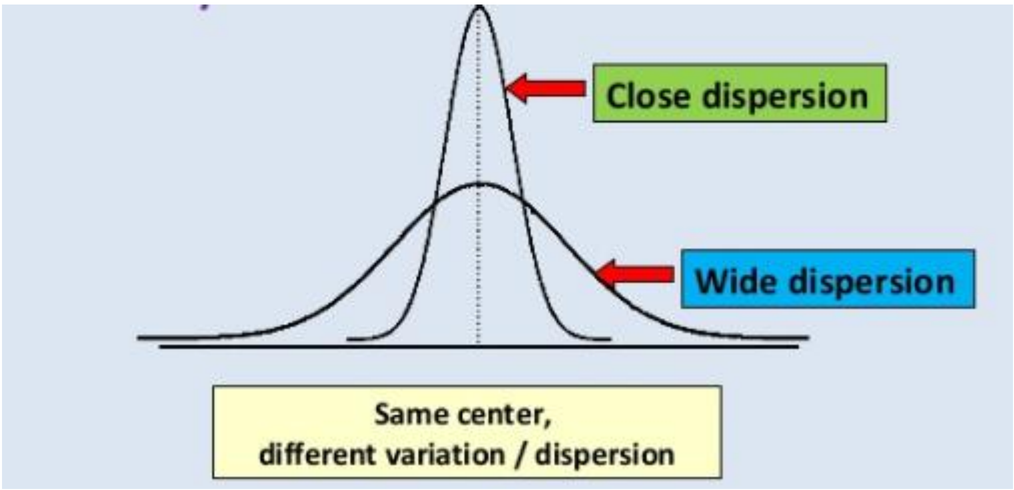


Population B, which is more variable than population A.



Less variability





Two frequency distributions with equal means but different amounts of dispersion.

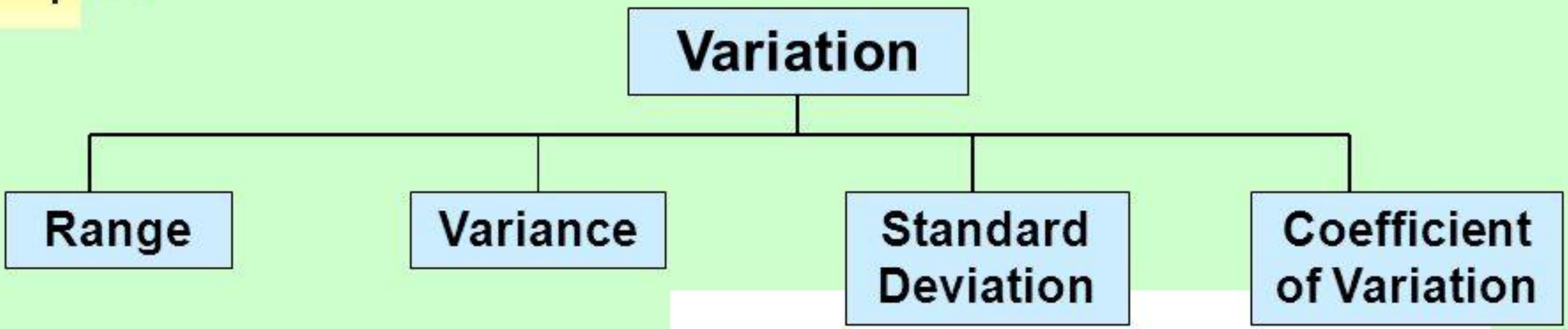
Indicators of dispersion

- Variation in experiments or in any assay is due to several reasons as

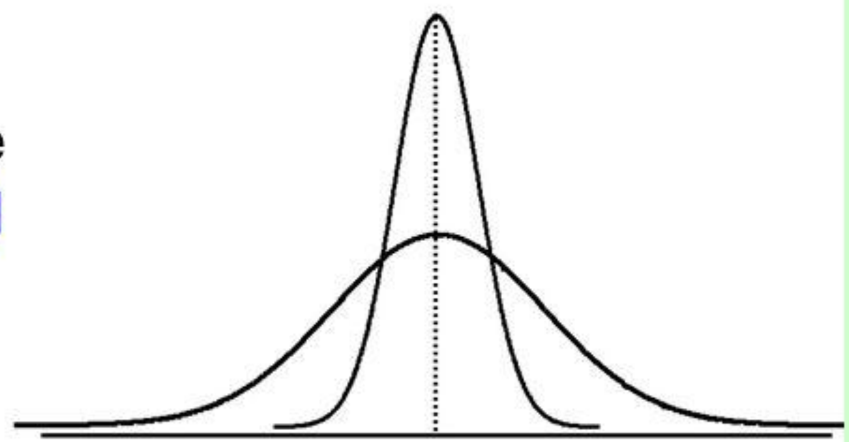
- The instrument used for analysis.
- The analyst performing the assay.
- The particular sample chosen.
- Unidentified error commonly known as noise.

Variation is an inherent characteristic of experimental observations.

Measures of Dispersion



- Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.



Same centre, different variation

1.The Range (R):

- The **R** is the difference between the largest and smallest value in a set of observations.
- Range = (X)Largest value- Smallest value

- **Properties of the R**
- Simplicity of its computation
- Poor measure of dispersion, as its value is concerned only onto two values, so imparts minimal information about a data set and therefore is of limited use.
- Data:
- 43,66,61,64,65,38,59,57,57,50.
- Find Range?
- Range=66-38=28

2.The Variance:

- It measure dispersion relative to the scatter of the values about the mean.

a) Sample Variance: S^2

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

$\bar{x}$ = is sample mean

n = count of values for the sample

b) Population Variance (σ)^2:

- $$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

- where Mu(μ) is Population mean

2. The Variance:

EXAMPLE:

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

- **Find Sample Variance of ages (**5,7,8,12,6,9,11,6).

Solution:

$\overline{x}$ **= (5+7+8+12+6+9+11+6)/8 = 8**

- $S^2 = [(5-8)^2 + (7-8)^2 + (8-8)^2 + (12-8)^2 + (6-8)^2 + (11-8)^2 + (6-8)^2]$ / 7
- S^2 = 9+1+0+16+4+1+9+4 / 7
- = 44/7
- = 6.28

3. The Standard Deviation:

- **The variance represents squared units and, therefore, is not an appropriate measure of dispersion when we wish to express this concept in terms of the original units.**
- **To obtain a measure of dispersion in original units, we merely take: the square root of the variance.**

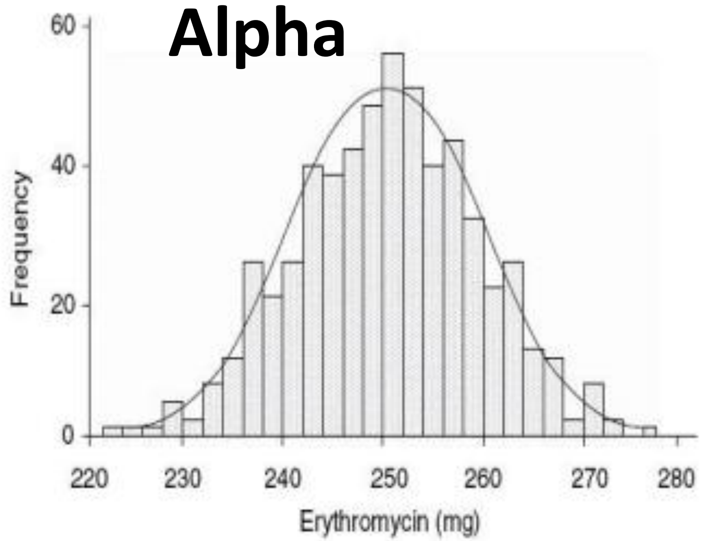
the square root of variance= $\sqrt{Variance}$

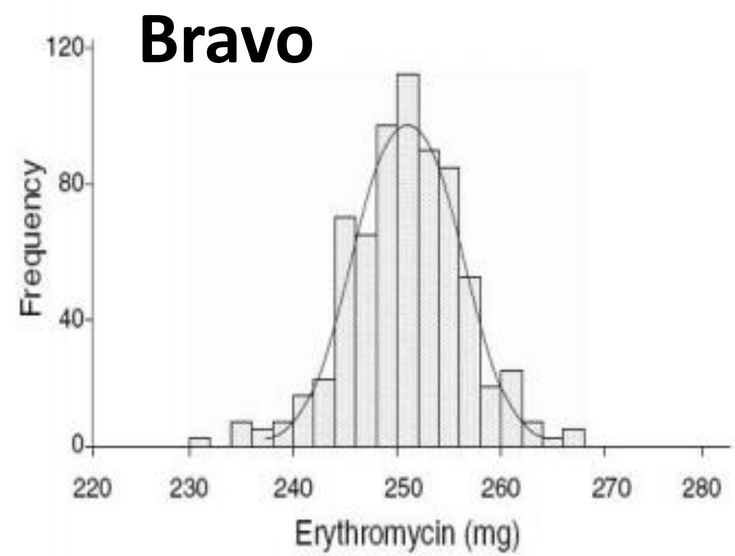
A. Sample Standard Deviation = S = $\sqrt{S^2}$

A. Population Standard Deviation = σ = $\sqrt{\sigma^2}$

3. The Standard Deviation

- Example:We have two tableting machines ('Alpha' and 'Bravo') producing erythromycin tablets with a nominal content of 250 mg.
- Five hundred tablets are randomly selected from each machine and their erythromycin contents assayed.





Histogram of erythromycin of 500 mg tablets from both Alfa and Bravo tableting machines; respectively.

The Standard Deviation

- An ‘indicator of dispersion’ is required in order to convey this difference in variability, which is **standard deviation**.

Alpha machine			Bravo machine		
Erythro content (mg)	Deviation from mean	Deviation squared	Erythro content (mg)	Deviation from mean	Deviation squared
249	**0.3**	0.09	251	**−0.1**	0.01
242	**−6.7**	44.89	247	**−4.1**	16.81
252	**3.3**	10.89	257	**5.9**	34.81
235	**−13.7**	187.69	250	**−1.1**	1.21
257	**8.3**	68.89	254	**2.9**	8.41
244	**−4.7**	22.09	251	**−0.1**	0.01
264	**15.3**	234.09	252	**0.9**	0.81
249	**0.3**	0.09	255	**3.9**	15.21
255	**6.3**	39.69	244	**−7.1**	50.41
240	**−8.7**	75.69	250	**−1.1**	1.21
Mean 248.7		Total 684.1	Mean 251.1		Total 128.9

Sum of squared deviations = 684.1
684.1/9 = 76.01
SD = square root 76.01
= **8.72 mg (SD)**

Sum of squared deviations = 128.9
128.9/9 = 14.32
SD = square root 14.32
= **3.78 mg (SD)**

3. The Standard Deviation

- The Alpha machine produces rather variable tablets and so several of the tablets deviate considerably (e.g. -13.7 or +15.3 mg) from the overall mean. So producing a **high** final SD (8.72 mg).

- In contrast, the Bravo machine is more consistent and individual tablets never have a drug content much above or below the overall average.

- The small figures in the column of individual deviations, leading to a **lower** SD (3.78 mg).

Reporting Standard deviation

Reporting the SD:

- The ± symbol is used in reporting the SD
- The symbol ± reasonably interpreted as meaning ‘more or less’.
- ± is used to indicate variability.
- With the tablets from our two machines, we would report their drug contents as:

- **Alpha machine: 248.7±8.72 mg (SD)**
- **Bravo machine: 251.1±3.78 mg (SD)**

Units of SD:

- The SD is not a unitless number. It has the same units as the individual pieces of data.
- Since our data consisted of erythromycin contents measured in milligrams, the SD is also in milligrams.

4. Coefficient of Variation (C.V)

- The standard deviation is useful as a measure of variation within a given set of data.

- When one desires to compare the dispersion in two sets of data, however, comparing the two standard deviations may lead to fallacious results.

- In this case is a measure of relative variation rather than absolute variation is required. Such a measure is found in the **coefficient of variation**, which expresses the standard deviation as a percentage of the mean.

Coefficient of variation expresses variation relative to the magnitude of data

4.The Coefficient of Variation (C.V):

- Is a measure use to compare the dispersion in two sets of data which is independent of the unit of the measurement .

$$C.V = \frac{S}{\overline{X}}(100)$$

- S: Sample standard deviation.
- $\overline{X}$: Sample mean.

- Suppose two samples of human males yield the following data:

	Sampe1	Sample2
Age	25-year-olds	11year-olds
Mean weight	145 pound	80 pound
Standard deviation	10 pound	10 pound

Question: examine the sample that has more variability?

A comparison of the standard deviations might lead one to conclude that the two samples possess equal variability.

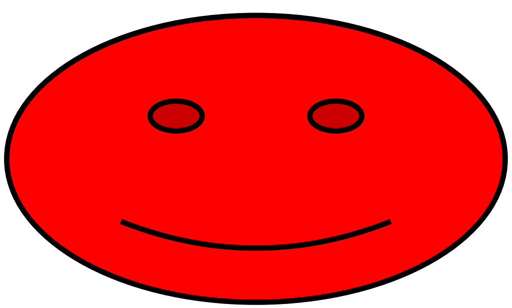
Solution:

- C.V (Sample1)= (10/145)*100= 6.9
- C.V (Sample2)= (10/80)*100= 12.5

Then, it is clear that **Sample2** (age of 11-years old) has more variation than **Sample1** (25-year-olds).

In other words, variation is much higher in the sample2 than sample1.

Now you got quite a different impression.



Coefficient of Variation

- Since the coefficient of variation is independent of the scale of measurement so:

➢ It is a useful statistic for comparing the variability of two or more variables (results) measured on different scales. for example, use the coefficient of variation to compare the variability in weights of one sample of subjects whose weights are expressed in pounds with the variability in weights of another sample of subjects whose weights are expressed in kilograms

- It is useful in comparing the results obtained by different persons who are conducting investigations involving the same variable.
- Because the coefficient of variation is a ratio, it is unitless (unlike the SD).