Pharmaceutical statistics

2025-2024

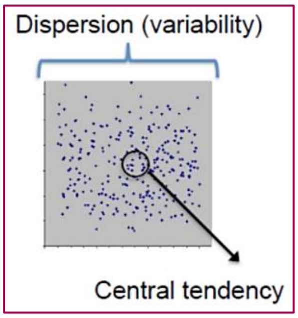
Dr. Rola Aqeel

ARKAN ACADEMY

Indicators of dispersion

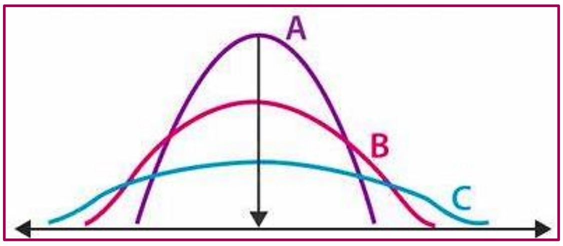
- If all observations are <u>the same</u>, there is *no variability*
- If they are <u>not all the same</u>, then dispersion is present in the data.

- ***Variation*** is an inherent characteristic of natural and experimental observations due to several reasons.
- It is always important to get an estimate of:
  (how much given objects tend to *differ from that central tendency*)



- In any experiment, variation will ***depend on***:
  - The <u>instrument</u> used for analysis.
  - The <u>analyst</u> performing the assay.
  - The sample chosen.
  - Unidentified error commonly known as.

- ***Why do we need indicators of dispersion?***
  - A, B and C have the <u>same mean</u> (central tendency indicator).
  - Based on similarity of the mean, can we say the data sets are the same? No
  - What is the difference between these data sets?
    Wideness of distribution (*Variability*).
  - How can we describe the (differences)?
    Using *indicators of dispersion*.



- ***Indicators of dispersion:***
  - *Standard deviation*
  - *Coefficient of variation*
  - *Variance*
  - *Quartiles*
  - *Box and whisker plot*
  - *Range (we don't use it because it is highly affected by extreme values/ outliers)*

Standard deviation

- Lets start with an example:
  - We have batch of drug was compressed in <u>2 types of machines</u> ***Alph*** and ***Bravo*** with a nominal content of 250mg then we took a <u>random sample 500 tablets</u> from each then we measured their drug content assays then we draw curve between (*drug content* and *frequency*)

  - The drug content distribution plots for Alpha & Bravo are <u>normal distribution</u>.
  - What about the ***central tendency*?** Both have the same central tendency values *almost the same mean*

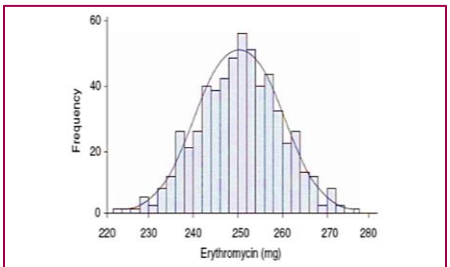


Figure 2.5 Histogram of erythromycin content of 500 tablets from an Alpha tabletting machine (relatively variable product – large standard deviation)

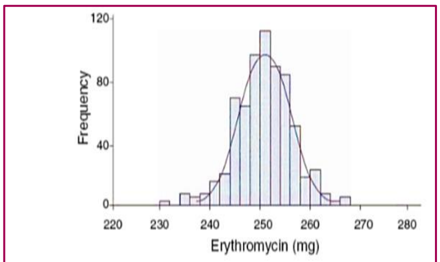


Figure 2.6 Histogram of erythromycin content of 500 tablets from a Bravo machine (more constant product – smaller standard deviation)

- Although tablets from both machines had equal mean, do you think the two machine still differ? How?
- What about ***dispersion***, is it the same or different? *Alpha* has a wider range of tablet weights this means *higher variability.*
- Which of these 2 curves have fatter tails? *Alpha,* Fatter tail means tails with *higher frequency* of the extreme values.
- Which of them has shorter middle value? *Alpha*

- ***So, we conclude that:***
  - Alpha tablet wights distribution has a wider range and distribution, higher variability, fatter tails, shorter middle value.
  - The two machines are very similar in terms of average drug content for the tablets, both producing tablets with a mean very close to 250 mg. However, the two products clearly *differ*.
  - With the *Alpha* machine, there is a considerable proportion of tablets with a content differing by more than 20 mg from the nominal dose (i.e. below 230 mg or above 270 mg).
  - Whereas with the *Bravo* machine, such outliers are much rarer.
  - An indicator of dispersion is required to convey this difference in variability and to decide which one has better performance!!

- ***Standard deviation formula:***
  - SD = $\sqrt{\frac{\sum_{i=1}^{n}(xi-\bar{x})^2}{n-1}}$
  - This is the Standard deviation of a sample.
  - For a population it is donated as σ
  - $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X-\mu)^2}{N}}$
  - What is the unit of ***SD*** or σ? The same unit of the observations.
  - ***Standard deviation*** is a widely used measure of variability and central dispersion

- Now let's go back to our *example* about the tableting machines:
  - We calculated the ***SD*** of the content of 10 tablets from an ***Alpha*** and ***Bravo*** tableting machines.
  - The ***Alpha*** machine produces rather variable tablets and so several of the tablets deviate considerably from the overall mean.
  - These relatively large figures then feed through the rest of the calculation, 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.
    (more appropriate to place in the R & D section and more expensive).
  - The small figures in the column of individual deviations, leading to a ***lower SD*** **(3.78 mg)**

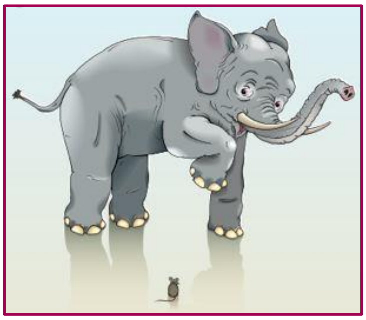
	Alpha machine			Bravo machine		
	Erythro content (mg)	Deviation from mean $X_i - \bar{X}$	$(X_i - \bar{X})^2$ Deviation squared	Erythro content (mg)	Deviation from mean	Deviation squared
	249	**0.3**	0.09	251	**−0.1**	0.01
	242	**−6.7**	44.89	247	**−4.1**	16.81
	252	**3.3**	10.89	257	**5.9**	34.81
	235	**−13.7**	187.69	250	**−1.1**	1.21
XX$_i$	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
$\bar{X}\bar{X}$=	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)

- ***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 (Mean ± SD mg)
    - ✓ ***Bravo*** *machine*: ***251.1±3.78*** mg (Mean ± SD mg)
  - The two machines produce tablets with almost identical mean contents, but those from the ***Alpha*** machine are *two* to *three* times more variable.

- ***Standard deviation and Coefficient of variation***



  - Elephant tail =150±**10** cm
  - Mouse tail=7±**3** cm
    - ✓ Which is more variable: the elephant tail length results or the one for the mouse? *elephant tail length*
  - **CV=SD/Mean*100**
    - ✓ Elephant tail: CV=10/150*100=**6.7%**
    - ✓ Mouse tail: CV= 3/7*100=**42.8%** *(more variable)*
  - Coefficient of variation (CV) expresses variation relative to the magnitude of data.
  - Useful to compare variation in two or more sets of data with different mean values
  - ***CV has no unit*** (it is a ratio).
  - A *higher* CV indicates greater variability in the data relative to the mean.
  - A *lower* CV means less variability, suggesting that the data points are more closely clustered around the mean.

  ★ Example 1:
  - Set A: 10, 20, 30, 40, 50 mean = 30
  - Set B: 1, 2, 3, 4, 5 mean = 3
  - To know which set is more variable we calculate CV because the two sets have different means.

  ★ Example 2:
    - ✓ 5 students got marks out of 100
      - 50, 60, 70, 80, 90
      - Mean = 70 & SD = 10
      - CV = 14.2%
    - ✓ 5 students got marks out of 10
      - 5, 6, 7, 8, 9
      - Mean = 7 & SD = 1
      - CV = 14.2%
    - ✓ The have the same variability

- ***Variance***
  - The Variance: σ^2 (population) or S^2 (sample) is a measure of spread that is related to the deviations of the data values from their mean.
  - *For population:* $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$
  - *For sample:* $S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
  - Unit: same as mean but squared (If mean in mg, variance will be in mg^2)

ARKAN

ACADEMY

علم في كل مكان

Arkan academy

Arkanacademy

www.arkan-academy.com

+962 790408805