

3.2 Measures of Variation

Variation = how much does the data vary

-Measure of Variation

1. Range : is the difference between the maximum and minimum value.

Ex: 23 12 45 65 76

76 is the maximum

12 is the minimum

so $76 - 12 = 64$

We had a range of 76 minus 12 = 64

Variance and Standard Deviation*

variance:

-How much the data vary

-think: how well does the mean represent the spread of the data

Standard deviation:

-'Standard' - following a standard, same

-'Deviation' - like a deviated septum

Sample Defining Formulas

$$\text{Sample variance} = s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$\text{Sample standard deviation} = s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Sum of square =

$$\sum(x - \bar{x})^2$$

Sample and Population

Sample Defining Formulas

$$\text{Sample variance} = s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$\text{Sample standard deviation} = s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Population Defining Formulas

$$\text{Population variance} = \sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

$$\text{Population standard deviation} = \sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

Coefficient of Variation

Coefficient of Variation

- You will notice the word “coefficient” used a lot in statistics
- Coefficient of Variation is CV for short
- CV shows how much the data varies compared to the mean
- Always expressed in a %

$$\text{Sample CV} = \frac{s}{x} \times 100$$

$$\text{Population CV} = \frac{\sigma}{\mu} \times 100$$

Patients:
 $s = 3.74$, $x\bar{} = 6$

$$\frac{3.74}{6} \times 100 = 62\%$$

CV is used for Comparison

- It's hard to explain with only one group of patients.
- CV has no units – so you could compare two different ways of measuring a lab value, for example.
- The CV is the measure of the spread of the data relative to the average of the data.
 - In the first sample, the s is only 50% of the mean.
 - In the second sample, the s is 62% of the mean.

Other Patients:
 $s = 4$, $\bar{x} = 8$

$$\frac{4}{8} \times 100 = 50\%$$

Patients:
 $s = 3.74$, $\bar{x} = 6$

$$\frac{3.74}{6} \times 100 = 62\%$$

CV, It indicates data consistency:

- a lower CV = higher precision/consistency,
- a higher CV = greater [1]variability.

Chebyshev's Theorem

What Chebyshev Figured Out

- First, he started thinking like this:
 - If you have an \bar{x} and an s , you can create lower and upper limits by subtracting the s and adding the s to the \bar{x} .
 - You can do this with a μ and σ , too – population version
- For example, if I had an μ of 100, and an σ of 5:
 - If I subtracted 1 σ from 100, I'd get 95 as the lower limit
 - If I add 1 σ to 100, I get 105 as the upper limit.
 - I could even try this by doing 2 σ – meaning subtracting 10 for the lower limit and adding 10 for the upper limit
- He realized if he used some rules along with this, there would be an interpretation of these limits that would be useful.

Chebyshev's Theorem

- He figured out the upper and lower limits, when figured out this way, explained *at least* what % of the data would be between these limits in his dataset
- He wanted this to work for all distributions, not just normal
- He used a formula to figure out this % that was based on the s. He used “k” to mean the number of “s”’s (or “ σ ”’s) in the equation:

$$1 - \frac{1}{k^2} = \% \text{ of data between } \bar{x} \text{ minus } k \text{ and } \bar{x} \text{ plus } k$$

For example, if a factory produces bolts with a mean length of 5cm and a standard deviation of 0.1cm, Chebyshev's Theorem guarantees that at least 75% of bolts will be between 4.8 cm and 5.2 cm (within 2 standard deviations).

Let's Try Chebyshev's Intervals!

- Calculate limits for Patient Sample:
 - 75% limits: $6 +/-(2 * 3.74) = -1.48 \text{ to } 13.48$
 - 88.9% limits: $6 +/-(3 * 3.74) = -5.22 \text{ to } 17.22$
 - 93.8% limits: $6 +/-(4 * 3.74) = -8.96 \text{ to } 20.96$
- Interpretation
 - *At least* 75% of the data are between -1.48 and 13.48
 - *At least* 88.9% of the data are between -5.22 to 17.22
 - *At least* 93.8% of the data are between -8.96 and 20.96.

Patient Sample (waiting room minutes)
 $s = 3.74, \bar{x} = 6$

s or 6	% of data in interval
2	75%
3	88.9%
4	93.8%

$$\bar{x} +/-(n * S)$$

1. lack of consistency or fixed pattern; liability to vary or change ↵