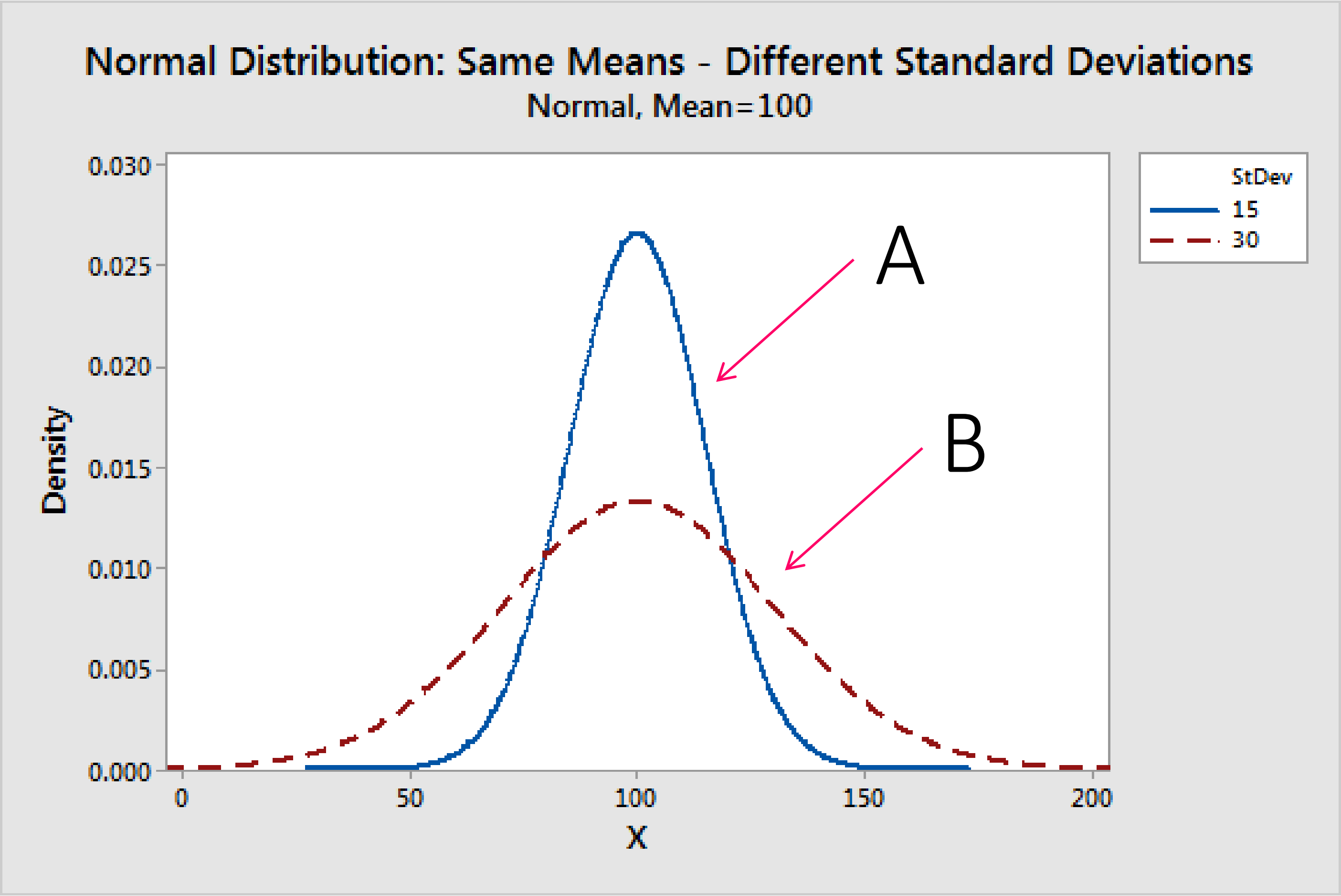
# Chapter 4 – Unit 3
# Measures of Dispersion

# Introduction

- We learnt that the mean, median and mode are measures of central tendency of a distribution.
- Two distributions may have same central locations but different dispersions.
- A distribution can be effectively described by measures of central tendency and measures of dispersion.
- The dispersion of a distribution provides additional information on the reliability of the measures of central tendency.
    - If the data are widely dispersed, the central location is said to be less representative of the data as a whole.
    - The central location for data with little dispersion is considered more reliable.

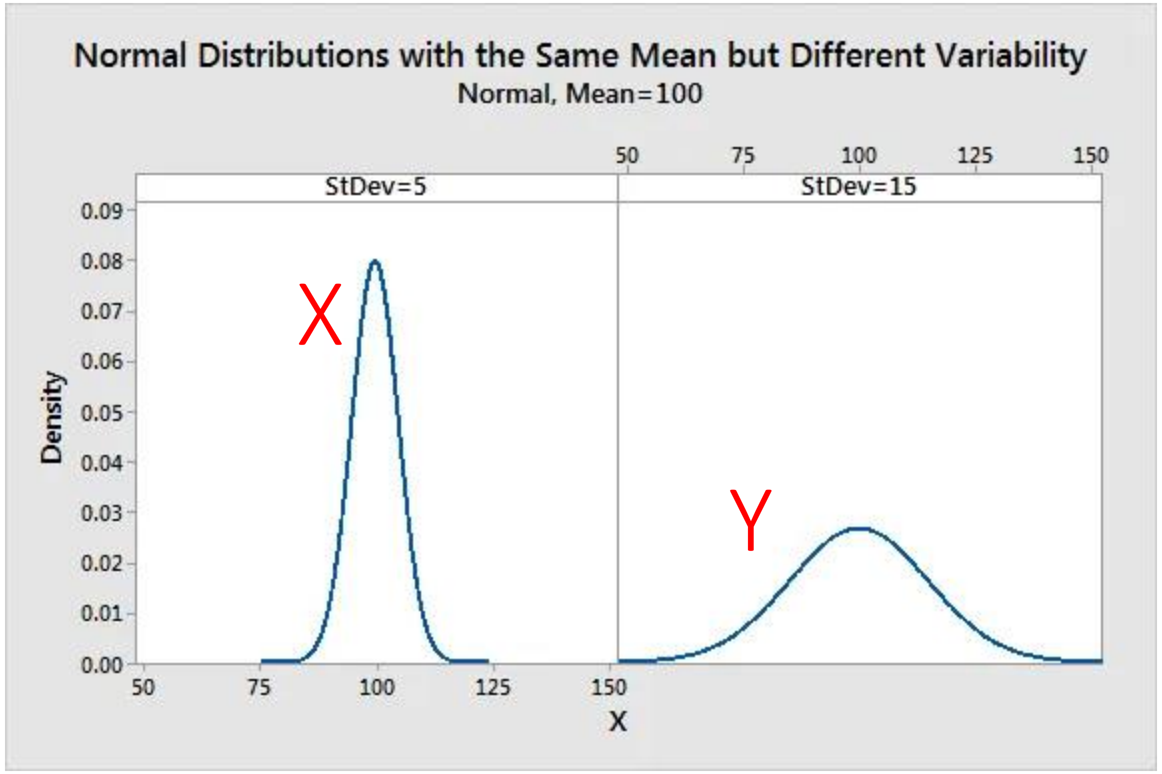
# Recall

| Measures of central tendency | Measures of Dispersion |
| --- | --- |
| Mean, Median, Mode, $\bar{x}$ $\tilde{x}$ $\hat{x}$ | Range<br>Variance<br>Standard deviation<br>IQR |




- Both distribution of A and B have the same mean = 100 but different standard deviation of 15 and 30 respectively.
- Therefore, distribution A is said to be more representative to the data as a whole compared to distribution B.

# Which distribution is well representative?




Answer:

Distribution X because (mean=100, sd=5) compared to Y (mean=100, sd=15)

# Range

# Range for ungrouped data

Range = Largest value – Smallest value

Example 5.1

Find the range for the following data.

| | | | | | | |
|---|---|---|---|---|---|---|
| 9.77 | 11.35 | 12.46 | 13.80 | 15.47 | 17.48 | 18.37 |
| 18.47 | 18.61 | 20.72 | 21.49 | 22.47 | 31.50 | 38.16 |

Range = 38.16 – 9.77 = 28.39

# Range for grouped data

Range = Upper boundary of highest class – Lower boundary of lowest class

Example 5.2

Table shows the daily wages of 80 workers in a factory. Determine the range.

| Daily wages (RM) | Class boundary | Number of workers |
|---|---|---|
| 10 – 19 | 9.5 – 19.5 | 6 |
| 20 – 29 | 19.5 – 29.5 | 10 |
| 30 – 39 | 29.5 – 39.5 | 30 |
| 40 – 49 | 39.5 – 49.5 | 20 |
| 50 – 59 | 49.5 – 59.5 | 10 |
| 60 – 69 | 59.5 – 69.5 | 4 |

Range = 69.5 – 9.5 = 60

# Interquartile Range

# Interquartile Range

= Upper quartile – Lower quartile

$= Q_3 - Q_1$

# Quartile Deviation

# Quartile Deviation

Also known as Semi-Interquartile Range

$$= \frac{1}{2}(Q_3 - Q_1)$$

# Sample variance
# &
# Standard deviation

# Sample variance and standard deviation for ungrouped data

- Sample variance for ungrouped data

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]$$

- Sample standard deviation for ungrouped data

$$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]}$$

# Basic Info

Standard deviation is the square root of the variance.

$$s = \sqrt{s^2}$$

$$4 = \sqrt{16}$$

## Example 5.3

Given variance, $s^2 = 16$, Find standard deviation, s

$$s = \sqrt{s^2}$$

$$s = \sqrt{16}$$

$$s = 4$$

## Example 5.4

Given standard deviation, s = 9, Find variance, $s^2$

$$s^2 = 9^2$$

$$s^2 = 81$$

# Example 5.5

Find the variance and the standard deviation of the sample data below.

1, 7, 2, 5

Answer:

$$\sum x = 1 + 7 + 2 + 5 = 15$$

$$\sum x^2 = 1^2 + 7^2 + 2^2 + 5^2 = 79$$

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]$$

Variance,

$$s^2 = \frac{1}{4-1}\left[79 - \frac{(15)^2}{4}\right]$$

$$s^2 = 7.583$$

# Cont..

Answer:

Standard deviation,

$$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]}$$

$$s = \sqrt{\frac{1}{4-1}\left[79 - \frac{(15)^2}{4}\right]}$$

$$s = 2.754$$

or it just simply,

$$s = \sqrt{s^2}$$

$$s = \sqrt{7.583}$$

$$s = 2.754$$

# Sample variance and standard deviation of grouped data

- Variance, $s^2$

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

*or*

$$s^2 = \frac{1}{n-1}\left(\sum fx^2 - n\bar{x}^2\right)$$

*or*

$$s^2 = \frac{1}{n-1}\left(\sum fx^2 - \frac{(\sum fx)^2}{n}\right)$$

**Commonly used**

- Standard deviation, s

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

*or*

$$s = \sqrt{\frac{1}{n-1}\left(\sum fx^2 - n\bar{x}^2\right)}$$

*or*

$$s = \sqrt{\frac{1}{n-1}\left(\sum fx^2 - \frac{(\sum fx)^2}{n}\right)}$$

**Commonly used**

# Example 5.6

The age distribution for a sample of employees in SP Company is shown below

| Age (years) | Number of workers |
|---|---|
| 21 – 25 | 10 |
| 26 – 30 | 35 |
| 31 – 35 | 16 |
| 36 – 40 | 14 |
| 41 – 45 | 13 |
| 46 – 50 | 10 |
| 51 – 55 | 3 |
| | 100 |

Calculate the variance and the standard deviation.

# Steps 1: Find the midpoint (x)

$$= \frac{21 + 25}{2}$$
$$= 23$$

| Age (years) | Number of workers (f) | Mid-points (x) | | |
|---|---|---|---|---|
| 21 – 25 | 10 | 23 | | |
| 26 – 30 | 35 | 28 | | |
| 31 – 35 | 16 | 33 | | |
| 36 – 40 | 14 | 38 | | |
| 41 – 45 | 12 | 43 | | |
| 46 – 50 | 10 | 48 | | |
| 51 – 55 | 3 | 53 | | |
| | 100 | | | |

# Steps 2: Find the $x^2$

$23^2 = 529$

| Age (years) | Number of workers (f) | Mid-points (x) | $x^2$ | |
|---|---|---|---|---|
| 21 – 25 | 10 | 23 | 529 | |
| 26 – 30 | 35 | 28 | 784 | |
| 31 – 35 | 16 | 33 | 1089 | |
| 36 – 40 | 14 | 38 | 1444 | |
| 41 – 45 | 12 | 43 | 1849 | |
| 46 – 50 | 10 | 48 | 2304 | |
| 51 – 55 | 3 | 53 | 2809 | |
| | 100 | | | |

# Steps 3: Find the f.x

10 x 23 = 230

| Age (years) | Number of workers (f) | Mid-points (x) | $x^2$ | f.x |
|---|---|---|---|---|
| 21 – 25 | 10 | 23 | 529 | 230 |
| 26 – 30 | 35 | 28 | 784 | 980 |
| 31 – 35 | 16 | 33 | 1089 | 528 |
| 36 – 40 | 14 | 38 | 1444 | 532 |
| 41 – 45 | 12 | 43 | 1849 | 516 |
| 46 – 50 | 10 | 48 | 2304 | 480 |
| 51 – 55 | 3 | 53 | 2809 | 159 |
| | 100 | | | |

# Steps 4: Find the f.$x^2$

10 x 529 = 5290

| Age (years) | Number of workers (f) | Midpoint (x) | $x^2$ | f.x | f. $x^2$ |
|---|---|---|---|---|---|
| 21 – 25 | 10 | 23 | 529 | 230 | 5290 |
| 26 – 30 | 35 | 28 | 784 | 980 | 27440 |
| 31 – 35 | 16 | 33 | 1089 | 528 | 17424 |
| 36 – 40 | 14 | 38 | 1444 | 532 | 20216 |
| 41 – 45 | 13 | 43 | 1849 | 516 | 22188 |
| 46 – 50 | 10 | 48 | 2304 | 480 | 23040 |
| 51 – 55 | 3 | 53 | 2809 | 159 | 8427 |
| | | | | | |

# Steps 5: Sum & Apply the formula

| Age (years) | Number of workers (f) | Midpoint (x) | $x^2$ | f.x | f. $x^2$ |
|---|---|---|---|---|---|
| 21 – 25 | 10 | 23 | 529 | 230 | 5290 |
| 26 – 30 | 35 | 28 | 784 | 980 | 27440 |
| 31 – 35 | 16 | 33 | 1089 | 528 | 17424 |
| 36 – 40 | 14 | 38 | 1444 | 532 | 20216 |
| 41 – 45 | 13 | 43 | 1849 | 516 | 22188 |
| 46 – 50 | 10 | 48 | 2304 | 480 | 23040 |
| 51 – 55 | 3 | 53 | 2809 | 159 | 8427 |
| | 100 | | | 3425 | 124025 |

$$s^2 = \frac{1}{n-1}\left(\sum fx^2 - \frac{(\sum fx)^2}{n}\right)$$

$$= \frac{1}{100-1}\left(124025 - \frac{(3425)^2}{100}\right)$$

$$= 67.87$$

Interpretation = The sample variance is 67.87 workers

Apply the formula for standard deviation

$$s = \sqrt{\frac{1}{n-1}\left(\sum fx^2 - \frac{(\sum fx)^2}{n}\right)}$$

$$s = \sqrt{\frac{1}{100-1}\left(124025 - \frac{(3425)^2}{100}\right)}$$

$$s = \sqrt{67.87}$$

$$s = 8.24$$

or

$$s = \sqrt{s^2} = \sqrt{67.87}$$

$$= 8.24$$

Interpretation = The sample standard deviation is 8.24 workers

# Coefficient of Variation (CV)

# Coefficient of Variation/ Relative Dispersion

- Large standard deviations mean large variability within the data set.
- In some cases, large variability is desired, but in other cases small variability is preferred.
- CV is a useful measure when comparing distributions of different means and variances

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Where

s = sample standard deviation

$\bar{x}$ = sample mean

# Example 5.7

- Typist Ani can type 40 words per minute with standard deviation of 5 while typist Jura can type 160 words per minute with standard deviation of 10. which typist is more consistent in her work?

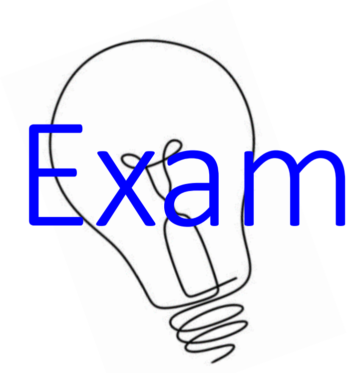
Keyword for CV

- CV for Ani $= \frac{5}{40} \times 100 = 12.5\%$

- CV for Jura $= \frac{10}{160} \times 100 = 6.25\%$

Thus, the typing ability of typist Jura is more consistent than typist Ani

# Exam tips



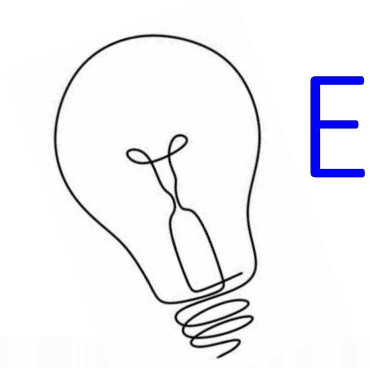
Keywords:

**More consistent** – choose lowest CV

**Less consistent** – choose highest CV

**More disperse** – choose highest CV

**Less disperse** – choose lowest CV



# Exam tips



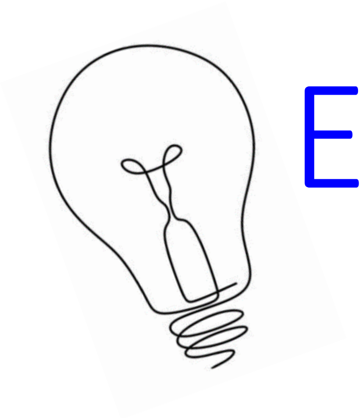
c) Using the histogram, estimate the modal time spend to brew the coffee for Coffee Shop X and comment on the value obtained.

(3 marks)

d) Given that the mean and standard deviation of the length of time to brew the coffee for Coffee Shop Z were 11 and 5.235 respectively. Determine which Coffee Shop has a more consistent distribution.

(3 marks)

Find CV and choose lowest CV



# Exam tips

CONFIDENTIAL 4 CS/JUN 2018/QMT181/212/216/STA104

e) Given the mean and standard deviation for the weight of 50 workers in Company B are 64.4 kg and 6.25 kg, respectively. Determine which company workers' weight distribution is more dispersed.

(3 marks)

Find CV and choose highest CV

# Skewness

- Other than measures of central tendency and dispersion, another important measure of distribution is the skewness of the distribution. A distribution can be symmetrical, skewed to the right and skewed to the left

- Measure of skewness

  If mean – mode = +ve (skewed to the right)

  If mean – mode = –ve (skewed to the left)

  If mean – mode = 0 (symmetrical)

# Pearson’s coefficient of skewness (PCS)

# Pearson coefficient of skewness

- Is usually used to measure the skewness of the distribution.

1 $$PCS = \frac{\bar{x} - \hat{x}}{s}$$

2 $$PCS = \frac{3(\bar{x} - \tilde{x})}{s}$$

- If skewness = 0 (symmetrical)
- If skewness = +ve (skewed to the right)
- If skewness = –ve (skewed to the left)

# Example 5.8

Given that the mean, mode and standard deviation of a set of data are 4,5 and 0.5 respectively, find the Pearson's coefficient of skewness and explain the distribution.

$$PCS = \frac{\bar{x} - \hat{x}}{s} = \frac{4-5}{0.5} = -2$$

U have mean, mode and standard deviation, use formula 1

This means that the distribution of the data is skewed to the left

b) The following table shows the frequency distribution of the number of books read by 300 students of a college university in 2014.

| Number of books read | Number of students |
|---|---|
| 0 - 2 | 48 |
| 3 - 5 | 67 |
| 6 - 8 | 112 |
| 9 - 11 | 55 |
| 12 - 14 | 18 |

i) Find the values of the mean and median. Explain the meaning of the values.

(8 marks)

ii) Calculate the standard deviation.

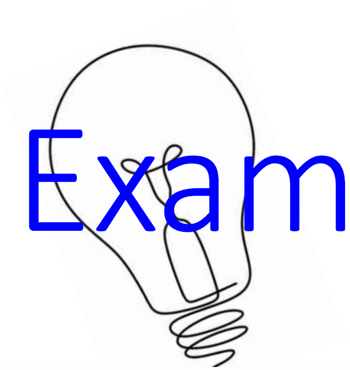
(3 marks)

iii) Calculate the coefficient of skewness by using an appropriate formula. What can you conclude?

(3 marks)

U have mean, median and standard deviation, use formula 2 to find PCS

# Exam tips



CONFIDENTIAL 4 CS/OCT 2016/QMT181/212/216/STA104

a) Calculate the mean and standard deviation for the data above.

(5 marks)

Median

b) Half of the callers spent at least 'Y' minutes for each call. Find the value of 'Y'.

(4 marks)

c) By using appropriate calculation, determine the skewness of the data distribution.

(3 marks)

U have mean, standard deviation and median, use formula 2 to find PCS

# THE END