MEASURES OF CENTRAL TENDENCY

MICHAEL KAKU MINLAH,ASSISTANT LECTURER, INSTITUTE OF PROFESSIONAL STUDIES

Chapter goals

- Compute and interpret the ***mean, median,*** and mode for a set grouped and ungrouped frequency distribution data.
- Compute the **range, variance,** and ***standard deviation*** and know what these values mean
- Compute and explain the ***coefficient of variation***

Definition-measures of central tendency

- Although frequency distributions may be complex at times, it is often very useful to be able to summarize or describe the distribution with a **single numerical value**.
- However, we need to take care to select a value that is the **most representative** of the **entire** distribution, that is of **all** of the individuals.

Definition-measures of central tendency

- ***Central tendency*** is a statistical measure that identifies a single score as representative of an entire distribution. The central tendency of the set of measurements-that is, the tendency of the data to cluster, or center, about certain numerical values.

Definition-measures of central tendency

- The goal of central tendency is to find the ***single score*** that is most typical or most representative of the entire group.
- There are three main measures of central tendency: the ***mean, median, and mode***.

MEAN

- Mean of a raw data
- Mean of a single value frequency
- Mean of a grouped frequency distribution

Definition of mean

- The **arithmetic mean**, or simply the **mean** or **average** is the central tendency of a collection of numbers taken as the sum of the numbers divided by the size of the collection.

- The arithmetic mean is a practical tool for comparing and measuring business data. It provides a way of assigning an average value to a set of numerical quantities. This average amount determines the midpoint of a data set also known as Central Tendency

Definition of mean

- While the arithmetic mean is often used to report central tendencies, it is greatly influenced by ***outliers***.

Formula for arithmetic mean

- The formula for the arithmetic mean for a raw data set is given by:
- For samples

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Where *n is the sample size.*

formula

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\Sigma X}{n}$$

Example of arithmetic mean for samples

- The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the mean.

  20, 26, 40, 36, 23, 42, 35, 24, 30

solution

$$\overline{X} = \frac{\Sigma X}{n} = \frac{20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30}{9} = \frac{276}{9} = 30.7 \text{ days}$$

Hence, the mean of the number of days off is 30.7 days.

Formula for arithmetic mean

- For population:

$$u = \frac{\sum_{i=1}^{N} X_i}{N}$$

- Where *N* is the population size

Arithmetic mean of a single value frequency distribution

- Example of a single value frequency distribution is given by the table below.

Number of vehicles serviceable (x)	Number of days(f)
0	2
1	5
2	11
3	4
4	4
5	1

Arithmetic mean of a single value frequency distribution

- The formula for arithmetic mean of single valued frequency distribution is given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i * x_i}{\sum_{i=1}^{n} f}$$

Calculation

Number of vehicles serviceable (x)	Number of days(f)	$f * x$
0	2	0
1	5	5
2	11	22
3	4	12
4	4	16
5	1	5
	$\sum f$ =27 $f * x$	$\sum f * x$ =60

calculation

- The mean number of cars serviceable is given by 60/27=2.22

Mean of a grouped frequency distribution

- The mean of a grouped frequency distribution is given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} f * x_m}{\sum_{i=1}^{n} f}$$

- Where x_m is the class midpoint

example

- Consider an example
- The following data relates to the number of successful sales made by the salesmen employed by a large microcomputer firm in a particular quarter.

example

Number of sales	Number of salesmen
0-4	1
5-9	14
10-14	23
15-19	21
20-24	15
25-29	6

solution

Number of sales (x)	Number of salesmen (*f*)	Midpoint (x_m)	$f * x_m$ $\sum_{i=1}^{n} f_i$
0-4	1	2	2
5-9	14	7	98
10-14	23	12	276
15-19	21	17	357
20-24	15	22	330
25-29	6	27	162
	$\sum_{i=1}^{n} f_i$ =80		$\sum_{i=1}^{n} f_i * x_{mi}$ =1225
Mean=($\sum_{i=1}^{n} f_i * x_{mi}$ =1225)/($\sum_{i=1}^{n} f_i$ =80)=15.3125			

interpretation

- The mean number of sales for that particular quarter is 15. 3125

Weighted mean

- The weighted mean is calculated as :

$$\bar{x}_{w} = \frac{\sum_{i=1}^{n} w_{i*} x_i}{\sum_{i=1}^{n} w_i}$$

- Where the weights are denoted by w_i

formula

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + \cdots + w_nX_n}{w_1 + w_2 + \cdots + w_n} = \frac{\Sigma wX}{\Sigma w}$$

where $w_1, w_2, \ldots, w_n$ are the weights and $X_1, X_2, \ldots, X_n$ are the values.

example

The costs of three models of helicopters are shown here. Find the weighted mean of the costs of the models.

example

Model	Number sold	Cost
Sunscraper	9	$427,000
Skycoaster	6	365,000
High-flyer	12	725,000

solution

$$\bar{x} = \frac{9(\$427,000) + 6(\$365,000) + 12(\$725,000)}{27} = \$545,666.67$$

Geometric mean

- It is mostly used in finding the averages of percentages, ratios, indexes or growth rates .
- It is given by the formula:

$$GM = \sqrt[N]{(X_1 * X_2 * X_3 * X_4 * .... * X_n)}$$

$$= \frac{1}{n}(\log X_1 + \log X_2 + \log X_3 + \log X_4 + ... + \log X_n)$$

example

- The profits earned by IPS company limited on four recent projects were 3%,2%, 4% and 6%. What is the geometric mean?
- The answer is given as:

$$GM = \sqrt[4]{(3*2*4*6)}$$

$$= \frac{1}{4}(\log 3 + \log 4 + \log 2 + \log 6)$$

$$= 3.46\%$$

median

- This refers to the midpoint of the data after the data has been ordered (preferably from lowest to highest).
- For a set of ***n*** observations arranged in order of magnitude, there are two cases:
  - If ***n*** is odd, then the median is given by the $\frac{1}{2}(n+1)th$ Observation.

median

- If on the other hand, ***n*** is even then the median is given by the mean of the

$$\frac{1}{2}\left(\frac{1}{2}n+\frac{1}{2}n+1\right)th\_observation$$

Median

- <u>WHEN ***n*** is ODD</u>
- Consider the example 12,15,22,17,20,26,22,26,12.
- After ordering we have 12,12,15,17,20,22,22,26,26.
- Given that ***n=***9, the mean corresponds to the 5th observation which is 20.

Median

- <u>WHEN ***n*** is EVEN</u>
- Consider the example. Given the data:
- 4,7,9,10,5,1,3,4,12,10
- Arranging in order of magnitude gives:
- 1,3,4,4,5,7,9,10,10,12
- The $\frac{1}{2}nth$, which is the 5th observation corresponds to 5 and the $\frac{1}{2}n+1$ observation corresponds to 7. The median is therefore 6.

Median of a single value frequency distribution

Ages(x)	frequency	Cumulative frequency
3	3	3
4	6	9
5	6	15
6	7	22
7	4	26
8	4	30

solution

- Since ***n*** is even the median will lie between the $\frac{1}{2}nth$ and the $\frac{1}{2}n+1$ observation. The $\frac{1}{2}nth$

Observation, from the cumulative frequency corresponds to 5 and the $\frac{1}{2}n+1$ observation corresponds to 6. the median is therefore the average of 5 and 6 which is 5.5

Median of a grouped frequency distribution

- Steps to follow;
- 1. find the median class. This is the class which corresponds to the $\frac{1}{2}nth$ observation.
- 2. after finding the median class, apply the formula for the median which is given by:

$$median = L_c + \left(\frac{\frac{n}{2} - F}{f_m} \right) c$$

MICHAEL KAKU MINLAH,ASSISTANT LECTURER, INSTITUTE OF PROFESSIONAL STUDIES

Median of a grouped frequency distribution

- Where L_c is the lower class limit of the median class. ***n*** is the sample size. *F* is the cumulative frequency of the class before the median class. f_m is the frequency of the median class and *c* is the class width of the median class.

example

Number of sales	Number of salesmen
0-4	1
5-9	14
10-14	23
15-19	21
20-24	15
25-29	6

solution

Number of sales	Number of sales men	Cumulative number of sales
0-4	1	1
5-9	14	15
10-14	23	38
15-19	21	59
20-24	15	74
25-29	6	80

solution

- The median class corresponds to n/2, which is given by 80/2=40. checking the cumulative frequency we observe that the median class corresponds to {15-19}. *F=38.* f_m is 21. c=5. L_c is 14.5. the median is therefore :

$$l_c = 14.5 + \left(\frac{40-38}{21} \right) 5 = 14.9762$$

mode

- A measure of central tendency
- Value that occurs ***most often***
- Not affected by extreme values (outliers)
- Used for either numerical or categorical data
- There may be no mode
- There may be several modes(uni-modal, bi-modal or multi-modal)

Mode of a grouped frequency distribution

- The formula is : $$\mathrm{mod}\,e = L_c + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)c$$
- Where L_c is the lower class boundary of the modal class. *C is* class width. Δ_1 is the difference between the frequency of the modal class and the frequency of the class preceding the modal class.
- Δ_2 is the difference between the frequency of the modal class and the frequency of the class following the modal class

Example and solution

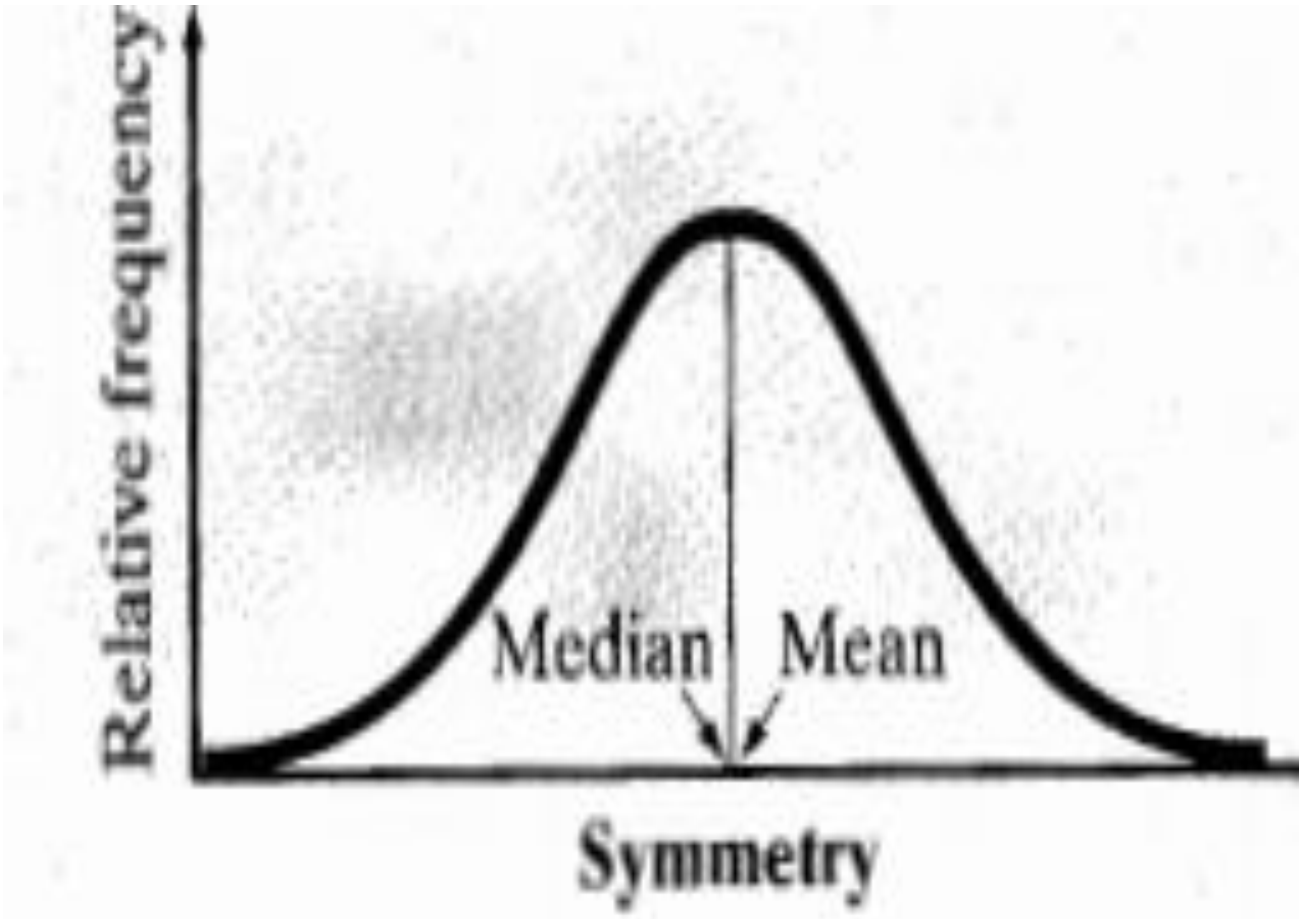
- From the example, the modal class is {10-14}.
- L_c is 9.5. Δ_1 is 9. Δ_2 is 2. *C* is 5. The mode is therefore :

$$\text{mode} = 9.5 + \left(\frac{9}{9+2}\right)5 = 13.5909$$

Shape of distribution

- Describes how data is distributed
- Is the data Symmetric or skewed?
- Histogram gives an idea about the shape of distribution
- What is the relation between mean, mode and median for symmetrical and skewed data
- For symmetric distributions the mean divides the distribution into two equal halves

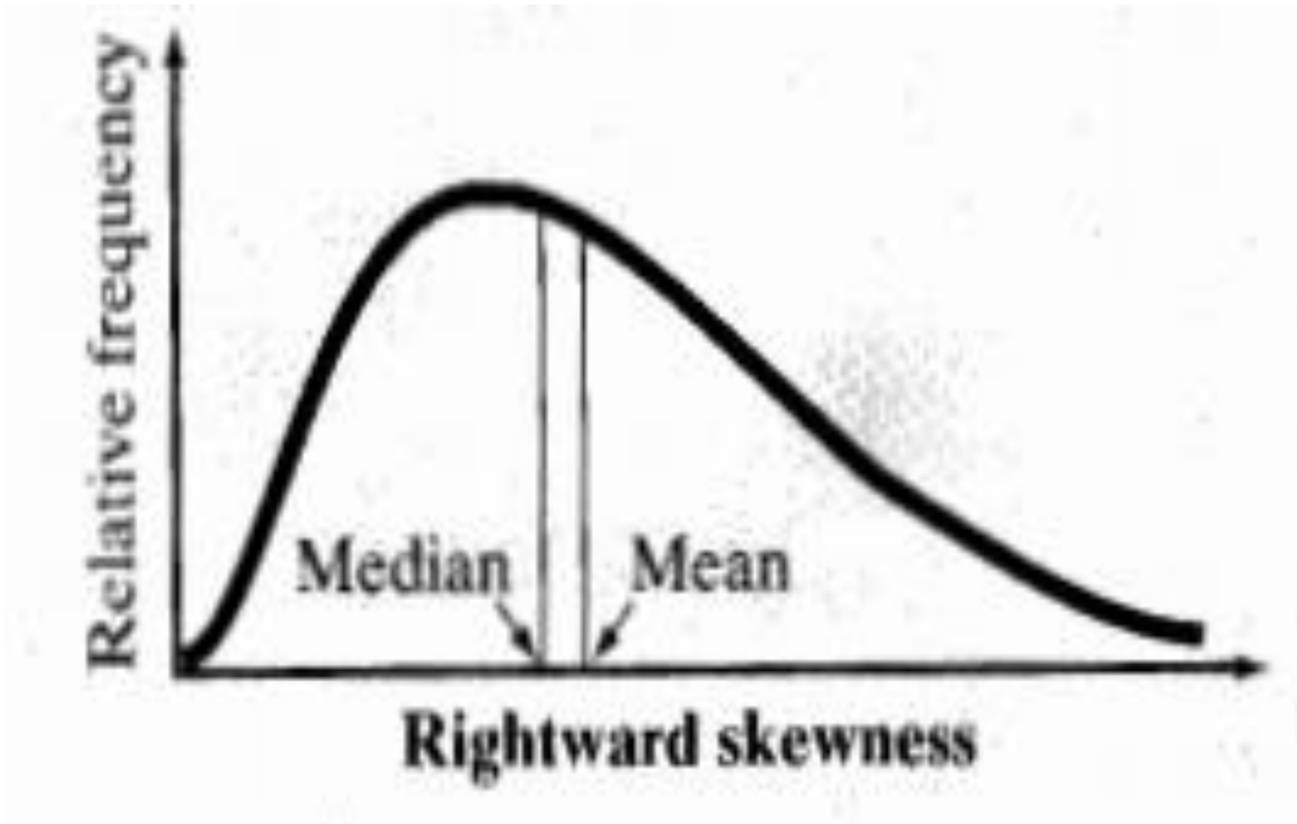
Normal distribution



Right- skewed data

- The tail of the distribution is to the right
- More data occurs to the left
- Mode <median<mean

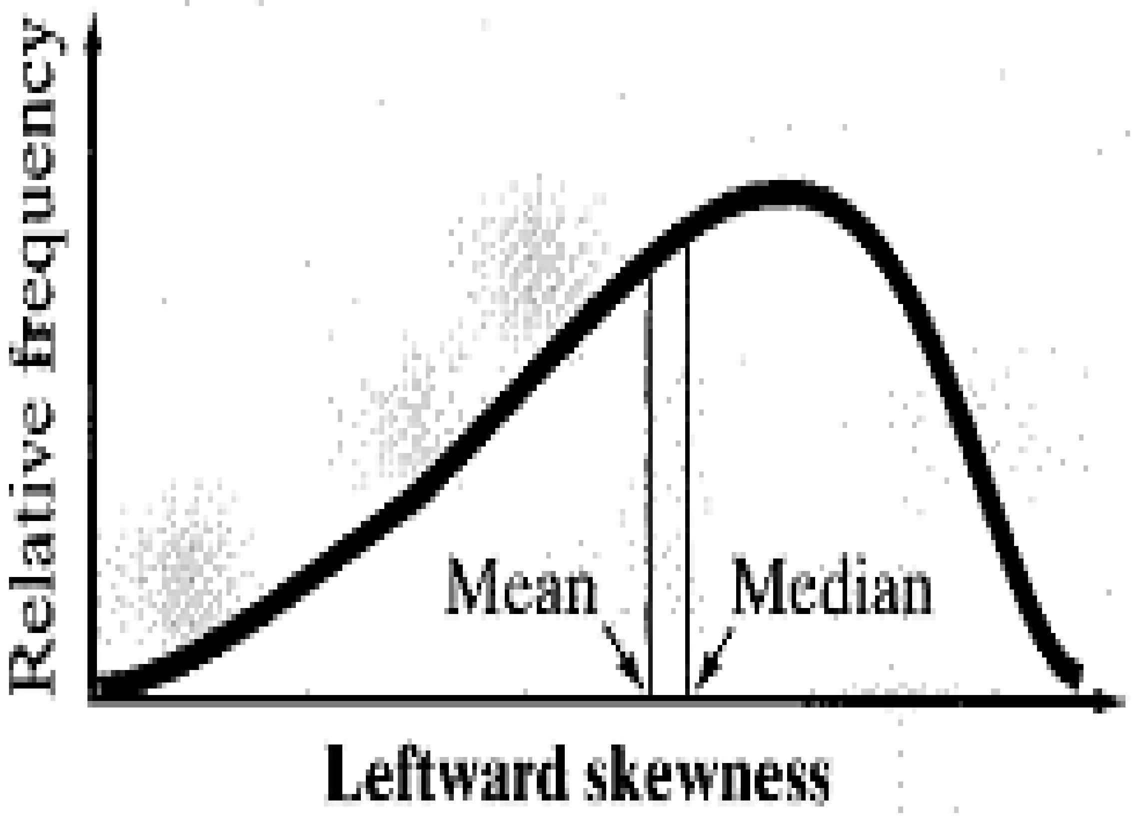
Right- skewed distribution



Left-skewed distribution

- The tail of the distribution is to the left
- More data occurs to the right
- mean<median< Mode

Left –skewed distribution



Measures of dispersion

- Measures the amount of dispersion in a given data set. It gives degree to which numerical data tend to spread about an average value. Examples include:
- Range
- Mean deviation
- Variance
- Coefficient of variation

range

- It is defines as the difference between the highest and the lowest value in the data set.

Mean deviation

- It measures the average absolute difference between each item and the mean. It is given as: $$MD = \frac{\sum | X - \bar{X} |}{\sum f}$$

- For a single valued frequency distribution it is given by: $$MD = \frac{\sum f | X - \bar{X} |}{\sum f}$$

example

- The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the absolute mean deviation.

  20, 26, 40, 36, 23, 42, 35, 24, 30

Mean is given by 30.67

solution

X	$X-\bar{X}$	$\lvert X-\bar{X} \rvert$
20	-10.67	10.67
26	-4.67	4.67
40	9.33	9.33
36	5.33	5.33
23	-7.67	7.67
42	11.33	11.33
35	4.33	4.33
24	-6.67	6.67
30	-0.67	0.67
		$\bar{X}=6.67$

Mean deviation for grouped frequency distribution

- Formula:

$$MD = \frac{\sum f \mid X_m - \bar{X} \mid}{\sum f}$$

variance

- Measures how spread out the data are around the mean. A high value depicts high variability whereas low value depicts low variability.
- We can calculate variance for both population and sample.
- Population variance is given by:
- $$\delta^2 = \frac{\sum_{i=1}^{N}(X_i - u)^2}{N} \quad \text{or} \quad \delta^2 = \frac{\sum_{i=1}^{N} X_i^{\,2} - \frac{(\sum_{i=1}^{N} X_i)^2}{N}}{N}$$

illustration

- A testing lab wishes to test two experimental brands of outdoor paint to see how long each will last before fading. The testing lab makes 6 gallons of each paint to test. Since different chemical agents are added to each group and only six cans are involved, these two groups constitute two small populations. The results (in months) are shown. Find the mean of each group.

illustration

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

Solution

The mean for brand A is

$$\mu = \frac{\Sigma X}{N} = \frac{210}{6} = 35 \text{ months}$$

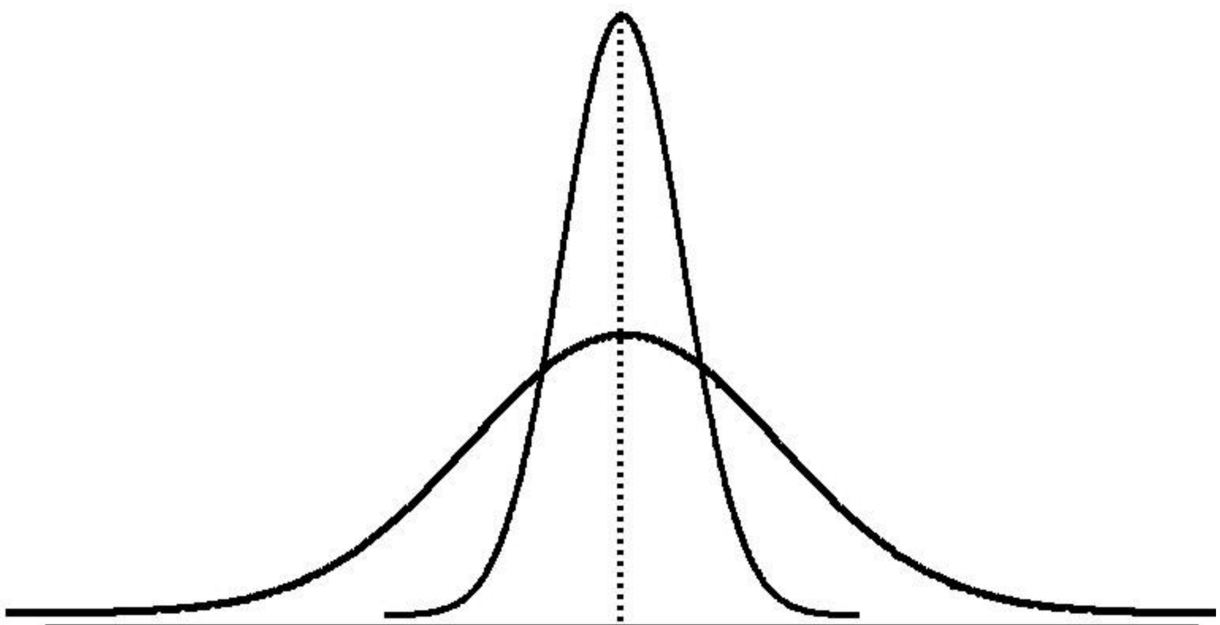
The mean for brand B is

$$\mu = \frac{\Sigma X}{N} = \frac{210}{6} = 35 \text{ months}$$

illustration

- Since both populations have the same mean, how can we determine which is more variable?
- We thus employ the concept of variation

Graphical depiction of variation: Same mean, different variance



Example-calculation of population variance

- 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

SOLUTION

X	X^2
11.2	125.4
11.9	141.61
12	144
12.8	163.84
13.4	179.56
14.3	204.49
$\sum X$ =75.6	$\sum X^2$ =958.94
$\delta^2 = \dfrac{\sum X^2 - \dfrac{\left(\sum X\right)^2}{N}}{N}$ $= \dfrac{958.94 - \dfrac{75.6^2}{6}}{6}$ =1.06	

Population variances for a simple frequency distribution

$$\delta^2 = \frac{\sum_{i=1}^{N} f_i X_i^{\,2} - \frac{(\sum_{i=1}^{N} f_i X_i)^2}{N}}{N}$$

Or

$$\delta^2 = \frac{\sum_{i=1}^{N} f_i (X_i - U)^2}{N}$$

example

Ages (X)	Frequency (f)
3	3
4	6
5	6
6	7
7	4
8	4

solution

Ages (X)	Frequency (f)	X^2	F*X	$F*X^2$
3	3	9	9	27
4	6	16	24	96
5	6	25	30	150
6	7	36	42	252
7	4	49	28	196
8	4	64	32	256
$\delta^2 = \frac{\sum f * X^2 - \frac{\left(\sum f * X\right)^2}{N}}{N}$ $= \delta^2 = \frac{977 - \frac{165^2}{30}}{30}$ =2.32				

Population variances for a grouped frequency distribution

$$\delta^2 = \frac{\sum_{i=1}^{N} f_i (X_{mi} - U)^2}{N}$$

or

$$\delta^2 = \frac{\sum_{i=1}^{N} f_i X_{mi}{}^2 - \frac{\left(\sum_{i=1}^{N} f_i X_{mi}\right)^2}{\sum f}}{N}$$

example

Number of sales	Number of salesmen
0-4	1
5-9	14
10-14	23
15-19	21
20-24	15
25-29	6

solution

Number of sales (X)	Number of salesmen (f)	X_m	X_m^2	$f^* X_m$	$f^* X_m^2$	
0-4	1	2	4	2	4	
5-9	14	7	49	98	686	
10-14	23	12	144	276	3312	
15-19	21	17	289	357	6069	
20-24	15	22	484	330	7260	
25-29	6	27	729	162	4374	
$\delta^2 = \frac{\sum f * x_m{}^2 - \frac{\left(\sum f * x_m\right)^2}{N}}{N}$ $\delta^2 = \frac{21505 - \frac{1225^2}{80}}{80}$ =34.34						

Sample variance

- $$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$ or $$s^2 = \frac{\sum_{i=1}^{n} X_i^{\,2} - \frac{\left(\sum_{i=1}^{n} X\right)^2}{n}}{n-1}$$
- Sample variance for single value frequency distribution is given as:
- $$s^2 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})^2}{n-1}$$ or $$s^2 = \frac{\sum_{i=1}^{n} f_i X_i^{\,2} - \frac{\left(\sum_{i=1}^{n} f_i X\right)^2}{n}}{n-1}$$

Sample variance for raw data

X	X^2
11.2	125.4
11.9	141.61
12	144
12.8	163.84
13.4	179.56
14.3	204.49
$\sum X$ =75.6	$\sum X^2$ =958.94
$s^2 = \dfrac{\sum X^2 - \dfrac{(\sum X)^2}{n}}{n-1}$ $= \dfrac{958.94 - \dfrac{75.6^2}{6}}{6-1}$ =1.28	

Sample variance for simple frequency distribution

Ages (X)	Frequency (f)	X^2	F*X	F*X^2
3	3	9	9	27
4	6	16	24	96
5	6	25	30	150
6	7	36	42	252
7	4	49	28	196
8	4	64	32	256
$s^2 = \frac{\sum f * X^2 - \frac{\left(\sum f * X\right)^2}{n}}{n-1}$ $= s^2 = \frac{977 - \frac{165^2}{30}}{29}$ =2.4				

Sample variance for a grouped frequency distribution

- $$s^2 = \frac{\sum_{i=1}^{n} f_i (X_{mi} - \bar{X})^2}{n-1}$$ or $$s^2 = \frac{\sum_{i=1}^{n} f_i X_{mi}{}^2 - \frac{\left(\sum_{i=1}^{n} f_i X_{mi}\right)^2}{n}}{n-1}$$

Sample variance for grouped data

Number of sales (X)	Number of salesmen (f)	X_m	X_m^2	$f^* X_m$	$f^* X_m^2$	
0-4	1	2	4	2	4	
5-9	14	7	49	98	686	
10-14	23	12	144	276	3312	
15-19	21	17	289	357	6069	
20-24	15	22	484	330	7260	
25-29	6	27	729	162	4374	
$s^2 = \dfrac{\sum f * x_m{}^2 - \dfrac{\left(\sum f * x_m\right)^2}{n}}{n-1}$ $s^2 = \dfrac{21505 - \dfrac{1225^2}{80}}{79}$ =34.77						

Comparing distributions

- When distributions have the same mean, the distribution with the larger variance is more variable
- What if the distributions have different means?
  - We employ the co efficient of variation

Coefficient of Variation

- Measures relative variation
- Always in percentage (%)
- Shows variation relative to mean
- Is used to compare two or more sets of data measured in different units

Co efficient of variation

- For population

$$CV = \frac{\delta}{u} * 100$$

- For samples

$$CV = \frac{s}{\bar{x}} * 100$$

example

- Stock A:
  - Average price last year = $50
  - Standard deviation = $5

- Stock B:
  - Average price last year = $100
  - Standard deviation = $5

solution

$$CV_A = \left(\frac{S}{\bar{X}} \right) * 100 = \frac{\$5}{\$50} * 100 = 10\%$$

$$CV_B = \left(\frac{S}{\bar{X}} \right) * 100 = \frac{\$5}{\$100} * 100 = 5\%$$

example

- 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.
- The co efficient of variation is given by:

$$CV = \frac{s}{\bar{x}} * 100$$

solution

- Sales:

$$CV_{sales} = \frac{5}{87} * 100 = 5.75\%$$

- Commission:

$$CV_{commission} = \frac{773}{5225} * 100 = 14.79\%$$