Introduction to Statistics

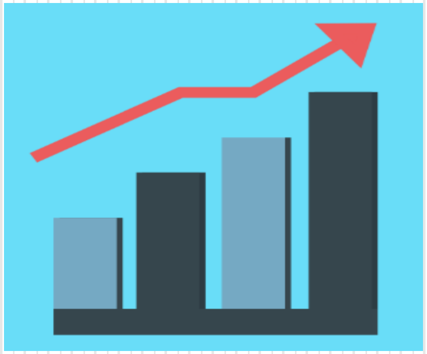
Dr. Farman Ali

Assistant Professor

DEPARTMENT OF SOFTWARE

SEJONG UNIVERSITY

Lecture-4



Course Syllabus

- Introduction to the course
- Sampling and data presentation
- Basic of probability
- Distributions
- Confidence intervals
- Hypothesis testing
- Correlation and simple linear regression
- Multiple regression

Outline

- **Measures of Central Tendency**
  - Midrange
  - Weighted Mean
- **Measures of Position**
  - Percentiles
  - Quartiles
  - Deciles
- **Measures of Variation**
  - Range
  - Variance
  - Standard deviations

Midrange

- Measures of average are called measures of central tendency.

- Summarize data, using measures of central tendency, such as the Mean, Median, Mode, Midrange, and Weighted Mean.

- **Midrange:** The arithmetic mean of the largest and the smallest values in a data set.
- The midrange is the average of the lowest and highest values in a data set.

$$MR = \frac{Lowest + Highest}{2}$$

Example: Find the midrange of the following data set.

12, 34, 56, 98, 10

$$= \frac{10 + 98}{2} \quad \Rightarrow \frac{108}{2} \quad \Rightarrow 54 \text{ is Midrange}$$

Median and Mid range From the Frequency Table

How To Find The Median and midrange Of A Frequency Table When The Number Of Observations?

Case 1: when the number of observations (n) is odd, then the median is the value at the $\left(\frac{n+1}{2}\right)^{th}$ position.

Example: The following is a frequency table of the score obtained in a mathematics quiz. Find the median score.

Score	0	1	2	3	4
Frequency	3	4	7	6	3

Solution:

Number of scores = 3 + 4 + 7 + 6 + 3 = 23 (odd number)

Since the number of scores is odd, the median is at the $\left(\frac{n+1}{2}\right)^{th} = \left(\frac{23+1}{2}\right)^{th} =$ 12^{th} position.

Median and Mid range From the Frequency Table

How To Find The Median and midrange Of A Frequency Table When The Number Of Observations?

Score	0	1	2	3	4
Frequency	3	4	7	6	3

To find out the 12^{th} position, we need to add up the frequencies as shown:

Score	0	1	2	3	4
Frequency	3	4	7	6	3
Position	3	$3 + 4 = 7$	$7 + 7 = 14$		

The 12th position is after the 7th position but before the 14th position. So, the median is 2.

0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4

Range= 4-0=4

Midrange= 4/2=2

The weighted Mean

- Mean that is calculated with extra weight given to one or more elements of the sample.
- The weighted mean is similar to an ordinary arithmetic mean (the most common type of average),except that instead of each of the data points contributing equally to the final average, some data points contribute more than others.

$$\bar{X} = \frac{\sum W\ X}{\sum W} \qquad \bar{X} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \ldots + w_nx_n}{w_1 + w_2 + w_3 + w_4 \ldots . + w_n}$$

The weighted Mean

Example1. Calculate the weighted mean of the sample of numbers shown below:

16, 20, 12, 16, 16, 10, 16, 20, 24, 20

The weighted Mean

Example1. Find weighted mean for following data set w = {2, 5, 6, 8, 9}, x = {4, 3, 7, 5, 6}:

Given data sets w = {2, 5, 6, 8, 9}, x = {4, 3, 7, 5, 6} and N = 5

$$Weighted\ mean = \frac{\sum(weights \times quantities)}{\sum(weights)}$$

= (w1x1 + w2x2 + w3x3 + w4x4 + w5x5) / (w1 + w2 + w3 + w4 + w5)

$= (2 \times 4 + 5 \times 3 + 6 \times 7 + 8 \times 5 + 9 \times 6) / (2 + 5 + 6 + 8 + 9)$

$= (8 + 15 + 42 + 40 + 54) / 30 \Rightarrow 159 / 30 = 5.3$

Therefore, the weighted mean is 5.3.

The weighted Mean

Example 2. A teacher provides the following weightage of 20% for class attendance, 30% for project work, 40% for tests, and 10% for home assignments. A student scores 80/100 for class attendance, 4/5 in project work, 35/50 in tests, and 8/10 in home assignments. Find the final score of the student.

Solution:

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

$$= \frac{20\%.\frac{80}{100} + 30\%.\frac{4}{5} + 40\%.\frac{35}{50} + 10\%.\frac{8}{10}}{20\% + 30\% + 40\% + 10\%}$$

$$= \frac{20\%.0.8 + 30\%.0.8 + 40\%.0.7 + 10\%.0.8}{100\%}$$

$$= \frac{0.2 \times 0.8 + 0.3 \times 0.8 + 0.4 \times 0.7 + 0.1 \times 0.8}{1}$$

$$= 0.16 + 0.24 + 0.28 + 0.08$$

$$= 0.76$$

Therefore, the final score of the student is 0.76.

The weighted Mean

Example 3. For a job application 0.8 weightage is given to academic qualification, 0.7 is given to personality, 0.4 is given to the location. The prospective candidate scores 4.5/5 for academic qualification, 3/5 for personality, and 2.8/5 for location. Find the final score received by the candidate.

Solution:

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

$$= \frac{0.8 \times \frac{4.5}{5} + 0.7 \times \frac{3}{5} + 0.4 \times \frac{2.8}{5}}{0.8 + 0.7 + 0.4}$$

$$= \frac{0.8 \times 0.9 + 0.7 \times 0.6 + 0.4 \times 0.56}{0.19}$$

$$= \frac{0.72 + 0.42 + 0.224}{0.19}$$

$$= \frac{1.364}{0.19}$$

$$= 7.18$$

Hence, the final score of the candidate is 7.18.

The weighted Mean

Example2. In a class of 20, eight students averaged a score of 86, seven students had an average of 74, and five students had an average test score of 98. what is the average test score for the entire class?

The weighted Mean

Example3. In a certain college, 20% of students have an average weight of 140 lbs, 35% of students have an average weight of 160lbs, 30% of students have an average weight of 175lbs, and 15% of students have an average weight of 195lbs. Based on this data, what is the average weight of all students in this high school.

The weighted Mean

Example: A student received the following grades in different subjects. Find the GPA using weighted mean.

Subjects	Credit w	Grade X
Physics	4	D 1 points
Chemistry	3	B 2 points
Biology	3	C 3 points
Computer	2	A 4 points

$$\bar{X} = \frac{\sum W\,X}{\sum W}$$

$$= \frac{4*1+3*2+3*3+2*4}{4+3+3+2}$$

$$= \frac{4+6+9+8}{12}$$

$$= 27/12$$

$$= 2.25\ approx$$

Learning Objectives: *Spread or Variability of Data Set:*

1. **Range**
2. **Variance**
3. **Standard Deviations**

- In order to describe the data set more accurately, Statisticians use measure of variations instead of measure of central tendency

Measure of Central Tendency

1. Mean
2. Median
3. Mode
4. Midrange
5. Weighted Mean

Measure of Position

1. Percentiles
2. Quartiles
3. Deciles

Measure of Variations

1. Range
2. Variance
3. Standard deviations

Measures of Position

- They are used to locate the relative position of the data value in the data set. The most common measures of position are percentiles, Deciles, and quartiles.
- **Percentile**: A percentile is a value below which a certain percentage of observations lie.
- Example: You are the fourth tallest person in a group of 20
- This means 80% of people are shorter than you in your class
- That means you are at the **80th percentile**.

- If your height is 1.85m then "1.85m" is the 80th percentile height in that group.
- Percentiles are not the same as percentages.
- Example: if a student gets 67 points in a test out of 100, it means she has 67%
- There is no indication of her position with respect to her class. May be her score is highest, the lowest, or Somewhere in between.
- If 67 points corresponds to 58^{th} percentile, then it means she performed better than 58 % of her class

Percentile

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

$\{140, 150, 160, 168, 170, 170, 175, 180, 182, 190\}$

We can call this "170" 5th percentile   This means 50% marks are less than 170

- **Example:** The marks of 10 students are given below.
  76, 56, 59, 87, 90, 34, 49, 48, 75, 62. Find the percentile rank of the score of 62.

$$Percentage = \frac{Obtained\ marks}{total\ marks} * 100 \qquad p^{th} percentile = \frac{(No\ of\ values\ below\ X)}{Total\ No\ of\ Values} * 100$$

- **Solution:** Arrange the data in order:

34, 48, 49, 56, 59, 62, 75, 76, 87, 90

Put the values in formula:

$$Percentile = \frac{5}{10} * 100$$

$= 50^{th}$ Percentile

A student whose score was 62, did better than 50%

Percentile

Finding a data value corresponding to a given percentile

- Arrange the data in order from lowest to highest

- Put the value into the formula $c = \dfrac{n.p}{100}$

n= total number of values
P= percentile

- if c is not a whole number, round up to the next whole number. Starting at the lowest values, count to the number that corresponds to the rounded-up values.

- If c is a whole number, use the value halfway between the cth and (c+1)st value when counting up from the lowest value.

Percentile

- **Example:** The marks of 10 students are given below.
  76, 56, 59, 87, 90, 34, 49, 48, 75, 62. Find the value corresponding to 25th percentile.
- **Solution:** Arrange the data in order:

34, 48, 49, 56, 59, 62, 75, 76, 87, 90

Put the values in formula:

$$c = \frac{n.p}{100}$$

$$c = \frac{10*25}{100}$$

$$c = 2.5$$

If c is not a whole number, round up to the next whole number. So C=3, now start at the lowest value and count over to the third value, which is 49, so the values 49 corresponds to 25th percentile.

Percentile

➢**Example:** For the following height data collected from students find the 10th and 95th percentiles. 91, 89, 88, 87, 89, 91, 87, 92, 90, 98, 95, 97, 96, 100, 101, 96, 98, 99, 98, 100, 102, 99, 101, 105, 103, 107, 105, 106, 107, 112.

➢**Solution:** The ordered observations of the data are 87, 87, 88, 89, 89, 90, 91, 91, 92, 95, 96, 96, 97, 98, 98, 98, 99, 99, 100, 100, 101, 101, 102, 103, 105, 105, 106, 107, 107, 112.

$$c = \frac{n.p}{100} \qquad c = \frac{30*10}{100} = 3$$

So the 10th ***percentile*** i.e P_{10} is the 3rd observation in sorted data is 88, which means that 10 percent of the observations in the data set are less than 88.

$$c = \frac{30*95}{100} = 28.5$$

29th observation is our ***95th percentile*** *i.e.* $P_{95}=107$.

Percentile grouped data

- The mth ***percentile*** (***a measure of the relative standing of an observation***) for grouped data is

$$P_m = l + \frac{\frac{m.n}{100} - cf}{f} \times c)$$

Where

L = lower limit of the percentile class

n = total number of observations (total frequency)

m = mth percentile to find

cf = cumulative frequency of the class preceding the percentile class

f = frequency of the percentile class

C= class length / class size

Percentile grouped data

Example: Calculate the P_{53} for the following data:

Marks	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Number of students	6	20	37	10	7

Solution: We need to calculate the cumulative frequencies to find the P_{53}.

Marks	Number of students	Cumulative frequency
0 - 20	6	6
20 - 40	20	26
40 - 60	37	63
60 - 80	10	73
80 - 100	7	80

Percentile grouped data

$$P_m = l + \frac{\frac{m.n}{100} - cf}{f} \times c)$$

Example: Calculate the P_{53} for the following data:

Where

L = lower limit of the percentile class =40

n = total number of observations (total frequency)= 80

m = mth percentile to find =53

cf = cumulative frequency of the class preceding the percentile class =26

f = frequency of the percentile class = 37

C= class length / class size= 20

Find the percentile class

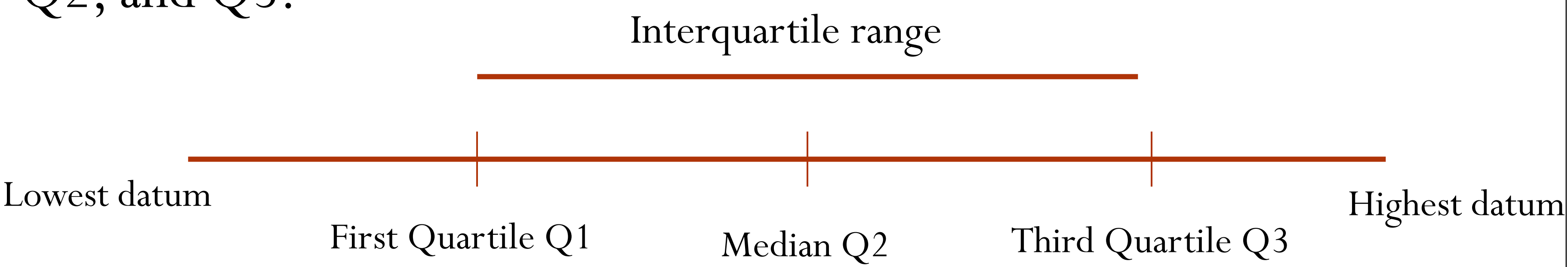
$$\frac{53.80}{100} = 42.4$$

$$P_{53} = 40 + \frac{42.4 - 26}{37} \times 20$$

$$P_{53} = 48.86$$

Quartile

➢ Quartile divide the data into 4 equal groups, and represented by Q1, Q2, and Q3.



Steps to find Quartiles:

1. Arrange the data in order from lowest to highest.

2. Find the median of the data values. This is the value for Q2.

3. Find the median of the data values that fall below Q2. This the value of Q1.

4. Find the median of the data values that fall above Q2. This is the values for Q3

Quartile

- **Example:** Find Quartiles Q1, Q2, and Q3.

  3, 4, 8, 5, 10, 9, 1, 4, 6, 12, 2.

- **Solution:** Arrange the data in order: 1, 2, 3, 4, 4, 5, 6, 8, 9, 10, 12

n=11 Q2=?

$= (n+1) * 50\% \quad \Rightarrow (11+1) * 0.5 \quad \Rightarrow 6$

The number which is at 6^{th} position is Q2. which is 5 is median.

- Arrange the data in order: 1, 2, 3, 4, 4, 5, 6, 8, 9, 10, 12

n=11 Q1=?

$= (n+1) * 25\% \quad \Rightarrow (11+1) * 0.25 \Rightarrow 3$

The number which is at 3^{th} position is Q1. which is 3.

- n=11 Q3=?

$= (n+1) * 75\% \quad \Rightarrow (11+1) * 0.75 \Rightarrow 9$

The number which is at 9^{th} position is Q3. which is 9.

Quartile

- **Example:** Find Quartiles Q1, Q2, and Q3.

  3, 4, 8, 5, 10, 9, 1, 4, 6, 12, 2, 14.

- **Solution:** Arrange the data in order: 1, 2, 3, 4, 4, 5, 6, 8, 9, 10, 12, 14

n=12 Q2=?

$= (n + 1) * 50\% \quad \Rightarrow (12 + 1) * 0.5 \quad \Rightarrow 6.5$

The number which is at 6^{th} and 7^{th} position is Q2. which is 5.5 is median

- Arrange the data in order: 1, 2, 3, 4, 4, 5, 6, 8, 9, 10, 12, 14

n=12 Q1=?

$= (n + 1) * 25\% \quad \Rightarrow (12 + 1) * 0.25 \Rightarrow 3.25$

The number which is at 3^{th} and 4^{th} position is Q1. which is 7/2= 3.5

- n=12 Q3=?

$= (n + 1) * 75\% \quad \Rightarrow (12 + 1) * 0.75 \Rightarrow 9.75$

The number which is at 9^{th} and 10^{th} position is Q3. which is 19/2= 9.5.

Quartile for grouped data

- The Quartile formula for grouped data is

$$Q_m = l + \left(\frac{\frac{m.n}{4} - cf}{f}\right) \times c)$$

Where

L = lower class boundary of the quartile class

n = total number of observations (total frequency)

m = mth Quartile to find

cf = cumulative frequency of the class preceding the Quartile class

f = frequency of the Quatile class

C= class length / class size

Quartile for grouped data

➢ Eighty randomly selected lightbulbs were tested to determine their lifetimes (in hours). The following frequency distribution was obtained. Find Q3.

$$Q_m = l + \left(\frac{\frac{m.n}{4} - cf}{f} \right) \times c$$

Classes	f	Class boundaries	cf
53 – 63	6	52.5 – 63.5	6
64 – 74	12	63.5 – 74.5	18
75 – 85	25	74.5 – 85.5	43
86 – 96	18	85.5 – 96.5	61
97 – 107	14	96.5 – 107.5	75
108 – 118	5	107.5 – 118.5	80

Find the Quartile class

$$\frac{3 * 80}{4} = 60$$

$$Q_3 = 85.5 + \left(\frac{60 - 43}{18} \right) \times 11$$

$$Q_3 = 95.89$$

Deciles

- Decile divide the distribution into 10 groups, and is denoted by D1, D2, D3, and so on.
- Formula: $D = \frac{m}{10} * (n + 1)$

- **Example:** Find Decile D7 and D8 of the following 10 students.
  34, 43, 80, 50, 60, 92, 51, 49, 65, 72.
- **Solution:** Arrange the data in order: 34, 43, 49, 50, 51, 60, 65, 72, 80, 92
- Formula: $D7 = \frac{m}{10} * (n + 1)$

$$D7 = \frac{7}{10} * (10 + 1)$$

$$D7 = \frac{77}{10} \Rightarrow 7.7 \Rightarrow 8 \ approx.$$

D7 is the 8^{th} element, which is 72

Deciles

- **Example:** Find Decile D7 and D8 of the following 10 students.
  34, 43, 80, 50, 60, 92, 51, 49, 65, 72.
- **Solution:** Arrange the data in order: 34, 43, 49, 50, 51, 60, 65, 72, 80, 92
- Formula: $D8 = \frac{m}{10} * (n + 1)$

$$D8 = \frac{8}{10} * (10 + 1)$$

$$D8 = \frac{88}{10} \Rightarrow 8.8 \Rightarrow 9 \; approx.$$

D8 is the 9th element, which is 80

Exercise: Suppose we have the following data: 2, 3, 5, 6, 7, 9, 9, 11, 12, 15

- What is the mean of these data?
- What is the median?
- What is the first quartile?
- What is the third quartile?

Decile for grouped data

- Eighty randomly selected lightbulbs were tested to determine their lifetimes (in hours). The following frequency distribution was obtained. Find D4.

$$D_m = l + \left(\frac{\frac{m.n}{10} - cf}{f}\right) \times c$$

Classes	f	Class boundaries	cf
53 – 63	6	52.5 – 63.5	6
64 – 74	12	63.5 – 74.5	18
75 – 85	25	74.5 – 85.5	43
86 – 96	18	85.5 – 96.5	61
97 – 107	14	96.5 – 107.5	75
108 – 118	5	107.5 – 118.5	80

Find the decile class

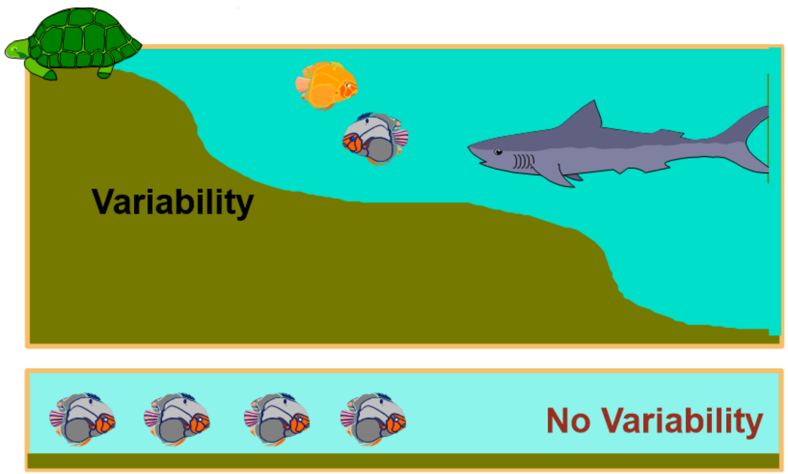
$$\frac{4 * 80}{10} = 32$$

$$D_4 = 74.5 + \left(\frac{32 - 18}{25}\right) \times 11$$

$$D_4 = 80.66$$

Measure of Variations

- The measures of central tendency are not adequate to describe data. Two data sets can have the same mean but they can be entirely different.
- Thus to describe data, one need to know the extent of variability. This is given by the measures of variations.



Measure of Variations

- There are 3 famous values used to measure the amount of dispersion or variation. (The spread of the group)

1. Range **2. Variance** **3.Standard deviation.**

Range: The range is the difference between the highest and lowest values in data set and is simplest of the three measures.
Symbolically,

Range=R=$Range = R = X_m - X_o$

where, X_m is the largest observation, X_o is the smallest observation

In case of Grouped Data: Range is the difference between the upper boundary of the highest class and the lower boundary of the lowest class.
Note: Range can’t be computed if there are any open-end classes in the frequency distribution

Measure of Variations

Two serious disadvantages of range

1. It ignores all the observations available from the intermediate observations.

2. Since it is based only on two extreme observation, so it might give misleading picture of the spread of the data

This is an absolute measure of dispersion.

Its relative measure known as the coefficient of dispersion.

$$\text{Coefficient of Dispersion} = \frac{(X_m - X_0)}{(X_m + X_0)}$$

This is a dimensionless number and thus has no unit and it is used for the purpose of comparison.

Example: The marks obtained by 9 students are given below:

45, 32, 37, 46, 39, 36, 41, 48, 36

Find the Range and the coefficient of dispersion.

Measure of Variations

- **Variance:** The Variance is the average of squares of the distance each value is from the mean. The Symbol for the population variance is σ^2(σ is the Greek lowercse letter sigma)

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

Where

X= individual value

μ = population mean

N= population size

- **Standard deviation:** Take the square root of the variance to find the standard deviation.

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

Measure of Variations

➢ **Example:**

Dataset 1

-10, 0, 10, 20, 30

Dataset 2

8, 9, 10, 11, 12

Range: $R = X_m - X_o$

Variance: $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$

Standard deviation: $\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$

Example: Light Bulbs

- Two different brands of light bulbs are tested to see how long each will last before it fuse. Seven bulbs of each brand make a small population. The results (in weeks) are shown. Find the mean and range of each group.

Brand A	Brand B
20	10
34	20
30	30
35	40
45	50
40	49
55	60

Mean of brand A

$$\mu = \frac{\sum X}{N}$$

$$\mu = \frac{259}{7}$$

$$\mu = 37\ weeks$$

Mean of brand **B**

$$\mu = \frac{\sum X}{N}$$

$$\mu = \frac{259}{7}$$

$$\mu = 37\ weeks$$

Still we can't decide which one brand bulb should be used?

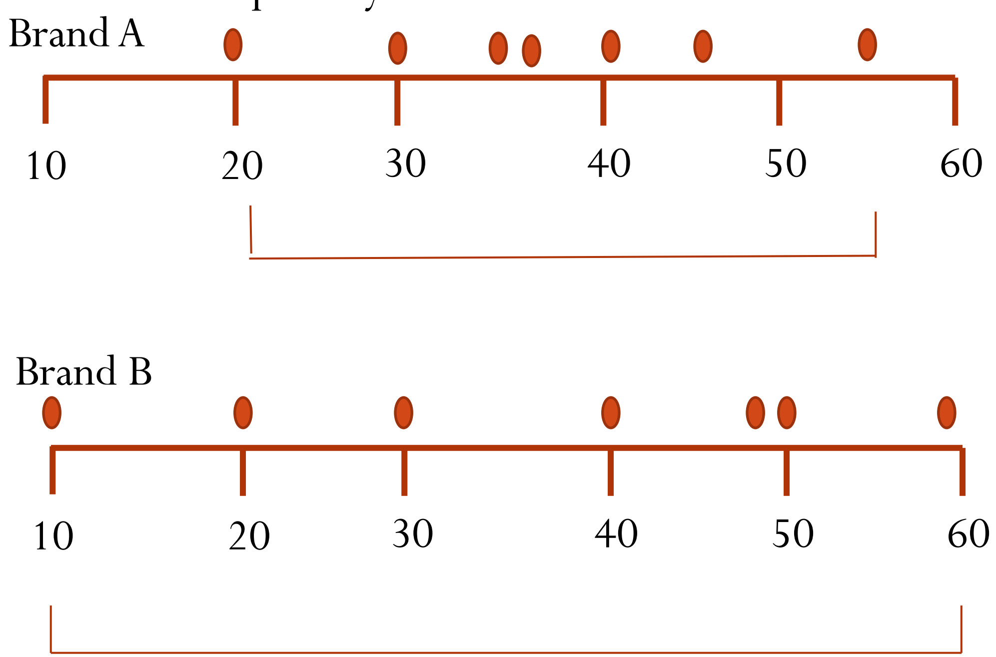
If I ask which one brand bulb you want to buy?

Example: Light Bulbs

- Two different brands of light bulbs are tested to see how long each will last before it fuse. Seven bulbs of each brand make a small population. The results (in weeks) are shown. Find the mean and range of each group.

Brand A	Brand B
20	10
34	20
30	30
35	40
45	50
40	49
55	60

Graphically examined the same data



Form graphical results Brand A is good, but the mean of both brands is same

Measure of Variations

- In order to describe the data set more accurately, statisticians use measure of variations.

- Range
- Variance
- Standard deviations

Use of the Variance and Standard Deviation

- To determine the spread of the data.
- To determine the consistency of a variable.
- Used in inferential statistics.

Comparison of the light bulbs

Brand A	Brand B
20	10
34	20
30	30
35	40
45	50
40	49
55	60

Range of brand A

R=highest-lowest

R=55-20

R=35

It concludes that 35 weeks separate the largest data value from the lowest data value

Range of brand B

R=highest-lowest

R=60-10

R=50

It concludes that 50 weeks separate the largest data value from the lowest data value

- To see the more meaningful statistic to measure the variability, statisticians use **variance** and **standard deviations**

- Population variance and standard deviations
- Sample variance and standard deviations

Measure of Variations

- How to find the population variance and population standard deviation.

1. Find the mean for the data $$\mu = \frac{\sum X}{N}$$
2. Find the deviation for each data value. $$X - \mu$$
3. Square each of the deviations $$(X - \mu)^2$$
4. Find the sum of the squares. $$\sum (X - \mu)^2$$
5. Divide by N to find the variance $$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$
6. Take the Square root of the variance to find the standard deviation.

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

- **Find the variance and standard deviation for the dataset for Brand A light bulb.**

Brand A	Brand B
20	10
34	20
30	30
35	40
45	50
40	49
55	60

20, 34, 30, 35, 45, 40, 55

1. Find the mean for the data

$$\mu = \frac{20+34+30+35+45+40+55}{7} = \frac{259}{7} \Rightarrow 37\ weeks$$

2. Subtract the mean from each data value $X - \mu$

$20 - 37 = -17$ $34 - 37 = -3$ $30 - 37 = -7$ $35 - 37 = -2$

$45 - 37 = 8$ $40 - 37 = 3$ $55 - 37 = 18$

3. Square each of the deviations $(X - \mu)^2$

$(-17)^2 = 289$ $(-3)^2 = 9$ $(-7)^2 = 49$ $(-2)^2 = 4$

$(-8)^2 = 64$ $(-3)^2 = 9$ $(-18)^2 = 324$

Measure of Variations

➢ **Find the variance and standard deviation for the dataset for Brand A light bulb.**

4. Find the sum of the squares. $\sum(X-\mu)^2$

$$289+9+49+4+64+9+324=748$$

5. Divide by N to find the variance $\sigma^2=\frac{\sum(X-\mu)^2}{N}$

$$=\frac{748}{7} \quad \Rightarrow 106.857$$

6. Take the Square root of the variance to find the standard deviation.

$$\sqrt{106.857} \approx 10.38 \qquad \sigma=\sqrt{\frac{\sum(X-\mu)^2}{N}}$$

Measure of Variations

➢ **Find the variance and standard deviation for the dataset for Brand B light bulb.**

Brand A	Brand B
20	10
34	20
30	30
35	40
45	50
40	49
55	60

10, 20, 30, 40, 50, 49, 60

1. Find the mean for the data

$$\mu = \frac{10+20+30+40+50+49+60}{7} = \frac{259}{7} \Rightarrow 37\ weeks$$

2. Subtract the mean from each data value $X - \mu$

$10 - 37 = -27$ $\quad 20 - 37 = -17$ $\quad 30 - 37 = -7$ $\quad 40 - 37 = 3$

$50 - 37 = 13$ $\quad 49 - 37 = 12$ $\quad 60 - 37 = 23$

3. Square each of the deviations $(X - \mu)^2$

$(-27)^2 = 729$ $\quad (-17)^2 = 289$ $\quad (-7)^2 = 49$ $\quad (3)^2 = 9$

$(13)^2 = 169$ $\quad (12)^2 = 144$ $\quad (23)^2 = 529$

Measure of Variations

- **Find the variance and standard deviation for the dataset for Brand A light bulb.**

4. Find the sum of the squares. $\sum(X-\mu)^2$

$$729 + 289 + 49 + 9 + 169 + 144 + 529 = 1918$$

5. Divide by N to find the variance $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$

$$= \frac{1918}{7} \Rightarrow 274$$

6. Take the Square root of the variance to find the standard deviation. $\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$

$$\sqrt{274} \approx 16.55$$

- **Conclusions:**
- As the standard deviation of brand A is 10.38, and for brand B is 16.55. The data of brand B is more variable compare to brand A.
- When the means are equal, the larger the variance or standard deviation is, the more variable the data.

Measure of Variations

➢ How to find the sample variance and sample standard deviation.

Where
X= individual value,
$\bar{X}$ *= Sample mean*
n= Sample size

1. Find the mean for the data $\bar{X} = \frac{\sum X}{n}$

2. Find the deviation for each data value. $X - \bar{X}$

3. Square each of the deviations $(X - \bar{X})^2$

4. Find the sum of the squares. $\sum(X - \bar{X})^2$

5. Divide by N to find the variance $S^2 = \frac{\sum(X - \bar{X})^2}{n-1}$

6. Take the Square root of the variance to find the standard deviation.

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$$

Example: A sample of 6 taxi drivers shows the time they spent in school zone in rush hours. Find the variance and standard deviation of the sample. 15, 22, 30, 10, 12, 25

Solution:

Summary

- We discussed Measures of Central Tendency
  - Midrange
  - Weighted Mean

- We discussed Measures of Position
  - Percentiles
  - Quartile
  - Deciles

- We discussed measure of Variation
  - Range
  - Variance
  - Standard deviations.

Thank You !