

DS-13/22 EU Session-2

DS-13 EU Statistics 2

Training Clarusway

Pear Deck - November 23, 2022 at 11:53AM

Part 1 - Summary

Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1



Use this space to take notes:

Slide 2

Course Info

Lesson Plan

STATISTICS BASICS



The goal of this course is to provide a comprehensive overview of the basics of statistics you will need to start your data science journey.

Custodian : Jason Acad.Coord. (lesson@clarusway.com)

In-class Sessions : 7 In-classes / 21 hours (Part-1 • 3 In-classes | Part-2 • 4 In-classes)

Lab Sessions : 3 Labs / 3 hours (Part-1 • 1 Lab | Part-2 • 2 Labs)

Certification Requirements:

1. Attend at least 70% of in-class sessions (at least 5 sessions of attendance)
2. Successfully complete and submit assignments (at least 2 assignments)

CLARUSWAY®
WAY TO REINVENT YOURSELF



Use this space to take notes:

Slide 3

Your Response

Able to finish the Central Tendency & Dispersion pre-class activity?



Students, drag the icon!

Pear Deck

Pear Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 4

Table of Contents

- ▶ Central Tendency (Measure of Centre)
 - ▷ Mean
 - ▷ Median
 - ▷ Mode
- ▶ Dispersion (Measure of Spread)
 - ▷ Range
 - ▷ Interquartile Range (IQR)
 - ▷ Standard Deviation (Variance)
 - ▷ Box Plot
- ▶ Practice with Python

CLARUSWAY®
WAY TO REINVENT YOURSELF.



Use this space to take notes:

Slide 5



1

Central Tendency (Measure of Centre)

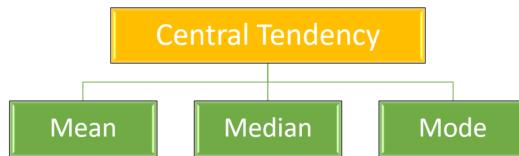
CLARUSWAY®
WAY TO REINVENT YOURSELF.

Use this space to take notes:

Slide 6

► Central Tendency (Measure of Centre) ➤

The central tendency concept is that one single value can best describe the data.



CLARUSWAY®
way to knowledge revealed

8

Use this space to take notes:

Slide 7

► Mean ➤

The mean is equal to the sum of the values in the dataset divided by the number of values.

1 Find the sum of all values in a group of values

2 Divide the sum by the number of values in the group.

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

CLARUSWAY®
way to knowledge revealed

9

Use this space to take notes:

Slide 8

► Mean Notation

Population Mean

ΣX = Sum of all X values
 N = Number of X values
 μ = Population mean

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

CLARUSWAY®
way to knowledge revealed



8

Use this space to take notes:

Slide 9

► Mean Notation

Population Mean

ΣX = Sum of all X values
 N = Number of X values
 μ = Population mean

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

CLARUSWAY®
way to knowledge revealed



Sample Mean

Σx = Sum of all x values
 n = Number of x values
 \bar{x} = Sample mean

$$\bar{x} = \frac{\Sigma x}{n}$$

9

Use this space to take notes:

Slide 10

► Mean Example



Compute the mean age.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |

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

CLARUSWAY®
way to knowledge revealed

10

Use this space to take notes:

Slide 11

► Mean Example



Compute the mean age.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |

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

$$\sum X = 49 + 64 + 36 + 64 + 47 = 260$$

$$N = 5$$

$$\mu = \frac{260}{5} = 52$$

CLARUSWAY®
way to knowledge revealed

11

Use this space to take notes:

Slide 12

► Median



The median is the middle score for a dataset that has been sorted from small to large.

1 List scores from smallest to largest

2 With an odd number of scores, the median is the middle score.

3 With an even number of scores, the median is the sum of the middle two scores divided by 2.

CLARUSWAY®
way to remember everything

12

Use this space to take notes:

Slide 13

► Median Example 1



Find the median age, given an odd number of scores.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |

1. List scores in ascending order.

36 47 49 64 64

13

Use this space to take notes:

Slide 14

► Median Example 1

Find the median age, given an odd number of scores.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |

1. List scores in ascending order.

36 47 49 64 64

2. With an odd number of scores, the median is the middle score.

36 47 49 64 64



Use this space to take notes:

CLARUSWAY®
way to remember everything

14

Slide 15

► Median Example 2

Find the median age, given an even number of scores.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| Martha L. Fox | 47 |

1. List scores in ascending order.

36 47 49 64

15

Use this space to take notes:

CLARUSWAY®
way to remember everything

Slide 16

► Median Example 2



Find the median age, given an even number of scores.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| Martha L. Fox | 47 |

1. List scores in ascending order.

36 47 49 64

2. With an even number of scores, the median is the sum of the middle two scores divided by 2.

$$\text{Median} = \frac{47 + 49}{2} = 48$$

CLARUSWAY®
way to education excellence

18

Use this space to take notes:

Slide 17

► Mean vs. Median



- The median is better if a small set of scores has an outlier.
- The mean is better if a large set of scores does not have an outlier.

CLARUSWAY®
way to education excellence

19

Use this space to take notes:

Slide 18

► Selection of Imputation Method for Missing Values

| | col1 | col2 | col3 | col4 | col5 |
|---|------|------|------|------|------|
| 0 | 2 | 5.0 | 3.0 | 6 | NaN |
| 1 | 9 | NaN | 9.0 | 0 | 7.0 |
| 2 | 19 | 17.0 | NaN | 9 | NaN |

→

| | col1 | col2 | col3 | col4 | col5 |
|---|------|------|------|------|------|
| 0 | 2.0 | 5.0 | 3.0 | 6.0 | 7.0 |
| 1 | 9.0 | 11.0 | 9.0 | 0.0 | 7.0 |
| 2 | 19.0 | 17.0 | 6.0 | 9.0 | 7.0 |

CLARUSWAY®
way to knowledge evolution

18

Use this space to take notes:

Slide 19

► Selection of Imputation Method for Missing Values

| | First Name | Gender | Salary | Bonus % | Senior Management | Team |
|----|------------|--------|----------|---------|-------------------|----------------------|
| 0 | Douglas | Male | 97308.0 | 6.945 | True | Marketing |
| 1 | Thomas | Male | 61933.0 | NaN | True | NaN |
| 2 | Jerry | Male | NaN | 9.340 | True | Finance |
| 3 | Dennis | n.a. | 115163.0 | 10.125 | False | Legal |
| 4 | NaN | Female | NaN | 11.598 | NaN | Finance |
| 5 | Angela | NaN | NaN | 18.523 | True | Engineering |
| 6 | Shawn | Male | 111737.0 | 6.414 | False | NaN |
| 7 | Rachel | Female | 142032.0 | 12.599 | False | Business Development |
| 8 | Linda | Female | 57427.0 | 9.557 | True | Client Services |
| 9 | Stephanie | Female | 36844.0 | 5.574 | True | Business Development |
| 10 | NaN | NaN | NaN | NaN | NaN | NaN |

CLARUSWAY®
way to knowledge evolution

19

Use this space to take notes:

Slide 20

► Mean vs. Median Example



Find the mean and the median of car prices.



CLARUSWAY®
way to knowledge revealed

20

Use this space to take notes:

Slide 21

► Mean vs. Median Example



Mean:

$$\mu = \frac{\sum X}{N}$$
$$\mu = \frac{\$4000 + \$15000 + \$20000 + \$33000 + \$1800000}{5}$$
$$\mu = \frac{\$1872000}{5} = \$374400$$



Median:

\$20000

21

Use this space to take notes:

Slide 22

Your Response

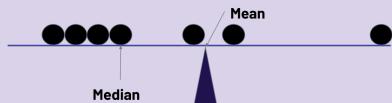
Slide 22

Your Response

Let's Practice



_____ is resistant to outliers.



Students choose an option

Peer Deck Interactive Slide

Do not remove this bar

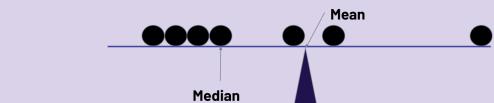
Use this space to take notes:

Slide 23

Let's Practice Answer



Median is resistant to outliers.



CLARUSWAY®

23

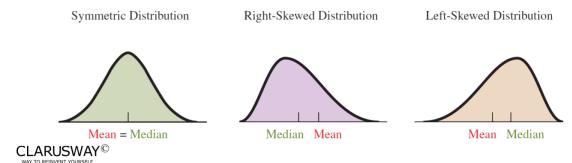
Use this space to take notes:

Slide 24

► Mean vs. Median

Generally, if the shape is

- ▶ Perfectly symmetric, the mean equals the median.
- ▶ Skewed to the right, the mean is larger than the median.
- ▶ Skewed to the left, the mean is smaller than the median.



24

Use this space to take notes:

Slide 25

Your Response

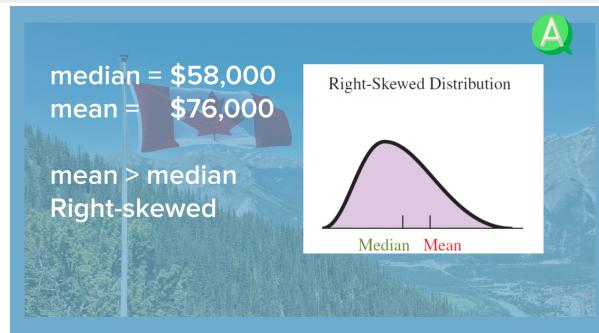
According to Statistics Canada, in 2004 the median household income in Canada was \$58,000 and the mean was \$76,000. What would you predict about the shape of the distribution?

Students, write your response!

Pear Deck Interactive Slide
Do not reproduce this slide

Use this space to take notes:

Slide 26



Use this space to take notes:

Slide 27

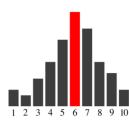
► Mode



The mode is the most frequent score in a dataset.

1 List scores from smallest to largest.

2 Count how many of each number. A number that appears most often is the mode.



CLARUSWAY®
www.clarusway.com

27

Use this space to take notes:

Slide 28

► Mode Example



Find the mode.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |
| Tim Berners-Lee | 64 |

CLARUSWAY®
way to remember everything

28

1. List scores from smallest to largest.

36 47 49 64 64 64

Use this space to take notes:

Slide 29

► Mode Example



Find the mode.

| Name | Age |
|-----------------|-----|
| Elon Musk | 49 |
| Bill Gates | 64 |
| Mark Zuckerberg | 36 |
| G. van Rossum | 64 |
| Martha L. Fox | 47 |
| Tim Berners-Lee | 64 |

CLARUSWAY®
way to remember everything

29

1. List scores from smallest to largest.

36 47 49 64 64 64

2. Count how many of each number.

| | | | |
|----|----|----|----|
| 1 | 1 | 1 | 3 |
| 36 | 47 | 49 | 64 |
| | | 64 | 64 |
| | | 64 | |

Mode

Use this space to take notes:

Slide 30

► Pro's and Con's of Mode



Advantages:

- The mode is easy to understand and calculate.
- The mode is not affected by extreme values.
- The mode is useful for categorical data.

Disadvantages:

- The mode is not defined when there are no repeats in a data set.
- The mode is not based on all values.
- The mode is unstable when the data consist of a small number of values.
- Sometimes data have one mode, more than one mode, or no mode at all.

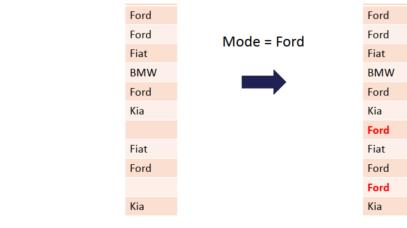
CLARUSWAY®
way to knowledge revealed

30

Use this space to take notes:

Slide 31

► Selection of Imputation Method for Missing Values



CLARUSWAY®
way to knowledge revealed

31

Use this space to take notes:

Slide 32

► Selection of Imputation Method for Missing Values



| | id | A | B | C | D |
|---|-----------|----------|----------|----------|----------|
| 0 | 0 | 10.0 | A | NaN | NaN |
| 1 | 1 | 9.0 | B | BB | 20.0 |
| 2 | 2 | 8.0 | A | CC | 18.0 |
| 3 | 3 | 7.0 | A | BB | 22.0 |
| 4 | 4 | NaN | NaN | BB | 18.0 |
| 5 | 5 | NaN | B | CC | 17.0 |
| 6 | 6 | 20.0 | A | AA | 19.0 |
| 7 | 7 | 15.0 | B | BB | NaN |
| 8 | 8 | 12.0 | NaN | NaN | 17.0 |
| 9 | 9 | NaN | A | AA | 23.0 |

CLARUSWAY®
WAY TO REINVENT YOURSELF

32

Use this space to take notes:

Slide 33



2

Dispersion (Measure of Spread)

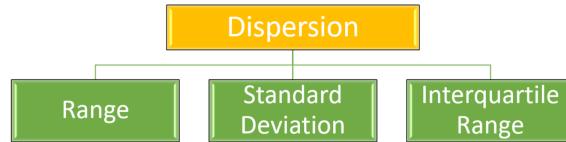
CLARUSWAY®
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 34

► Dispersion (Measure of Spread) ➤

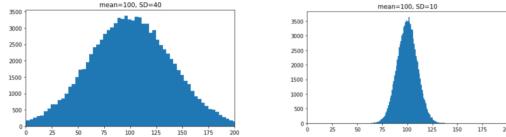
The most common measures of variability are the range, the interquartile range (IQR), variance, and standard deviation.



Use this space to take notes:

Slide 35

► Dispersion vs Central Tendency ➤



Means are the same but standard deviations are quite different because distributions are different.
Dispersion gives idea about distributions.

Use this space to take notes:

Slide 36

► Range



The range is the difference between the largest and smallest values in a set of values.

Example:

(2) 4 (9) 5 7 3

$$\text{Range} = \text{Largest} - \text{Smallest} = 9 - 2 = 7$$

CLARUSWAY®
way to knowledge revealed

38

Use this space to take notes:

Slide 37

► The Interquartile Range (IQR)



The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles.

- ▶ Quartiles divide a rank-ordered data set into four equal parts.
- ▶ The values that divide each part are called the first, second, and third quartiles.
- ▶ First, second, and third quartiles are denoted by Q1, Q2, and Q3, respectively.

CLARUSWAY®
way to knowledge revealed

39

Use this space to take notes:

Slide 38

► The Interquartile Range (IQR)

The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles.

- ▶ Q2 is the median of the entire data set.
- ▶ Q1 is the median of the data below Q2.
- ▶ Q3 is the median of the data above Q2.

$$\text{Interquartile Range} = Q3 - Q1$$

CLARUSWAY®
way to knowledge revealed

38

Use this space to take notes:

Slide 39

► IQR Example

Ordered data set.



CLARUSWAY®
way to knowledge revealed

39

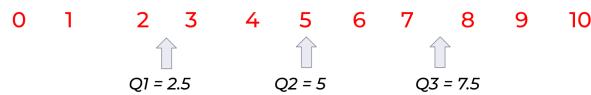
Use this space to take notes:

Slide 40

► IQR Example



Ordered data set.



$$\text{Interquartile Range} = 7.5 - 2.5 \\ IQR = 5$$

CLARUSWAY®
www.clarusway.com



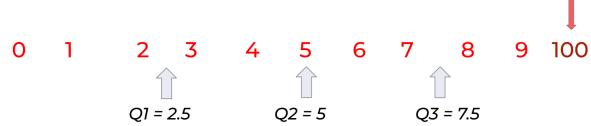
Use this space to take notes:

Slide 41

► IQR Example



Ordered data set. Add an outlier instead of 10.



$$\text{Interquartile Range} = 7.5 - 2.5 \\ IQR = 5$$

CLARUSWAY®
www.clarusway.com



Use this space to take notes:

Slide 42

Your Response

Slide 42

Your Response

► IQR Practice



Question:

What is the interquartile range of these numbers?

8 9 10 10 12 13 14 15 16



Peer Deck Interactive Slide

You can't copy this slide.

Use this space to take notes:

Slide 43

► IQR Practice

Answer



Question:

What is the interquartile range of these numbers?

8 9 (10) 10 12 13 (14) 15 16

$$\begin{aligned} \text{IQR} &= Q3 - Q1 \\ \text{IQR} &= 14 - 10 \\ \text{IQR} &= 4 \end{aligned}$$

CLARUSWAY®
WAY TO REVIVE YOURSELF

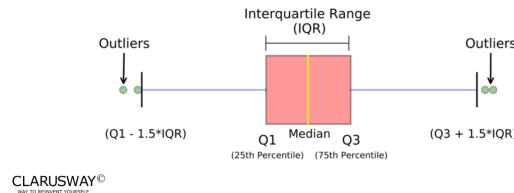
43

Use this space to take notes:

Slide 44

► 1.5xIQR Rule for Outliers ➤

1.5 IQR Rule: If an observation falls more than 1.5 IQRs above Q3 or below Q1, it is an outlier.



CLARUSWAY®
way to knowledge education

44

Use this space to take notes:

Slide 45

► Outlier Practice ➤



Question:

Are there any outliers?

8 9 (10) 10 12 13 (14) 15 22

$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ \text{IQR} &= 14 - 10 \\ \text{IQR} &= 4 \end{aligned}$$

CLARUSWAY®
way to knowledge education

45

Use this space to take notes:

Slide 46

► Variance (Population)

Variance is the average squared deviation from the mean.

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

element mean
 ↓
 N
 ↑
number of elements

CLARUSWAY®
way to knowledge revealed



48

Use this space to take notes:

Slide 47

► Variance (Sample)

Variance is the average squared deviation from the mean.

$$\text{sample variance} \longrightarrow S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

observation mean
 ↓
 n - 1
 ↑
number of observations

CLARUSWAY®
way to knowledge revealed



49

Use this space to take notes:

Slide 48

► Variance Example



Find the Variance.

0 1 5 6

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

CLARUSWAY®
way to knowledge™

48

Use this space to take notes:

Slide 49

► Variance Example



Find the Variance.

0 1 5 6

Mean: $\mu = \frac{\sum X}{N} = \frac{0+1+5+6}{4} = \frac{12}{4} = 3$

Dev Sum of Squares: $SS = \sum(X - \mu)^2$
 $SS = (0 - 3)^2 + (1 - 3)^2 + (5 - 3)^2 + (6 - 3)^2$

$SS = 9 + 4 + 4 + 9 = 26$

Variance: $\sigma^2 = \frac{\sum(X - \mu)^2}{N}$
 $\sigma^2 = \frac{26}{4} = 6.5$

CLARUSWAY®
way to knowledge™

49

Use this space to take notes:

Slide 50

► Standard Deviation

Standard deviation is the square root of the variance.

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

element mean
number of elements

CLARUSWAY®
WAY TO RECOGNIZE YOURSELF

50

Use this space to take notes:

Slide 51

► Standard Deviation Example

Students in a class were asked on a questionnaire at the beginning of the course,
"How many children do you think is ideal for a family?"
The observations, classified by student's gender, were

Men : 0 0 0 2 4 4 4
 Women : 0 2 2 2 2 2 4

Both men and women have a mean of 2 and a range of 4.

Do the distributions of data have the same amount of variability around the mean?

CLARUSWAY®
WAY TO REINVENT YOURSELF

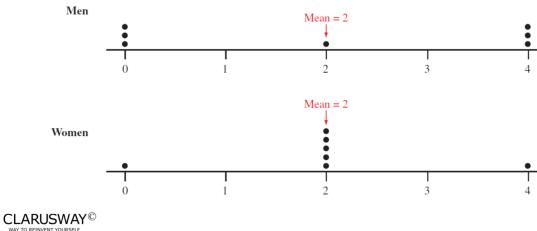
51

Use this space to take notes:

Slide 52

► Standard Deviation Example ➤

Men : 0 0 0 2 4 4 4
Women : 0 2 2 2 2 2 4



CLARUSWAY®
way to measure intelligence

52

Use this space to take notes:

Slide 53

► Standard Deviation Example ➤

Men : 0 0 0 2 4 4 4
Women : 0 2 2 2 2 2 4

$$\text{Men: } s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{24}{6}} = \sqrt{4} = 2.0$$

$$\text{Women: } s = 1.2$$

The observations for males tend to be farther from the mean than those for females, as indicated by $s = 2.0 > s = 1.2$. In summary, the men's observations vary more around the mean.

CLARUSWAY®
way to measure intelligence

53

Use this space to take notes:

Slide 54



▶ Box Plot



CLARUSWAY®
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 55

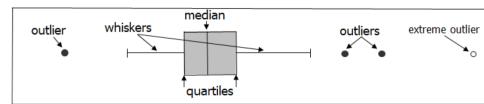
▶ What is Box Plot



A **box plot** is a method for graphically depicting groups of numerical data through their *quartiles*.

A box plot generally shows

- ★ median
- ★ 1st quartile (Q1)
- ★ 3rd quartile (Q3)
- ★ outliers



CLARUSWAY®
way to reinvent yourself

55

Use this space to take notes:

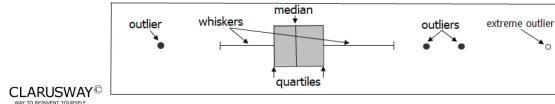
Slide 56

► Box Plot

Boxplots show distribution in one dimension

- ★ Only useful for continuous variables
- ★ Good for comparing distributions of a continuous variable between categorical groups

Box plots are also known as **box and whisker plots.**



Use this space to take notes:

Slide 57

► How to make a Box Plot (Min & Max)

Step 1: Order the data from smallest to largest.

25 28 29 29 30 34 35 35 37 38

| Weight, kg |
|------------|
| 38 |
| 25 |
| 37 |
| 28 |
| 35 |
| 29 |
| 35 |
| 29 |
| 34 |
| 30 |

CLARUSWAY®
www.clarusway.com

57

Use this space to take notes:

Slide 58

► How to make a Box Plot (Min & Max) ➤

| Weight, kg |
|------------|
| 38 |
| 25 |
| 37 |
| 28 |
| 35 |
| 29 |
| 35 |
| 29 |
| 34 |
| 30 |

CLARUSWAY®
www.clarusway.com

Step 1: Order the data from smallest to largest.

25 28 29 29 30 34 35 35 37 38

Step 2: Find the median.

25 28 29 29 30 34 35 35 37 38

Median = 32

58

Use this space to take notes:

Slide 59

► How to make a Box Plot (Min & Max) ➤

| Weight, kg |
|------------|
| 38 |
| 25 |
| 37 |
| 28 |
| 35 |
| 29 |
| 35 |
| 29 |
| 34 |
| 30 |

CLARUSWAY®
www.clarusway.com

Step 1: Order the data from smallest to largest.

25 28 29 29 30 34 35 35 37 38

Step 2: Find the median.

25 28 29 29 30 34 35 35 37 38

Median = 32

Step 3: Find the quartiles.

25 28 29 29 30 34 35 35 37 38

Q1 = 29

Q3 = 35

59

Use this space to take notes:

Slide 60

► How to make a Box Plot (Min & Max) ➤

| Weight, kg |
|------------|
| 38 |
| 25 |
| 37 |
| 28 |
| 35 |
| 29 |
| 35 |
| 29 |
| 34 |
| 30 |

CLARUSWAY®
way to measure everything

Step 1: Order the data from smallest to largest.

25 28 29 29 30 34 35 35 37 38

Step 2: Find the median.

25 28 29 29 30 34 35 35 37 38

Median = 32

Step 3: Find the quartiles.

25 28 29 29 30 34 35 35 37 38

Q1 = 29 Q2 = 35

Step 4: Find the min and the max.

Min = 25 Max = 38

60

Use this space to take notes:

Slide 61

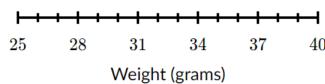
► How to make a Box Plot (Min & Max) ➤

| | |
|---------------|----|
| Min | 25 |
| Q1 | 29 |
| Median | 32 |
| Q3 | 35 |
| Max | 38 |

Step 1: Scale / label an axis that fits the five-number.

25 28 29 29 30 34 35 35 37 38

Min = 25 Max = 38



61

Use this space to take notes:

Slide 62

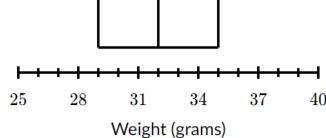
► How to make a Box Plot (Min & Max) ➤

| | |
|---------------|----|
| Min | 25 |
| Q1 | 29 |
| Median | 32 |
| Q3 | 35 |
| Max | 38 |

Step 2: Draw a box from Q_1 to Q_3 with a vertical line through the *median*.

25 28 29 29 30 34 35 35 37 38

Q_1 median Q_3



CLARUSWAY®
way to revolutionize education

62

Use this space to take notes:

Slide 63

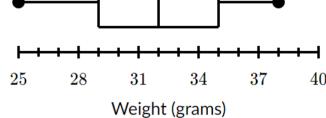
► How to make a Box Plot (Min & Max) ➤

| | |
|---------------|----|
| Min | 25 |
| Q1 | 29 |
| Median | 32 |
| Q3 | 35 |
| Max | 38 |

Step 3: Draw a whisker from Q_1 to the *min* and from Q_3 to the *max*.

25 28 29 29 30 34 35 35 37 38

min Q_1 median Q_3 max



CLARUSWAY®
way to revolutionize education

63

Use this space to take notes:

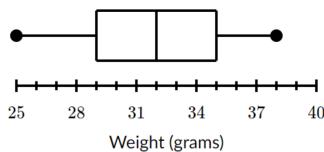
Slide 64

► How to make a Box Plot (Min & Max) ➤

| | |
|---------------|----|
| Min | 25 |
| Q1 | 29 |
| Median | 32 |
| Q3 | 35 |
| Max | 38 |

We don't need the labels on the final product:

25 28 29 29 30 34 35 35 37 38



CLARUSWAY®
WAY TO KNOW™ WORKED

64

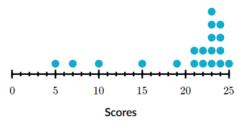
Use this space to take notes:

Slide 65

Your Response

► Identifying Outliers ➤

An **outlier** is a data point that lies outside the overall pattern in a distribution.



Question:

How many outliers do you see?



Students, enter a number!

Pear Deck Interactive Slides

64

Use this space to take notes:

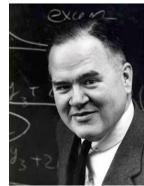
Slide 66

▶ Identifying Outliers (1.5xIQR Rule)

1.5 IQR Rule: If an observation falls more than 1.5 IQRs above Q3 or below Q1, it is an outlier.

According to John Tukey,
1 IQR seemed like too little and
2 IQRs seemed like too much.

CLARUSWAY®
way to knowledge revealed

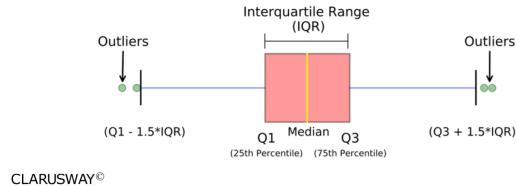


Use this space to take notes:

Slide 67

▶ Boxplot with 1.5xIQR Rule

1.5 IQR Rule: If an observation falls more than 1.5 IQRs above Q3 or below Q1, it is an outlier.



CLARUSWAY®
way to knowledge revealed



Use this space to take notes:

Slide 68

► How to make a Box Plot (1.5xIQR Rule) ➤

5, 7, 10, 15, 19, 21, 21, 22, 22, 23, 23, 23, 23, 23, 24, 24, 24, 24, 24, 25

Step 1: Find the median, quartiles, and interquartile range.

$$\text{Median} = 23$$

$$Q1 = 20$$

$$Q3 = 23.5$$

$$IQR = Q3 - Q1 = 23.5 - 20 = 3.5$$

```
[33] np.median(a)
```

23.0

```
[34] np.percentile(a, 25)
```

20.0

```
[35] np.percentile(a, 75)
```

23.5

CLARUSWAY®
way to knowledge revealed

68

Use this space to take notes:

Slide 69

► How to make a Box Plot (1.5xIQR Rule) ➤

5, 7, 10, 15, 19, 21, 21, 22, 22, 23, 23, 23, 23, 23, 24, 24, 24, 24, 25

Step 2: Calculate $1.5 \times IQR$ below the first quartile and check for low outliers.

$$Q1 - 1.5 \times IQR = 20 - (1.5 \times 3.5) \\ = 14.75$$

Low Outliers: 5 7 10

CLARUSWAY®
way to knowledge revealed

69

Use this space to take notes:

Slide 70

► How to make a Box Plot (1.5xIQR Rule) ➤

5, 7, 10, 15, 19, 21, 21, 22, 22, 23, 23, 23, 23, 23, 23, 24, 24, 24, 24, 24, 25

Step 3: Calculate $1.5 \times IQR = 23.5 + (1.5 \times 5)$
= 28.75

High Outliers: None

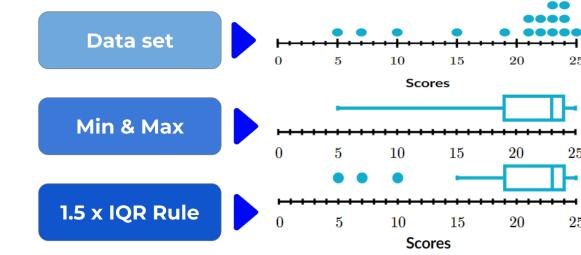
CLARUSWAY®
way to revision success

70

Use this space to take notes:

Slide 71

► Box Plot Overview ➤



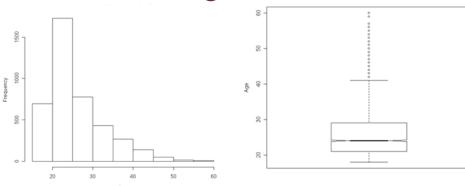
CLARUSWAY®
way to revision success

71

Use this space to take notes:

Slide 72

► Box Plot vs. Histogram



- Histogram shows distribution of the data in two dimensions – the boxplot is in one dimension
 - Histogram shows frequency of observations within ranges
 - Boxplot only shows summary statistics

CLARUSWAY®
way to knowledge forward



72

Use this space to take notes:

Slide 73

Your Response

► Let's Practice Draw Box Plot



| | | | | | | | | |
|---|---|---|---|---|----|----|----|----|
| 3 | 6 | 8 | 9 | 9 | 10 | 12 | 14 | 19 |
|---|---|---|---|---|----|----|----|----|



Students, draw anywhere on this slide!

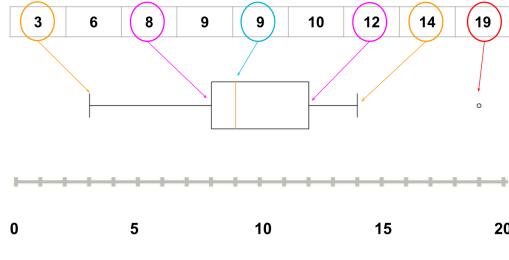
Pear Deck Interactive Slides

72

Use this space to take notes:

Slide 74

► Let's Practice Draw Box Plot



CLARUSWAY®
Way to REVIEW knowledge

74

Use this space to take notes:

Slide 75

Your Response

How well did you like this lesson?



Students, drag the icon!

Pear Deck

Pear Deck Interactive Slide
Do not review this slide

Use this space to take notes:

THANKS!

Any questions?

You can find us at:

- ▶ richard@clarusway.com
- ▶ jason@clarusway.com

CLARUSWAY®
WAY TO REINVENT YOURSELF



78

Use this space to take notes: