

# DS-13/22 EU Session-3

DS-13 EU Statistics 3

Training Clarusway

Pear Deck - November 24, 2022 at 11:53AM

## Part 1 - Summary

Use this space to summarize your thoughts on the lesson

## Part 2 - Responses

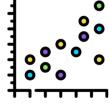
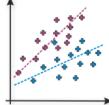
Slide 1



Use this space to take notes:

Slide 2	Your Response
<p>Did you finish Statistics (Covariance&amp;Correlation, Normal Distribution) pre-class activity?</p>  <p>Students choose an option</p> <p>Pear Deck Interactive Slide Do not remove this bar</p>	<p>You Chose</p> <ul style="list-style-type: none"> <li>• I finished partially.</li> </ul> <p>Other Choices</p> <ul style="list-style-type: none"> <li>• I finished completely.</li> <li>• No, I didn't finish.</li> </ul>

Use this space to take notes:

Slide 3
<h2>Table of Contents</h2>   <ul style="list-style-type: none"> <li>▶ Scatter Plot</li> <li>▶ Covariance</li> <li>▶ Correlation</li> <li>▶ Normal Distribution</li> </ul> <p>CLARUSWAY® WAY TO REINVENT YOURSELF</p>

Use this space to take notes:

## Slide 4



### Scatter Plot



CLARUSWAY®  
WAY TO REINVENT YOURSELF

Use this space to take notes:

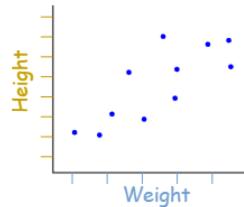
## Slide 5

### ▶ What is Scatter Plot



A **scatter plot** shows the relationship between two numerical variables.

CLARUSWAY®  
way to reinvent yourself

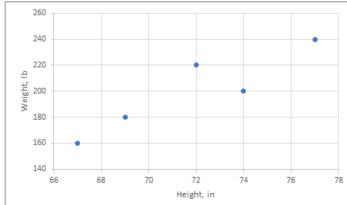


Use this space to take notes:

## Slide 6

### ▶ How to make a Scatter Plot

Height, in	Weight, lb
67	160
72	220
77	240
74	200
69	180



Scatter plots reveal patterns in bivariate data.

CLARUSWAY®  
way to knowledge revealed



Use this space to take notes:

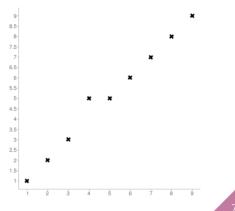
## Slide 7

### ▶ Patterns of Data in Scatter Plot

**Linearity**    **Slope**    **Strength**    **Unusual Features**

Linearity refers to whether a data pattern is linear (straight) or nonlinear (curved).

Linear



CLARUSWAY®  
way to knowledge revealed



Use this space to take notes:

## Slide 8

### ▶ Patterns of Data in Scatter Plot ➤

Linearity

Slope

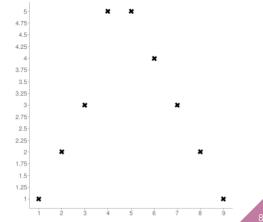
Strength

Unusual Features

Linearity refers to whether a data pattern is linear (straight) or nonlinear (curved).

Nonlinear

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 9

### ▶ Patterns of Data in Scatter Plot ➤

Linearity

Slope

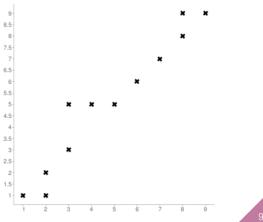
Strength

Unusual Features

Slope refers to the direction of change in variable Y when variable X gets bigger.

Positive slope

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 10

### ► Patterns of Data in Scatter Plot ➤

Linearity

Slope

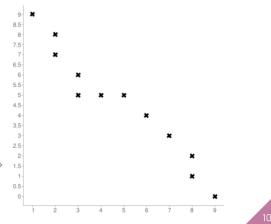
Strength

Unusual Features

Slope refers to the direction of change in variable Y when variable X gets bigger.

Negative slope

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 11

### ► Patterns of Data in Scatter Plot ➤

Linearity

Slope

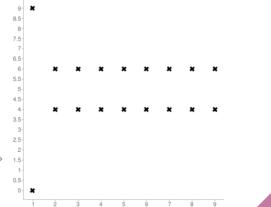
Strength

Unusual Features

Slope refers to the direction of change in variable Y when variable X gets bigger.

Zero slope

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 12

### ► Patterns of Data in Scatter Plot ➤

Linearity

Slope

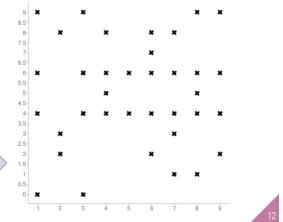
Strength

Unusual Features

Strength refers to the degree of "scatter" in the plot.

Weak relationship

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 13

### ► Patterns of Data in Scatter Plot ➤

Linearity

Slope

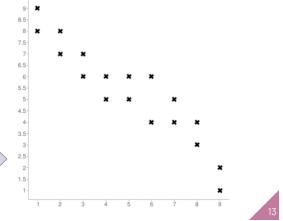
Strength

Unusual Features

Strength refers to the degree of "scatter" in the plot.

Strong relationship

CLARUSWAY®  
way to knowledge education



Use this space to take notes:

## Slide 14

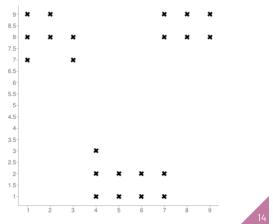
### ► Patterns of Data in Scatter Plot ➤

Linearity      Slope      Strength      Unusual Features

Unusual Features such as  
Clusters, Gaps, and Outliers

Clusters

CLARUSWAY®  
way to knowledge revealed



14

Use this space to take notes:

## Slide 15

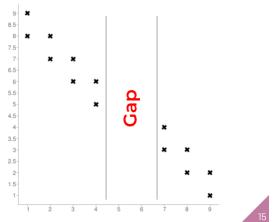
### ► Patterns of Data in Scatter Plot ➤

Linearity      Slope      Strength      Unusual Features

Unusual Features such as  
Clusters, Gaps, and Outliers

Gap

CLARUSWAY®  
way to knowledge revealed



15

Use this space to take notes:

## Slide 16

### ► Patterns of Data in Scatter Plot ➤

Linearity

Slope

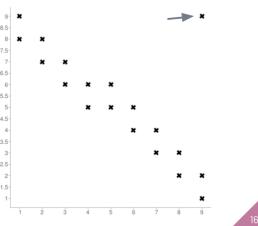
Strength

Unusual Features

Unusual Features such as Clusters, Gaps, and Outliers

Outlier

CLARUSWAY®  
way to knowledge forward



Use this space to take notes:

## Slide 17

### Your Response

### ► Let's Practice ?

Draw lines or circles.

**Linearity**  
Linear  
Nonlinear

**Strength**  
Weak  
Strong

**Slope**  
Positive  
Negative  
Zero

**Unusual F.**  
Clusters  
Gap  
Outlier  
None



Students, draw anywhere on this slide!

Pear Deck Interactive Slides

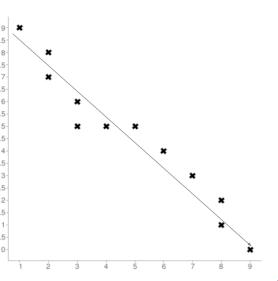
Use this space to take notes:

## Slide 18

### ► Let's Practice

- |                  |                   |
|------------------|-------------------|
| <b>Linearity</b> | <b>Strength</b>   |
| Linear           | Weak              |
| Nonlinear        | Strong            |
|                  |                   |
| <b>Slope</b>     | <b>Unusual F.</b> |
| Positive         | Clusters          |
| Negative         | Gap               |
| Zero             | Outlier           |
| None             |                   |

CLARUSWAY®  
way to knowledge™



18

Use this space to take notes:

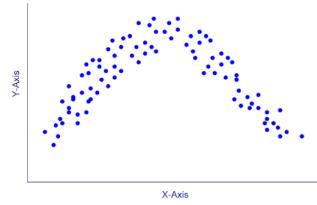
## Slide 19

### Your Response

### ► Let's Practice

Draw lines or circles.

- |                  |                   |
|------------------|-------------------|
| <b>Linearity</b> | <b>Strength</b>   |
| Linear           | Weak              |
| Nonlinear        | Strong            |
|                  |                   |
| <b>Slope</b>     | <b>Unusual F.</b> |
| Positive         | Clusters          |
| Negative         | Gap               |
| Zero             | Outlier           |
| None             |                   |



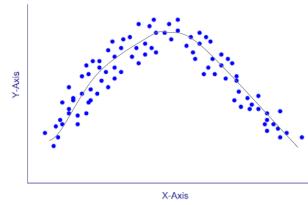
Pear Deck Interactive Slides  
The new way to teach.

Use this space to take notes:

## Slide 20

### Let's Practice

Linearity	Strength
Linear	Weak
Nonlinear	Strong
Slope	Unusual F.
Positive	Clusters
Negative	Gap
Zero	Outlier
	None



CLARUSWAY®  
way to review knowledge

20

Use this space to take notes:

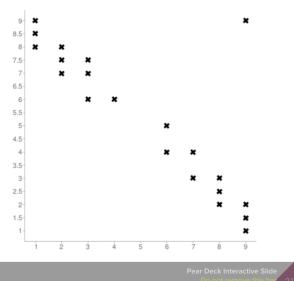
## Slide 21

### Your Response

### Let's Practice

Draw lines or circles.

Linearity	Strength
Linear	Weak
Nonlinear	Strong
Slope	Unusual F.
Positive	Clusters
Negative	Gap
Zero	Outlier
	None



Students, draw anywhere on this slide!

Pear Deck Interactive Slide

### Let's Practice

Draw lines or circles.

Linearity	Strength
Linear	Weak
Nonlinear	Strong
Slope	Unusual F.
Positive	Clusters
Negative	Gap
Zero	Outlier
	None



ISWANT, draw anywhere on this slide!

Pear Deck Interactive Slide

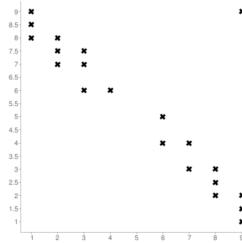
Use this space to take notes:

## Slide 22

### ► Let's Practice

- Linearity**
  - Linear
  - Nonlinear
- Strength**
  - Weak
  - Strong
- Slope**
  - Positive
  - Negative
  - Zero
- Unusual F.**
  - Clusters
  - Gap
  - Outlier
  - None

CLARUSWAY®  
WAY TO REINVENT YOURSELF

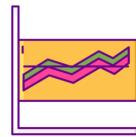


22

Use this space to take notes:

## Slide 23

### 2 ► Covariance



CLARUSWAY®  
WAY TO REINVENT YOURSELF

Use this space to take notes:

## Slide 24

### ► Intro

- ★ We are interested in how variables behave together.
- ★ For example, suppose we want to investigate the relationship between smoking and lung capacity.
- ★ We might ask a group of people about their smoking habits, and measure their lung capacities.

CLARUSWAY®  
way to healthier breathing



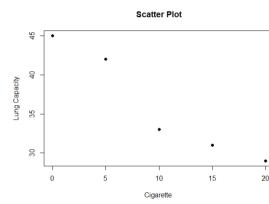
24

Use this space to take notes:

## Slide 25

### ► Scatter Plot

Cigarette (X)	Lung Capacity (Y)
0	45
5	42
10	33
15	31
20	29



Scatter plots reveal patterns in bivariate data.

CLARUSWAY®  
way to healthier breathing

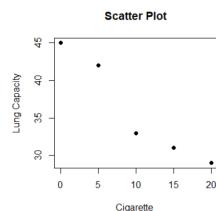
25

Use this space to take notes:

## Slide 26

### ▶ Scatter Plot

- We can see easily from the graph that as smoking goes up, lung capacity tends to go down.
- The two variables covary in opposite directions.



CLARUSWAY®  
way to healthier smoking



28

Use this space to take notes:

## Slide 27

### ▶ Covariance

- ★ **Covariance** is a measure of how much two random variables change together.
- ★ The covariance gives info on the strength of the linear relationship between two variables.
- ★ Since it is not a scaled metric, it is not useful to compare two different values of covariances.

$$\text{Cov}(x, y) = \sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{N}$$

CLARUSWAY®  
way to healthier smoking



29

Use this space to take notes:

## Slide 28

### ► Covariance Interpretation

x and y tend to move in the same direction

$$\text{Cov}(x,y) > 0$$

x and y tend to move in opposite directions

$$\text{Cov}(x,y) < 0$$

x and y are independent

$$\text{Cov}(x,y) = 0$$

CLARUSWAY®  
way to knowledge revealed

28

Use this space to take notes:

## Slide 29

### ► Covariance Example

Table shows years of education (S) and hourly earnings in dollars (Y).

Observation	S	Y	Observation	S	Y
1	15	17.24	11	17	15.38
2	16	15.00	12	12	12.70
3	8	14.91	13	12	26.00
4	6	4.50	14	9	7.50
5	15	18.00	15	15	5.00
6	12	6.29	16	12	21.63
7	12	19.23	17	16	12.10
8	18	18.69	18	12	5.55
9	12	7.21	19	12	7.50
10	20	42.06	20	14	8.00

CLARUSWAY®  
way to knowledge revealed

29

Use this space to take notes:

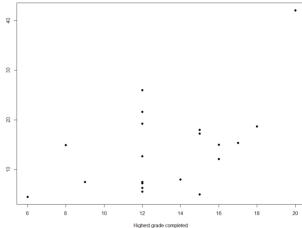
## Slide 30

### ► Covariance Example

- Figure shows the data plotted as a scatter plot.
- You can see that there is a weak positive association between the two variables.

CLARUSWAY®  
www.clarusway.com

Scatter Plot



30

Use this space to take notes:

## Slide 31

### ► Covariance Example

Observation	S	Y	(S - $\bar{S}$ )	(Y - $\bar{Y}$ )	(S - $\bar{S}$ )(Y - $\bar{Y}$ )
1	15	17.24	1.75	3.016	5.277
2	16	15.00	2.75	0.775	2.133
3	8	14.91	-5.25	0.685	-3.599
4	6	15.00	-3.25	-0.596	1.937
5	15	18.00	1.75	3.776	6.697
6	12	6.29	-1.25	-7.915	9.918
7	12	19.23	-1.25	5.006	-6.257
8	18	18.69	4.75	4.466	21.211
9	12	7.21	-1.25	-7.015	8.768
10	20	4.66	6.75	27.856	187.979
11	17	15.58	1.75	1.186	4.333
12	12	12.70	-1.25	-1.525	1.900
13	12	26.00	-1.25	11.776	-14.719
14	9	7.50	-4.25	-6.725	28.579
15	15	5.00	1.75	-9.225	-16.143
16	12	21.63	-1.25	7.406	-9.257
17	10	13.00	-3.25	-2.182	-5.872
18	12	5.55	-1.25	-8.675	10.843
19	12	7.50	-1.25	-6.725	8.406
20	14	8.00	0.75	-6.225	-4.668

CLARUSWAY®  
www.clarusway.com

$$\begin{aligned}\text{Cov}(x, y) &= \frac{1}{n} [(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})] \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})\end{aligned}$$

Therefore, covariance is calculated as

$$305.888 / 19 = 16.0993$$

31

Use this space to take notes:

## Slide 32



3

### Correlation



CLARUSWAY®  
WAY TO REINVENT YOURSELF

Use this space to take notes:

## Slide 33

### Correlation



The **correlation** summarizes the direction of the association between two quantitative variables and the strength of its linear (straight-line) trend.

#### Direction

- Positive : Move in same direction
- Negative : Move in opposite directions

#### Strength

- Weak : Widely spread
- Strong : Concentrated around a line

CLARUSWAY®  
way to reinvent yourself

33

Use this space to take notes:

## Slide 34

### ► Correlation Notation

Sample  
Correlation

$r$



Population  
Correlation

$\rho$  or  $R$

CLARUSWAY®  
way to knowledge revealed

34

Use this space to take notes:

## Slide 35

### ► Correlation Formula

Sample  
Correlation

$$r = \frac{\text{Cov}(x, y)}{s_x s_y}$$

Population  
Correlation

$$\rho = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

CLARUSWAY®  
way to knowledge revealed

35

Use this space to take notes:

## Slide 36

### ► Examples



38

Use this space to take notes:

## Slide 37

### ► Correlation Coefficient

The correlation coefficient,  $r$ , developed by Karl Pearson in the early 1900s.



39

Use this space to take notes:

## Slide 38

### ► Correlation Coefficient ➤

Denoted by  $r$ , it takes values between -1 and +1.

#### Strength

- Greater absolute value, stronger relationship
- Strongest relationship is correlation of -1 or 1
- Weakest relationship is correlation of zero.

#### Direction

- Sign of coefficient describes direction
- Positive sign: Variables move in same direction
- Negative sign: Move in opposite directions

CLARUSWAY®  
way to knowledge revealed

38

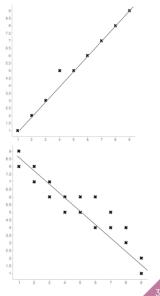
Use this space to take notes:

## Slide 39

### ► Linear Relationship ➤

When the data points follow a roughly straight-line trend, the variables are said to have an approximately **linear relationship**.

#### On the line



#### Around the line

CLARUSWAY®  
way to knowledge revealed

39

Use this space to take notes:

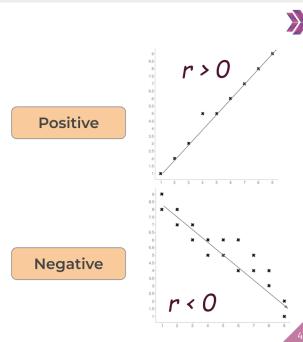
## Slide 40

### ► Direction

**Slope** determines the sign of the correlation coefficient.

When the slope of the line in the plot is negative, the correlation is negative; and vice versa.

CLARUSWAY®  
www.clarusway.com



Use this space to take notes:

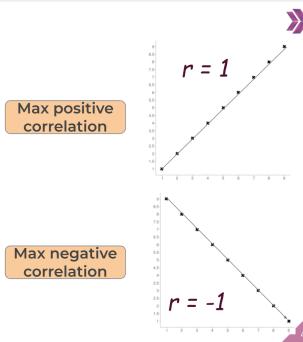
## Slide 41

### ► Strength

**Strength** is determined by scatter in data pattern.

The strongest correlations ( $r = 1.0$  and  $r = -1.0$ ) occur when data points fall exactly on a straight line.

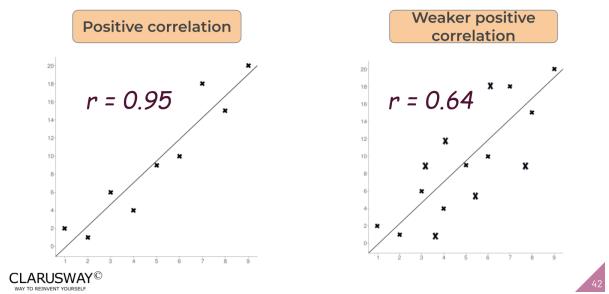
CLARUSWAY®  
www.clarusway.com



Use this space to take notes:

## Slide 42

### ► Strength

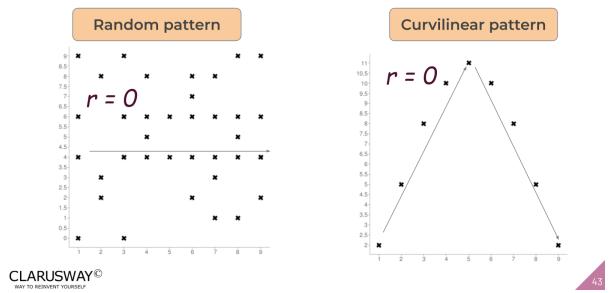


42

Use this space to take notes:

## Slide 43

### ► Zero Correlation



43

Use this space to take notes:

## Slide 44

## Your Response

## Slide 44

### Patterns of Data in Scatter Plot

Draw lines or circles.

Linearity	Strength
Linear	Weak
Nonlinear	Strong

Slope
Positive
Negative
Zero



## Your Response

### Patterns of Data in Scatter Plot

Draw lines or circles.

Linearity	Strength
Linear	Weak
Nonlinear	Strong

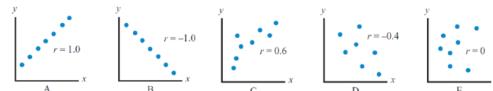
Slope
Positive
Negative
Zero



Use this space to take notes:

## Slide 45

### Examples of Approximate r Values

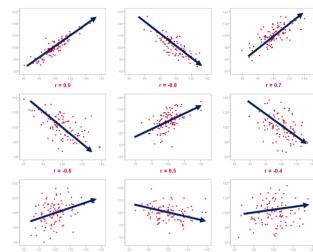


CLARUSWAY®  
WAY TO REINVENT YOURSELF.

Use this space to take notes:

## Slide 46

### ► Examples of Approximate r Values ➤



CLARUSWAY®  
WAY TO KNOW YOU BETTER

48

Use this space to take notes:

## Slide 47

### Your Response

### ► Let's Practice ➤

Which pair of variables exhibits the **strongest** linear relationship?  
Which pair of variables exhibits the **weakest** linear relationship?

Pair	Variable 1	Variable 2	Correlation
A	Internet Users	Facebook Users	0.507
B	Internet Users	Broadband Subscribers	0.949
C	Internet Users	Population	0.744
D	Facebook Users	Broadband Subscribers	0.619
E	Facebook Users	Population	0.097
F	Broadband Subscribers	Population	0.533



Students choose an option  
Answer now!

Pear Deck Interactive Slides  
The future of presentation software.

Use this space to take notes:

## Slide 48

### Your Response

## Slide 48

## Your Response

### Let's Practice

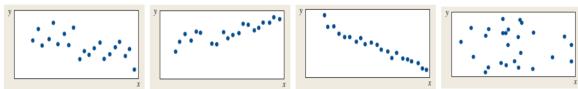
Match the scatter plots below with the correlation values.

$r = -0.9$

$r = 0$

$r = -0.5$

$r = 0.6$



Peer Deck Interactive Slide  
Do not remove this bar

Use this space to take notes:

## Slide 49



### 4 Normal Distribution

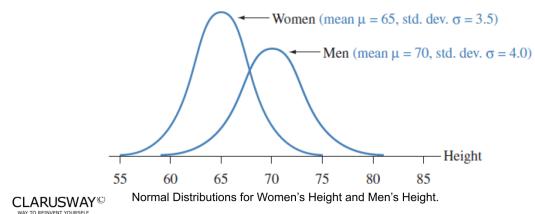
CLARUSWAY®  
www.clarusway.com

Use this space to take notes:

## Slide 50

### ► Normal Distribution

The normal distribution is symmetric, bell-shaped, and characterized by its mean  $\mu$  and standard deviation  $\sigma$ .



50

Use this space to take notes:

## Slide 51

### ► Normal Distribution

- ★ Symmetric around their mean.
- ★ The mean, median, and mode are equal.
- ★ The area under the normal curve is equal to 1.0.
- ★ Denser in the center and less dense in the tails.
- ★ Defined by two parameters:  $\mu$  and  $\sigma$ .  $\mathcal{N}(\mu, \sigma^2)$
- ★ Bell-shaped curve or Gaussian curve.



51

Use this space to take notes:

## Slide 52

## Your Response

## Slide 52

## Your Response

?

Which one is wrong about normal distributions?

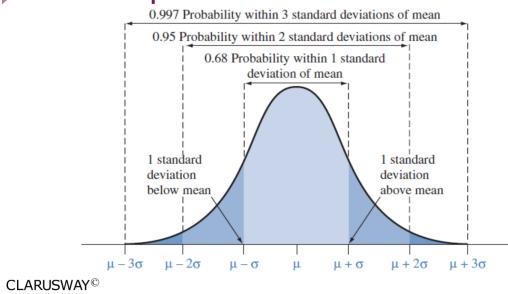
Students choose an option

Peer Deck Interactive Slide  
Do not remove this bar

Use this space to take notes:

## Slide 53

### ► The Empirical Rule



53

Use this space to take notes:

## Slide 54

### ► The Empirical Rule

If a distribution of data is bell shaped, then approximately

- ▶ 68% of the observations fall within 1 standard deviation of the mean.
- ▶ 95% of the observations fall within 2 standard deviations of the mean.
- ▶ 99.7% of the observations fall within 3 standard deviations of the mean.

CLARUSWAY®  
WAY TO KNOW™ WORKED



54

Use this space to take notes:

## Slide 55

### Your Response



A distribution has a mean of 40 and a standard deviation of 5.  
68% of the distribution can be found between what two numbers?



Students choose an option

Pear Deck Interactive Slide

Do not remove this bar

Use this space to take notes:

## Slide 56

### Your Response

## Slide 56

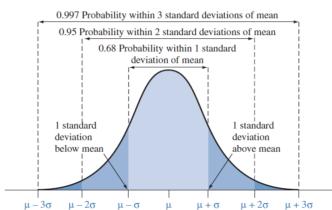
## Your Response

### ► The Empirical Rule

Question:

About what percentage (%) would fall more than 2 standard deviations from the mean?

Enter a number between 0-100



SWOTMS, enter a number!

Peer Deck Interactive Slide

You can't complete this slide

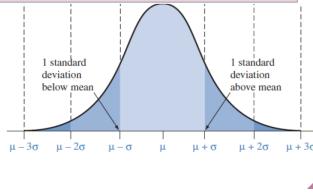
Use this space to take notes:

## Slide 57

### ► Finding Standard Score ( z-score)

- The normal random variable of a standard normal distribution is called a standard score or a z-score.
- Every normal random variable  $x$  can be transformed into a z score via the following equation:

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$



CLARUSWAY®  
WAY TO REVIVE YOURSELF

57

Use this space to take notes:

## Slide 58

### ▶ Finding z-score ➤

- SAT scores are approximately normally distributed with mean  $\mu = 500$  and standard deviation  $\sigma = 100$ .
- If your SAT score from one of the three components was  $x = 650$ , how many standard deviations from the mean was it?

$$z = \frac{x - \mu}{\sigma} = \frac{650 - 500}{100} = 1.50$$

CLARUSWAY®  
www.clarusway.com

58

Use this space to take notes:

## Slide 59

### ▶ Finding z-score ➤

- SAT scores are approximately normally distributed with mean  $\mu = 500$  and standard deviation  $\sigma = 100$ .
- What percentage of SAT scores was higher than yours?

CLARUSWAY®  
www.clarusway.com



Table 4. Normal Curve Areas  
Standard normal probability in right-hand tail  
(for negative values of  $z$ , areas are found by symmetry)

$z$	Second decimal place of $z$									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	3438	3399	3358	3317	3276	3236	3195	3155	3115	3075
0.2	2420	2418	2419	2409	2405	2403	2394	2384	2374	2369
0.3	1430	1428	1429	1418	1415	1413	1403	1394	1384	1373
0.4	3444	3409	3372	3334	3300	3264	3228	3192	3156	3121
0.5	3086	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2032	2003	1977	1949	1922	1894	1867
0.9	1430	1428	1429	1418	1415	1413	1403	1394	1384	1373
1.0	1387	1362	1339	1315	1292	1271	1251	1230	1210	1190
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1332	1311	1290	1269	1248	1227	1206	1185	1165	1145
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1.4	0886	0795	0778	0764	0749	0735	0722	0708	0694	0681
1.5	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1.8	0303	0293	0283	0273	0263	0254	0244	0234	0224	0214
1.9	0267	0251	0234	0214	0198	0182	0166	0150	0134	0120

59

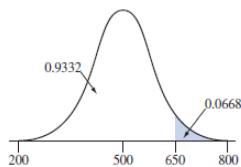
Use this space to take notes:

## Slide 60

### ▶ Finding z-score ➤

- SAT scores are approximately normally distributed with mean  $\mu = 500$  and standard deviation  $\sigma = 100$ .
- What percentage of SAT scores was higher than yours?

CLARUSWAY®  
way to knowledge revealed



60

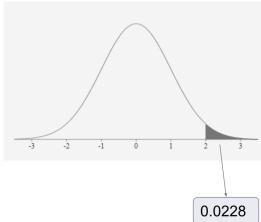
Use this space to take notes:

## Slide 61

### ▶ Normal Distribution - Example ➤

- Let  $Z$  denote a normal random variable with mean 0 and standard deviation 1.
- Find
  - $P(Z > 2)$

CLARUSWAY®  
way to knowledge revealed



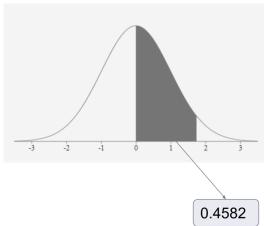
61

Use this space to take notes:

## Slide 62

### ► Normal Distribution - Example ➤

- Let  $Z$  denote a normal random variable with mean 0 and standard deviation 1.
- Find
  - $P(0 \leq Z \leq 1.73)$



CLARUSWAY®  
WAY TO KNOW™ WORKED

62

Use this space to take notes:

## Slide 63

### Your Response



An average light bulb manufactured by the Acme Corporation lasts 300 days with a standard deviation of 50 days.

Assuming that bulb life is normally distributed,

What is the probability that an Acme light bulb will last at most 365 days?



Students, write your response!

Pear Deck Interactive Slide  
Do not remove this bar

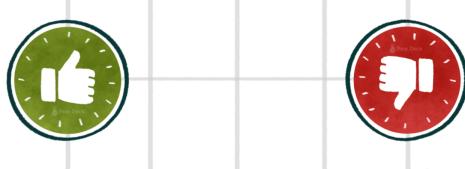
Use this space to take notes:

## Slide 64

### Your Response

## Slide 64

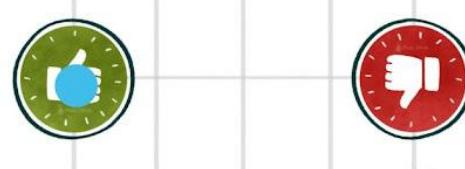
How well did you like this lesson?



Students, drag the icon! Pear Deck  
Interactive Slide  
Do not remove this bar

## Your Response

How well did you like this lesson?



Students, drag the icon! Pear Deck  
Interactive Slide  
Do not remove this bar

Use this space to take notes:

## Slide 65

# THANKS!

Any questions?

You can find me at:

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



Use this space to take notes: