DS-13/22 EU Session-1

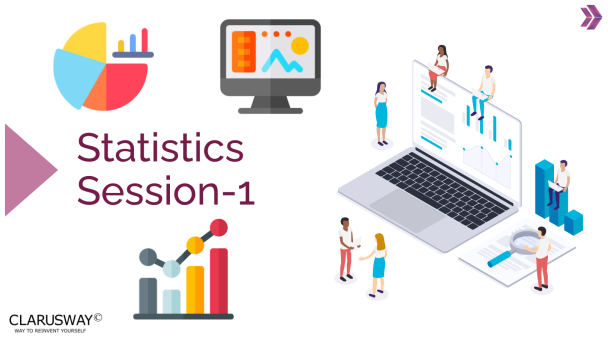
DS-13 EU Statistics 1
Training Clarusway
Pear Deck - November 21, 2022 at 0:00PM

Part 1 - Summary

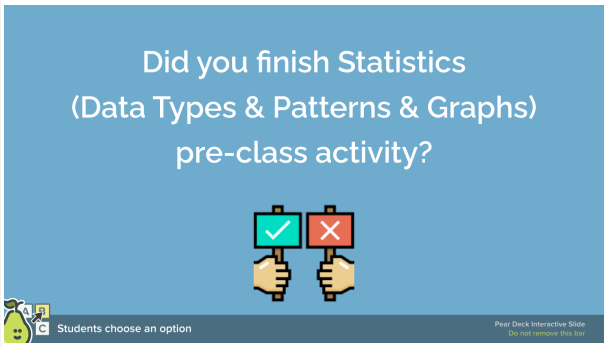
Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1



Use this space to take notes:

Slide 2	Your Response
	You Chose • **I finished completely.** Other Choices • I finished partially. • No, I didn't finish.

Use this space to take notes:

Slide 3


Use this space to take notes:

Slide 4

CLARUSWAY©
WAY TO REINVENT YOURSELF

4

Use this space to take notes:

Slide 5

Urgent & Important *do it now*	Important not urgent *decide when to do it*
Urgent not important *delegate it*	Not important not urgent *delete it*

CLARUSWAY©
WAY TO REINVENT YOURSELF

5

Use this space to take notes:

Slide 6

CLARUSWAY©
WAY TO REINVENT YOURSELF

6

Use this space to take notes:

Slide 7

CLARUSWAY©
WAY TO REINVENT YOURSELF

7

Use this space to take notes:

Slide 8

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. (jason@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

8

Use this space to take notes:

Slide 9

Scope of the Course

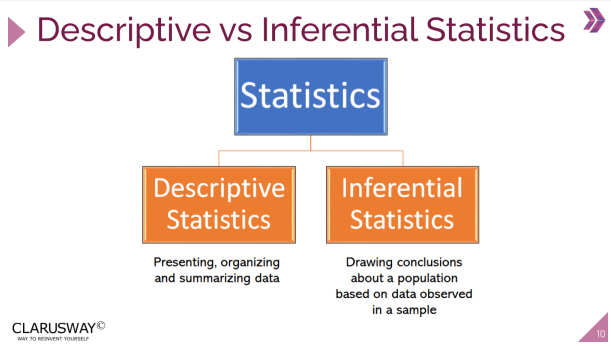
- Data Types & Patterns & Graphs
- Central Tendency & Dispersion
- Correlation & Normal Distribution
- Central Limit Theorem and Confidence Intervals
- Basic Concepts of Hypothesis Testing
- Hypothesis Tests about Means
- Analysis of Categorical Variables

CLARUSWAY©
WAY TO REINVENT YOURSELF

9

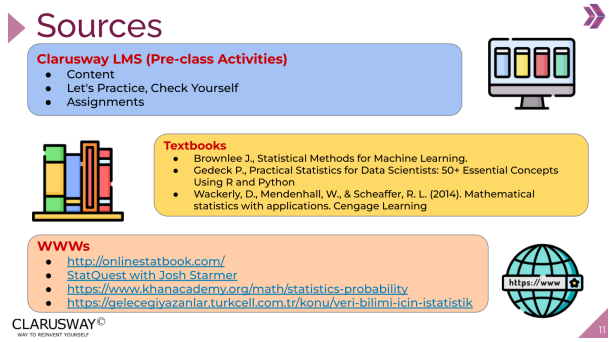
Use this space to take notes:

Slide 10



Use this space to take notes:

Slide 11



Link(s) on this slide:

- http://onlinestatbook.com/
- https://www.youtube.com/c/joshstarmer
- https://www.khanacademy.org/math/statistics-probability

Use this space to take notes:

Slide 12

Table of Contents

CLARUSWAY©
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 13

1 What is “Statistics”?

CLARUSWAY©
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 14

What is “Statistics”?

WIKIPEDIA → Statistics is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data.

Merriam-Webster's Collegiate Dictionary → A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

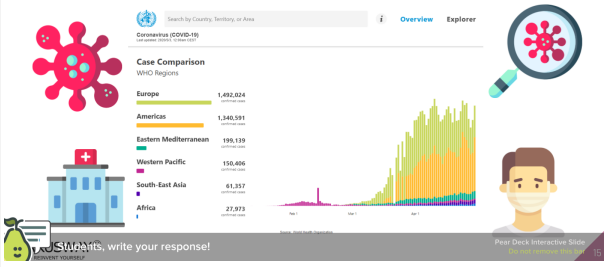
CLARUSWAY© → All the authors imply that statistics is a theory of information, with inference making as its objective.

14

Use this space to take notes:

Slide 15

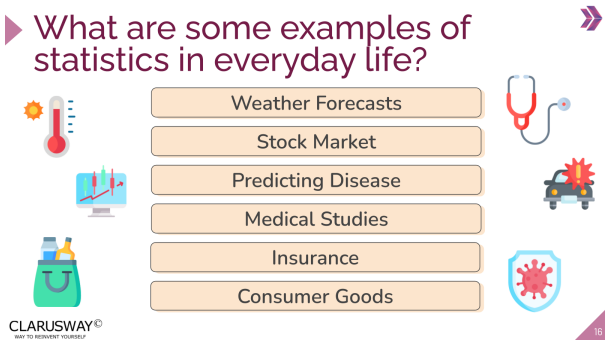
What are some examples of statistics in everyday life?



Your Response

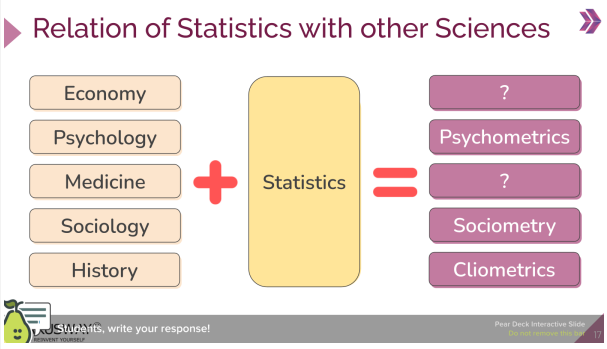
Use this space to take notes:

Slide 16



Use this space to take notes:

Slide 17



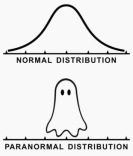
Your Response

Use this space to take notes:

Slide 18

Funny Statistics

"Excellent health statistics - smokers are less likely to die of age related illness"



CLARUSWAY©

Use this space to take notes:

Slide 19

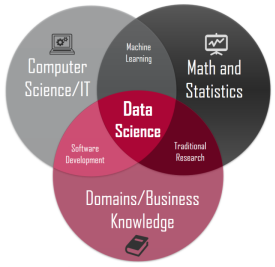
2 Why Should You Learn Statistics?

CLARUSWAY©

Use this space to take notes:

Slide 20

Data Science vs. Statistics



"A Data Scientist is one who knows more statistics than a programmer and more programming than a statistician"

CLARUSWAY©

20

Use this space to take notes:

Slide 21

3 Stats with Python

CLARUSWAY©

Use this space to take notes:

Slide 22

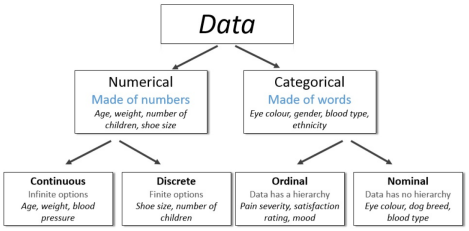
Use this space to take notes:

Slide 23

Use this space to take notes:

Slide 24

Types of Data



CLARUSWAY©

24

Use this space to take notes:

Slide 25

Numerical Data

Continuous Data

- Continuous data can have an infinite continuum of possible values.
  - height
  - weight
  - age
  - the amount of time it takes to complete an assignment

Discrete Data

- Any variable with a finite number of possible values is discrete.
  - the number of pets in a household
  - the number of children in a family
  - the number of foreign languages in which a person is fluent

CLARUSWAY©

25

Use this space to take notes:

Slide 26

Types of Data

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
339	7	178	84	0	0	39.9	0.331	41
403	9	72	78	25	0	31.6	0.280	38
551	3	84	68	30	106	31.9	0.591	25
197	3	107	62	13	48	22.9	0.678	23
563	6	99	60	19	54	26.9	0.497	32
239	0	104	76	0	0	18.4	0.582	27
141	5	106	82	30	0	39.5	0.286	38
523	9	130	70	0	0	34.2	0.652	45
696	3	169	74	19	125	29.9	0.268	31
238	9	164	84	21	0	30.8	0.831	32

CLARUSWAY©
WAY TO REINVENT YOURSELF

26

Use this space to take notes:

Slide 27

Categorical Data

Ordinal Data

- Ordinal data requires an order
  - small medium, large
  - good, average, poor
  - strongly agree, agree, disagree
- The distance between ordered categories is not measurable.
- No arithmetic can be done with the ordinal data as they show sequence only.

Nominal Data

- Nominal data simply names something without an order being given.
  - employee's status
  - color
  - race
- Data obtained on nominal scale is in terms of frequency.

CLARUSWAY©
WAY TO REINVENT YOURSELF

27

Use this space to take notes:

Slide 28

Types of Data

	gender	race/ethnicity	parental level of education	lunch
0	female	group B	bachelor's degree	standard
1	female	group C	some college	standard
2	female	group B	master's degree	standard
3	male	group A	associate's degree	free/reduced
4	male	group C	some college	standard

CLARUSWAY©
WAY TO REINVENT YOURSELF

28

Use this space to take notes:

Slide 29

Types of Data

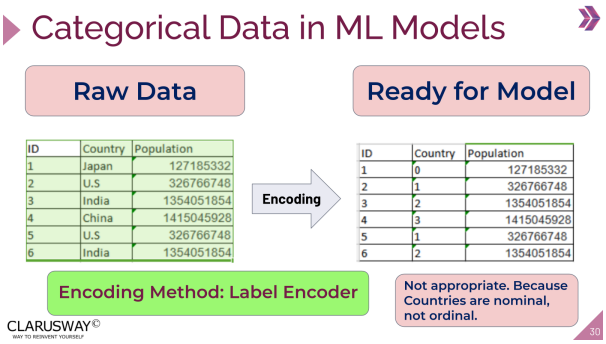
	division	level of education	training level	work experience	salary	sales
0	printers	some college	2	6	91684	372302
1	printers	associate's degree	2	10	119679	495660
2	peripherals	high school	0	9	82045	320453
3	office supplies	associate's degree	2	5	92949	377148
4	office supplies	high school	1	5	71280	312802

CLARUSWAY©
WAY TO REINVENT YOURSELF

29

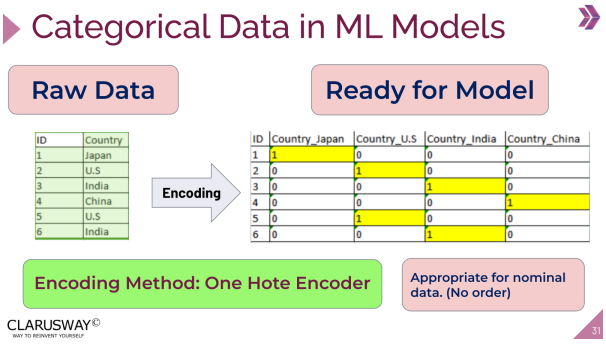
Use this space to take notes:

Slide 30



Use this space to take notes:

Slide 31

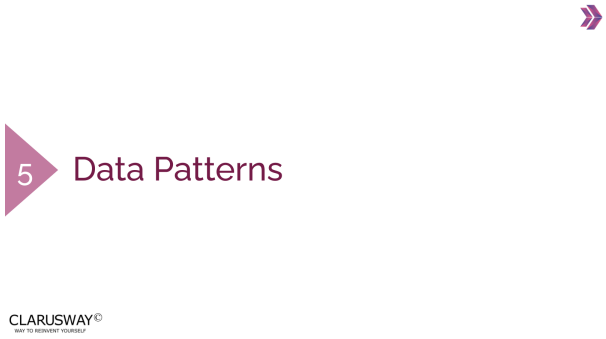


Use this space to take notes:

Slide 32	Your Response
	You Chose • **Race**

Slide 32	Your Response
	Other Choices ● Height

Use this space to take notes:

Slide 33


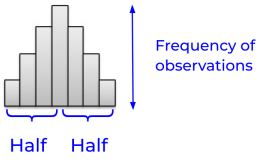
Use this space to take notes:

Slide 34

Data Patterns in Statistics

Center | Spread | Shape | Unusual Features

The center of a distribution, graphically, is located at the median of the distribution.



CLARUSWAY©

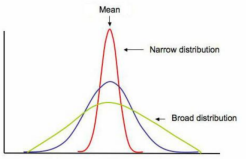
Use this space to take notes:

Slide 35

Data Patterns in Statistics

Center | **Spread** | Shape | Unusual Features

The spread of a distribution refers to the variation of the data.



CLARUSWAY©

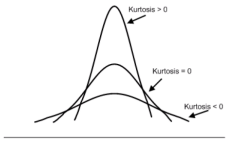
Use this space to take notes:

Slide 36

Data Patterns in Statistics

Center | Spread | **Shape** | Unusual Features

Kurtosis - Some distributions may have multiple observations on one side of the graph than the other side.



CLARUSWAY©
WAY TO REINVENT YOURSELF

36

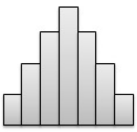
Use this space to take notes:

Slide 37

Data Patterns in Statistics

Center | Spread | **Shape** | Unusual Features

Symmetry - In symmetric distribution, graph can be divided at the center in such a way that each half is a mirror image of the other.

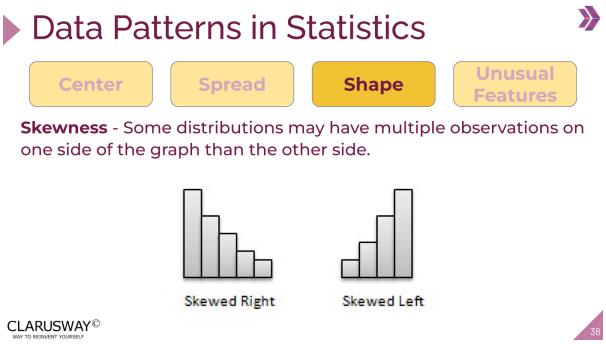


CLARUSWAY©
WAY TO REINVENT YOURSELF

37

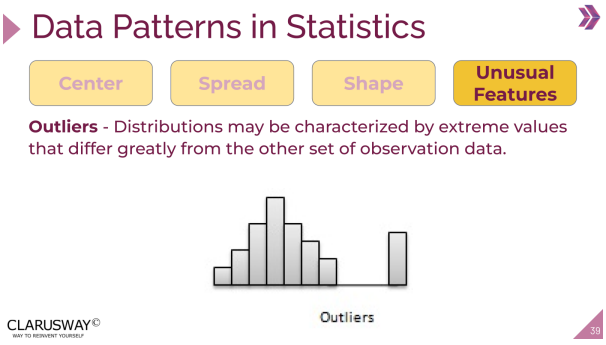
Use this space to take notes:

Slide 38



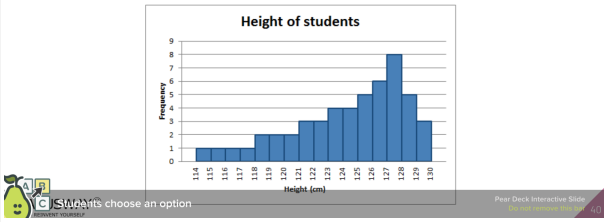
Use this space to take notes:

Slide 39



Use this space to take notes:

Slide 40	Your Response
	You Chose • **The distribution is left-skewed with**

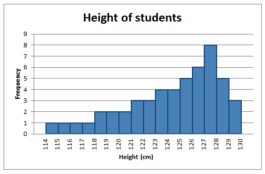
Slide 40	Your Response
Let's Practice Which of the following statements is true about the figure?  	**no outliers.** Other Choices • The distribution is right-skewed with no outliers. • The distribution is right-skewed with one outlier. • The distribution is left-skewed with one outlier. • The distribution is symmetric.

Use this space to take notes:

Slide 41

Let's Practice Answer

Which of the following statements is true about the figure?

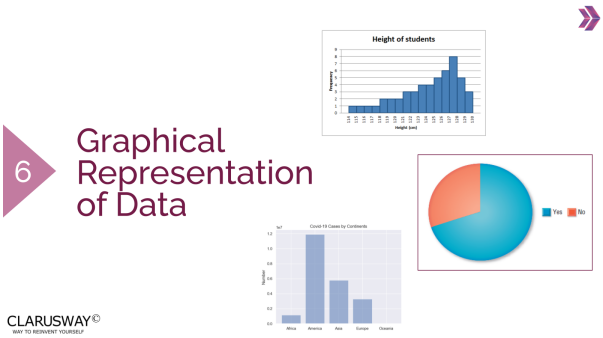


C: The distribution is left-skewed with no outliers.

CLARUSWAY©

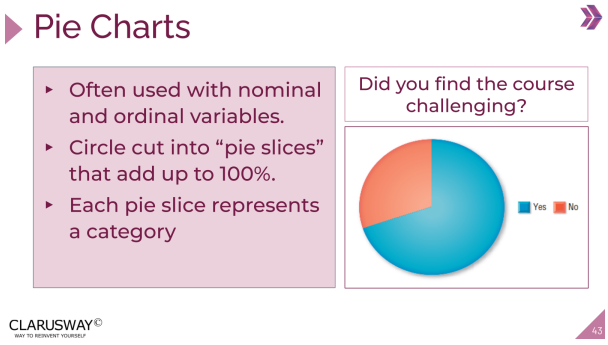
Use this space to take notes:

Slide 42



Use this space to take notes:

Slide 43



Use this space to take notes:

Slide 44

Pie Charts

The following table shows the numbers of hours spent by a child on different events on a working day.

Represent the data on a pie chart

Activity	No. of Hours
School	6
Sleep	8
Playing	2
Study	4
T. V.	1
Others	3

CLARUSWAY©

Use this space to take notes:

Slide 45

Pie Charts

The central angles for various observations can be calculated as:

Activity	No. of Hours	Measure of central angle
School	6	$(^{6}/_{24} \times 360)° = 90°$
Sleep	8	$(^{8}/_{24} \times 360)° = 120°$
Playing	2	$(^{2}/_{24} \times 360)° = 30°$
Study	4	$(^{4}/_{24} \times 360)° = 60°$
T. V.	1	$(^{1}/_{24} \times 360)° = 15°$
Others	3	$(^{3}/_{24} \times 360)° = 45°$

CLARUSWAY©

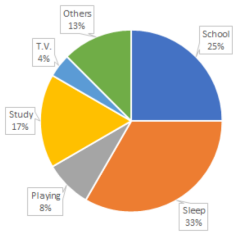
Use this space to take notes:

Slide 46

Pie Charts

Now, we shall represent these angles within the circle as different sectors.

Activity	No. of Hours
School	6
Sleep	8
Playing	2
Study	4
T. V.	1
Others	3



CLARUSWAY©

46

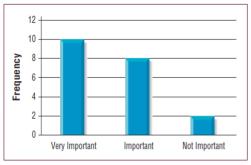
Use this space to take notes:

Slide 47

Bar Charts

- Often used with nominal and ordinal variables.
- A series of bars represent the different attributes of a variable.
- The height of each bar reflects frequencies for each attribute.

How important is *Data Science* to you?



CLARUSWAY©

47

Use this space to take notes:

Slide 48

Bar Charts

The following table shows the numbers of Covid-19 data for continents.

Represent the data on a bar chart

continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471

CLARUSWAY©
WAY TO REINVENT YOURSELF

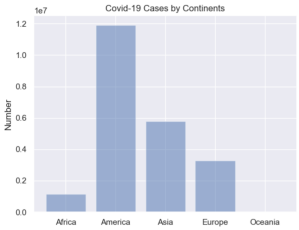
48

Use this space to take notes:

Slide 49

Bar Charts

continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471



CLARUSWAY©
WAY TO REINVENT YOURSELF

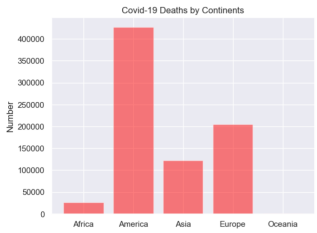
49

Use this space to take notes:

Slide 50

Bar Charts

continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471



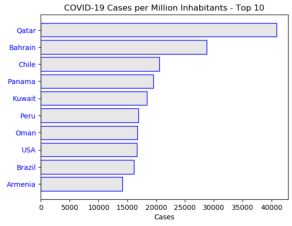
CLARUSWAY©
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 51

Bar Charts

countriesAndTerritories	cases	deaths	popData2019	casesPer1M
Qatar	115661	193	2832071.0	40839.724710
Bahrain	47185	175	1641164.0	28750.935312
Chile	388855	10546	18952035.0	20517.849402
Panama	82790	1809	4246440.0	19496.331044
Kuwait	77470	505	4207077.0	18414.210151
Peru	549321	26658	32510462.0	16896.745423
Oman	83418	597	4974992.0	16767.464149
USA	5482416	171821	329064917.0	16660.591016
Brazil	3407354	109888	211049519.0	16144.808177
Armenia	41846	832	2957728.0	14148.021725



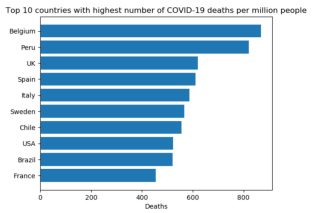
CLARUSWAY©
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 52

Bar Charts

countriesAndTerritories	cases	deaths	popData2019	casesPer1M	deathsPer1M
Belgium	78804	9959	11455519.0	6879.129614	869.362619
Peru	549321	26658	32510462.0	16896.745423	819.982195
UK	320286	41381	66647112.0	4805.699608	620.897122
Spain	364196	28670	46937060.0	7759.241844	610.817976
Italy	254636	35405	60359546.0	4218.653334	586.568362
Sweden	85219	5790	10230185.0	8330.152387	565.972170
Chile	388855	10546	18952035.0	20517.849402	556.457394
USA	5482416	171821	329064917.0	16660.591016	522.149251
Brazil	3407354	109888	211049519.0	16144.808177	520.674013
France	221267	30451	67012883.0	3301.857644	454.405163



CLARUSWAY©
WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 53

Let's Practice

The graph shows the number of complaints for six different airlines as reported to the US Department of Transportation in February 2013. Alaska, Pinnacle, and Airtran Airlines have far fewer complaints reported than American, Delta, and United.

Can we conclude that American, Delta, and United are the worst airline carriers since they have the most complaints?

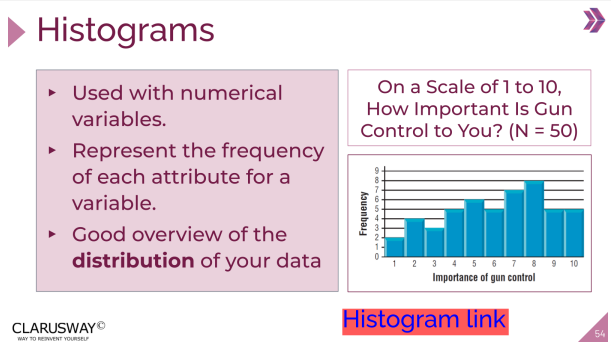
Students, write your response!

Pear Deck Interactive Slide

Your Response

Use this space to take notes:

Slide 54

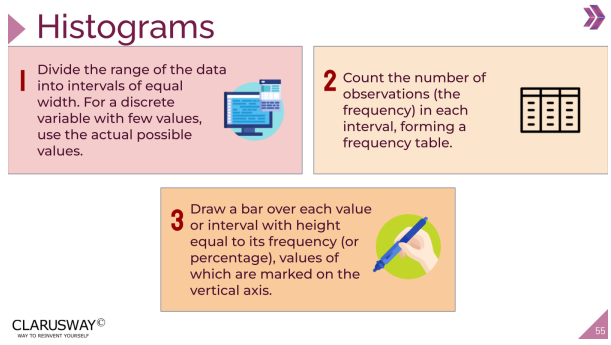


Link(s) on this slide:

- https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/histograms/v/histograms-intro

Use this space to take notes:

Slide 55



Use this space to take notes:

Slide 56

7 Populations & Samples

CLARUSWAY©

Use this space to take notes:

Slide 57

Populations

The study of statistics revolves around the study of data sets.

Populations → Include each element from the set of observations that can be made.

The set of all stars within the Milky Way galaxy

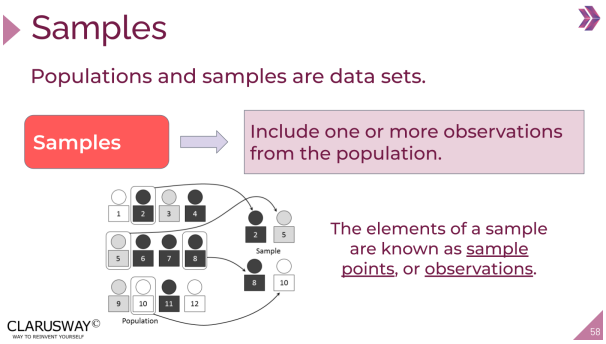
All people living in the USA

CLARUSWAY©

57

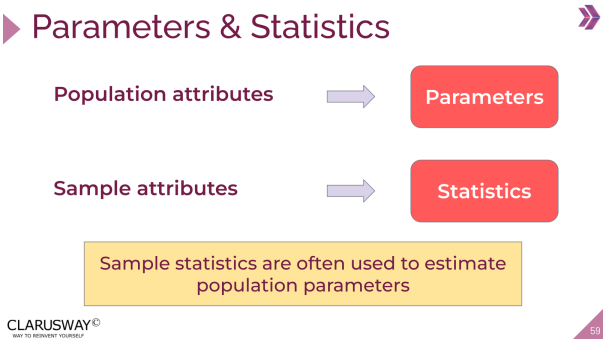
Use this space to take notes:

Slide 58



Use this space to take notes:

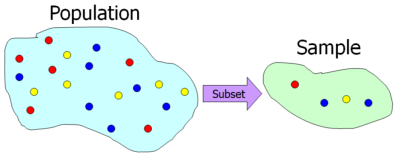
Slide 59



Use this space to take notes:

Slide 60

Parameters & Statistics



- Populations have Parameters (like μ, σ^2, θ, p)
- Samples have Statistics, functions of observed data, like $\bar{x}, \tilde{x}, s^2, \hat{\theta}, \hat{p}$

CLARUSWAY©
WAY TO REINVENT YOURSELF

60

Use this space to take notes:

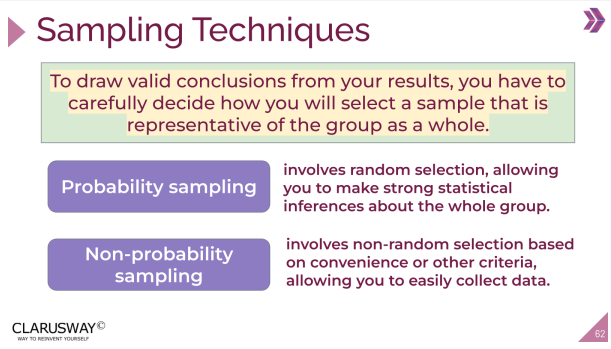
Slide 61

7 Sampling Techniques

CLARUSWAY©
WAY TO REINVENT YOURSELF

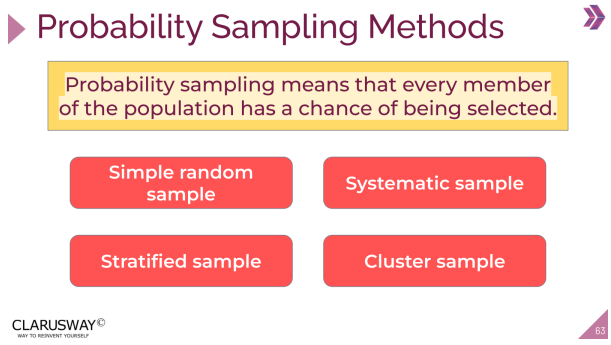
Use this space to take notes:

Slide 62



Use this space to take notes:

Slide 63



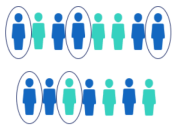
Use this space to take notes:

Slide 64

Simple Random Sample

In a simple random sample, every member of the population has an equal chance of being selected.

Simple random sample



CLARUSWAY©

64

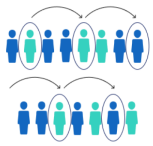
Use this space to take notes:

Slide 65

Systematic Sample

Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

Systematic sample



CLARUSWAY©

65

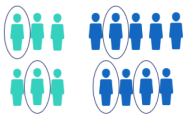
Use this space to take notes:

Slide 66

Stratified Sample

Stratified sampling involves dividing the population into subpopulations that may differ in important ways.

Stratified sample



CLARUSWAY©
WAY TO REINVENT YOURSELF

66

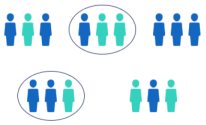
Use this space to take notes:

Slide 67

Cluster Sample

Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample.

Cluster sample



CLARUSWAY©
WAY TO REINVENT YOURSELF

67

Use this space to take notes:

Slide 68

Non-probability Sampling Methods

In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included.

- Convenience sample
- Voluntary response sample
- Purposive sample
- Snowball sample

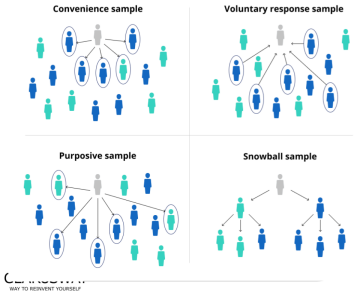
CLARUSWAY©

68

Use this space to take notes:

Slide 69

Non-probability Sampling Methods

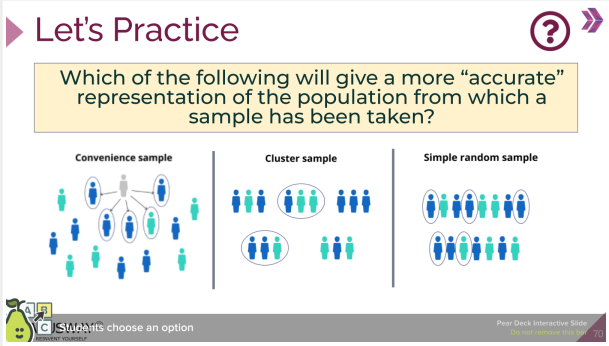


Researchers use this method in studies where it is impossible to draw random probability sampling due to time or cost considerations.

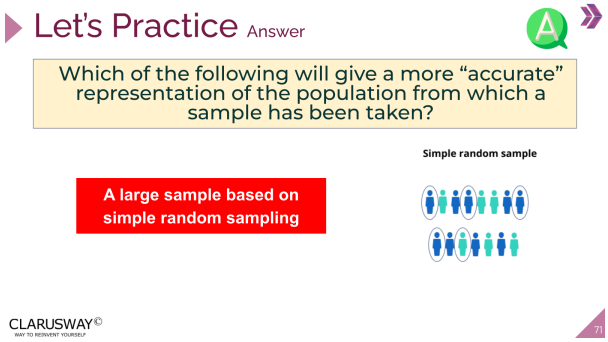
69

Use this space to take notes:

Slide 70	Your Response

Slide 70	Your Response
	

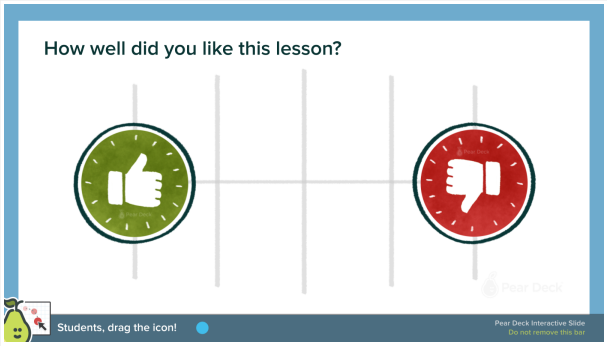
Use this space to take notes:

Slide 71


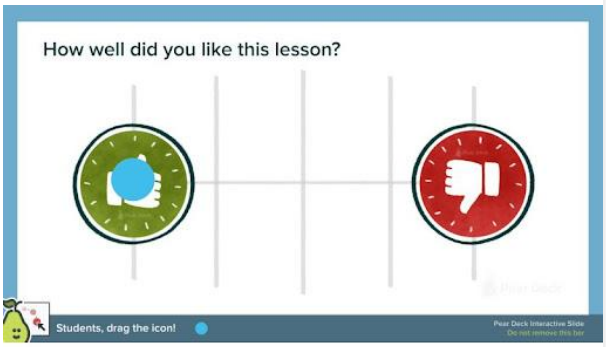
Use this space to take notes:

Slide 72	Your Response

Slide 72



Your Response



Use this space to take notes:

Slide 73

Use this space to take notes: