

Data Analysis = Convert raw data into useful information to gain insight and make decisions.

Synonyms: Data Analytics, Analytics, Business Intelligence, Statistical Analysis, Mathematical Analysis, Data Analysis

Goal: help make data-driven decisions, which tends to be more accurate and help to achieve goals more consistently

Raw Data = data stored in its smallest form in a cell

Not Raw Data:

Data about Google Stock
10/25/2021, \$2,775.46, 1,054,085, Google

Raw Data:

Date	Close	Volume	Company
10/25/2021	\$ 2,775.46	1,054,085	Google

Proper Data Set = Data Set = Table

Table is made up of:

Field = column in table

Field Name = name at top of field that describes what data goes into field

Record = one row in table

Table requirements in Excel:

- 1) Field names in first row
- 2) Records of related data in subsequent rows
- 3) Empty cells or Excel Row/Column Headers, all the way around table

Close Price	Volume	Company
\$2,775	1,054,085	Google
\$2,793	1,412,937	Google
\$2,929	2,592,546	Google
\$2,923	1,620,903	Google
\$2,965	1,447,725	Google
\$3,320	2,225,956	Amazon
\$3,376	2,698,342	Amazon
\$3,392	2,702,224	Amazon
\$3,447	5,708,733	Amazon
\$3,372	6,486,077	Amazon
\$308	17,554,469	Microsoft
\$310	28,107,349	Microsoft
\$323	52,588,690	Microsoft
\$324	26,297,943	Microsoft
\$332	34,765,982	Microsoft
\$16	67,808,039	Ford
\$16	64,882,295	Ford
\$16	96,094,256	Ford
\$17	215,237,578	Ford
\$17	100,560,723	Ford
\$202	3,294,689	Caterpillar
\$200	2,933,573	Caterpillar
\$196	3,406,655	Caterpillar

← Fields
 ← Records
 ←
 ←

Raw Data ==>>
 Measurements for the 5 elements

Textbook definitions:

Data = Facts and figures collected, measured, summarized, and analyzed for presentation and interpretation.

Table = field names in first row, records in subsequent rows

Elements = unique list of entities on which variable data are collected

Each company name is an element in this element field

Variable = a characteristic or attribute of interest for the element

There are 5 variable fields in this table

Company	Ticker	Year Incorporated	Industry	No. Employees	P/E Ratio (\$)
Google	GOOG	2015	Software	150,028	28.12
Amazon	AMZN	1996	Retail	1,335,000	58.72
Microsoft	MSFT	1993	Software	181,000	36.84
Ford	F	1919	Auto	186,000	20.09
Caterpillar	CAT	1986	Machinery	97,300	21.82

There are 6 fields in the above table: 1 element field and 5 variable fields

For each record there are 6 cells: 1 element cell and 5 variable cells

Observation = variable data for one element

Observation for the element Google =

GOOG	2015	Software	150,028	28.12
------	------	----------	---------	-------

Measurements are made for each variable to provide data in the observation.

Measurements made on past 5 days of raw data:

Company	Mean Price	σ Price	Sum Volume
Google	\$2,877	\$77	8,128,196
Amazon	\$3,382	\$41	19,821,332
Microsoft	\$319	\$9	159,314,433
Ford	\$16	\$1	544,582,891
Caterpillar	\$201	\$3	18,389,846

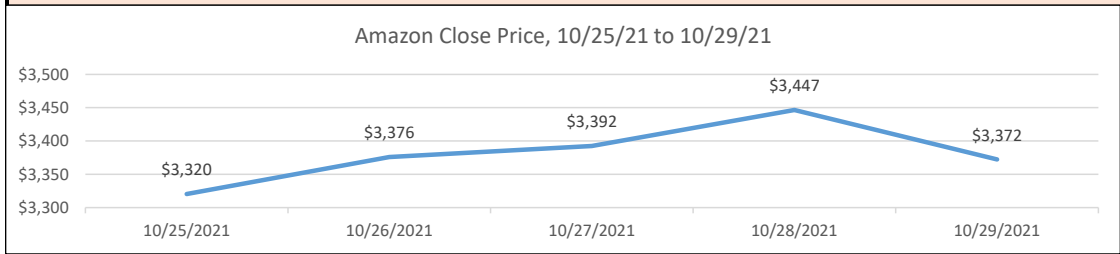
Categorical data = data grouped by a category						Quantitative data = numeric data					
Nominal: Category No Rank <u>Calcs:</u> Count % of total based on the counts		Ordinal: Category Rank Don't know distance between each rank <u>Calcs:</u> Count % of total based on the counts Average (Mean) OK for number category				Interval: Rank Know distance between each rank Zero is either: Not in the scale (like IQ scores) Just point on the scale (like Celsius) <u>Calcs:</u> Count % of total based on the counts Averages Differences Ratio Not Okay			Ratio: Rank Know distance between each rank Scales where zero = nothing exists Ratio Okay <u>Calcs:</u> Count % of total based on the counts Averages Differences Ratio OK		
Student	Phone	Eye Color	Rank In Class (1st to 4th)	Rating of Teacher (Bad, Good, Great)	Rating of Teacher (1,2,3)	Temperature Fahrenheit	SAT Score (400 to 1200)	Score On Final (0 - 100)	Temperature Kelvin	Money In Bank	
Chantel	iPhone	Brown	1st	Good	2	-10	1170	81	249.8	\$12,000.00	
Michael	Samsung	Brown	4th	Great	3	0	590	90	255.4	\$6,000.00	
Mo	iPhone	Brown	3th	Bad	1	20	1180	45	266.5	\$4,369.00	
Sioux	iPhone	Hazel	2nd	Good	2	10	1099	84	260.9	\$3,131.00	

Types of numbers:
Discrete numbers are counting numbers & have gaps between each successive number, like 1, 2, 3..., or 1.2, 1.3, 1.4.... Answers the question How Many?
Continuous numbers can occupy any value over a continuous range and depend on the measuring instrument. Like: time, weight, temperature, money (\$ don't seem continuous, but many statisticians treat it as such). Answers the question How Much?

Cross-sectional data = Data collected at the same or nearly the same point in time
 This cross-sectional data was collected on 11/5/2021:
 Useful to compare elements such as company.

Company	Ticker	Year Incorporated	Industry	No. Employees	P/E Ratio (x)
Google	GOOG	2015	Software	150,028	28.12
Amazon	AMZN	1996	Retail	1,335,000	58.72
Microsoft	MSFT	1993	Software	181,000	36.84
Ford	F	1919	Auto	186,000	20.09
Caterpillar	CAT	1986	Machinery	97,300	21.82

Time-series data = Data collected over time
 This data was collected for Amazon over a 5-day period
 Useful for seeing trends in the past that may help us to estimate what trends there may be in the near future



Different classifications of data in statistics:

Categorical data = data grouped by a category (also known as Qualitative data)
Nominal = data with no rank or order. Like eye color: Brown, Hazel, Brown; or Phone name: iPhone, Samsung, iPhone. With this data you can count and then calculate the % of total based on the counts. If you use a number as a category label, you cannot do arithmetic (+, -, *, /, ^) on it - you can only count.
Ordinal = data with rank but you don't know distance between each rank, Like: Bad, Good, Great; or *, **, ***, ****; or 1st, 2nd, 3rd. You usually count how many are in each category and then calculate the % of total. In some cases when a number is used to represent the category, you can calculate an average (mean).
Quantitative data = numeric data
Interval = data with rank and a fixed distance between each rank, but where zero is either not in the scale (like IQ or SAT scores) or zero is just a point on the scale (Like Fahrenheit temperature or Celsius temperature). Calculations like counting, % of count total, differences, averages and differences are OK, but ratios between two numbers is not OK.
Ratio = data with rank and a fixed distance between each rank and zero means nothing exists. Like money, weight, height, time, Kelvin temperature. All are scales where zero means nothing exists. Calculations like counting, % of count total, differences and ratios are OK.
Types of numbers:
Discrete numbers are counting numbers & have gaps between each successive number, like 1, 2, 3..., or 1.2, 1.3, 1.4.... Answers the question How Many?
Continuous numbers can occupy any value over a continuous range and depend on the measuring instrument. Like: time, weight, temperature, money (\$ don't seem continuous, but many statisticians treat it as such). Answers the question How Much?

Statistics

Numerical facts like:

USA unemployment rate reported Jan. 2021 was 6.3%

Sioux Radcoolinator student ranked at the 90th percentile for the test

3rd quarter YouTube advertising revenue was \$7.20 billion vs. \$7.4 billion expected.

Subject of Statistics defined:

Statistics is the art and science of collecting, analyzing, presenting and interpreting data.

Descriptive Statistics:

Data that is summarized and presented

Tabular: table of information

Graphical: charts, graphs, visualizations

Numerical: like an average (mean), median, mode

Inferential Statistics:

The process of using data obtained from a sample to make estimates and test hypotheses about the characteristics/attribute of a population

Take a sample from the population and draw reasonable conclusions that can help to estimate the unknown future.

Define Terms:

Population: The set of all elements of interest in a particular study

(In many situations it is too costly to get data from all the elements in the population)

Example ==>> Census: Collecting data for a population

Sample: A subset of the population

Example ==>> Sample survey: Collecting data for a sample