

An Introduction to Statistics

~ Part 1 ~

An Introduction to Statistics

~ Part 1 ~

- What does a statistician do?
- What *is/are* statistics???
- Data?
- Terminology

What does a statistician do?

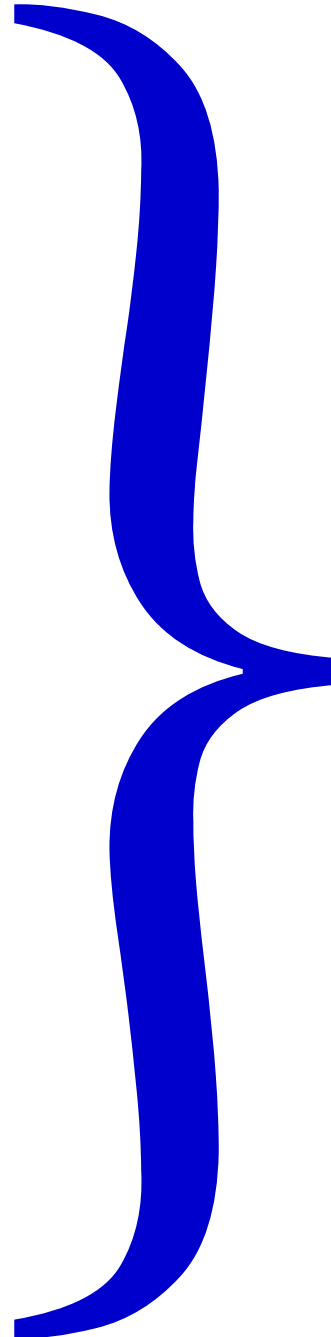
- A statistician makes sense of information collected about the world

What does a statistician do?

- A statistician makes sense of information collected about the world
 - Analyze effectiveness of drug/treatment
 - Determining behavior patterns
 - Analyze spread of a virus
 - ◉ [West Nile](#) | [EEE](#) | [Coronavirus](#)
 - Analyze performance of
 - ◉ Individuals (ex. Baseball players)
 - ◉ Businesses
 - ◉ Stocks

- Economics -
- Psychology -

- Biology -
- Anthropology -
- History -
- Philosophy -
- Engineering -
- Accounting -
- Statistics -



**In general,
what
are these
subjects
about???**

- Economics - about money
- Psychology - about *why* we think
what we think (we think)
- Biology - about life
- Anthropology - about who
- History - about what, where, when
- Philosophy - about why
- Engineering - about how
- Accounting - about how much
- Statistics -

- Economics - about money
- Psychology - about *why* we think
what we think (we think)
- Biology - about life
- Anthropology - about who
- History - about what, where, when
- Philosophy - about why
- Engineering - about how
- Accounting - about how much
- Statistics - *about variation*

What does it mean to Vary?

What does it mean to Vary?

- to make a partial change in
- make different in some attribute or characteristic
- to make differences between items in
- to exhibit or undergo change
- to take on successive values
- to exhibit divergence in structural or physiological characters from the typical form

Merriam-Webster Dictionary,
<https://www.merriam-webster.com/dictionary/vary>

What does it mean to Vary?

- to make a partial change in
- make different in some attribute or characteristic
- to make differences between items in
- to exhibit or undergo change
- to take on successive values
- to exhibit divergence in structural or physiological characters from the typical form

Merriam-Webster Dictionary,
<https://www.merriam-webster.com/dictionary/vary>

What is Variation?

- the act, process, or accident of varying in condition, character, or degree
- an instance of this
- amount, rate, extent, or degree of change
- a different form of something; variant

Dictionary.com,
<https://www.dictionary.com/browse/variation>

What is Variation?

- the act, process, or accident of varying in condition, character, or degree
- an instance of this
- amount, rate, extent, or degree of change
- a different form of something; variant

Dictionary.com,
<https://www.dictionary.com/browse/variation>

What is Statistics?

- Statistics *is* a branch of mathematics dealing with ...

What is Statistics?

- Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical/categorical data

What is Statistics?

- **Statistics *is* ...**

What is Statistics?

- Statistics is a way of reasoning together with collection tools and methods designed to help us to understand data and information about the world.

What is Statistics?

- Statistics is a way of reasoning together with collection tools and methods designed to help us to understand data and information about the world
- Two types

What is Statistics?

- Statistics is a way of reasoning together with collection tools and methods designed to help us to understand data and information about the world
- Two types
 - Descriptive statistics
 - Inferential statistics

Descriptive Statistics

- Methods for **organizing and summarizing** information and data

Descriptive Statistics

- **Methods for organizing and summarizing information and data**
- **Examples**
 - **Tables**
 - **Graphs**
 - **Summary statistics**
 - **Summary parameters**

Descriptive Statistics

- **Methods for organizing and summarizing information and data**
- **Examples**
 - **Tables - relative frequency distribution**
 - **Graphs - histogram, pie chart**
 - **Summary statistics/parameters**
 - **Mean, median, mode, standard deviation**

Inferential Statistics

- **Methods which**
 - **Generalize/extend information and results** obtained from a *sample* to a *population*
and
 - **Measure the reliability** of the results

Inferential Statistics

- **Methods which**
 - **Generalize/extend information and results obtained from a sample to a population**
 - and
 - **Measure the reliability of the results**
- **Examples**
 - **Confidence intervals**
 - **Hypothesis testing**

What are statistics?

- **Statistics are calculations made from sample data.**

What *is* a statistic?

- A statistic *is* a numerical summary based on data from a sample.

What are parameters?

- Parameters are calculations made from population data.

What is a parameter?

- A parameter is a numerical summary based on data from a population.

Sample vs. Population

- Note the S: *statistics* are for *samples*
- Note the P: *parameters* are for *populations*

Population

- The entire group being studied

Sample

- A subset of the population being studied

Individual or Subject

- A member of the population being studied

Individual or Subject

- A member of the population being studied
 - Person
 - Place
 - Thing
 - ◉ Tangible
 - ◉ Intangible

What are data?

- **Systematically recorded information together with context**
- **Note: Data is plural and datum is singular.**

Context

- Tells
 - *Who* was measured
 - *What* was measured
 - *Where* data were collected
 - *When* data were collected
 - *Why* study was performed
 - *How* data were collected

Data is useless without ...

Data is useless without ...

- **Context**

Data is useless without ...

- **Context**

- ◊ Who
- ◊ What
- ◊ Where
- ◊ When
- ◊ Why
- ◊ How

Data can be ...

- **Numbers**
 - **Measurements or counts**
- **Dates**
- **Words**
- **Codes or other labels**
- **Symbols**
- **Pictures**

Data are ...

- Values of a variable

Variable

- A variable is a characteristic or information collected about each individual in the study
 - What is measured or recorded

Data Vary

- Values of a variable change

Data Vary

- Values of a variable change
 - from individual to individual
 - over time for certain individuals

Data Vary

- Values of a variable change
 - from individual to individual
 - over time for certain individuals
- Two different statistical analyses of the same variable can lead to different results

Data Table

- An arrangement of data

Data Table

- Rows represent
 - Individuals - those studied/examined
 - ◉ Respondents of a survey
 - ◉ Subjects of an experiment
 - ◉ Participants of a study
 - ◉ Experimental units (animals, plants, websites, inanimate objects)
- Rows are referred to as
 - Records (for a database)
 - Most generally, cases or units

Data Table

- **Columns contain**
 - **Characteristics/information collected/recorded about each individual**
 - ◉ **Values of each variable**
 - **Data associated with the variables**

Data Table

- Caution: The heading for a column may be a code or word used to represent the actual variable

- For the following two examples,
 - what are the records or cases?
 - what information was collected?

You are hired to analyze the per capita income, in dollars, as well as the percentages of the labor force employed in agriculture, industry, and service occupations for the twenty (20) OECD* countries for 1960.

COUNTRY	Per Capita Income	Agriculture	Industry	Service
UNITED KINGDOM	1105	4	56	40
BELGUIM	1005	6	52	42
SWITZERLAND	1361	11	56	33
NETHERLANDS	810	11	49	40
CANADA	1536	13	43	45
SWEDEN	1644	14	53	33
LUXEMBOURG	1242	15	51	34
WEST GERMANY	1035	15	60	25
DENMARK	1049	18	45	37
FRANCE	1013	20	44	36
NORWAY	977	20	49	32
AUSTRIA	681	23	47	30
ICELAND	839	25	47	29
ITALY	504	27	46	28
JAPAN	344	33	35	32
IRELAND	529	36	30	34
SPAIN	290	42	37	21
PORTUGAL	238	44	33	23
GREECE	324	56	24	20
TURKEY	177	79	12	9

*OECD - Organization for Economic Cooperation and Development

<http://www.oecd.org/general/listofocdmembercountries-ratificationoftheconventionontheoecd.htm>

For your summer internship with U.S. News and World Report, you are asked to analyze data on the Class of 1999 for the top twenty (20) universities in the United States. From your analysis, you obtain information on the percentage of incoming freshmen who graduate within four years (the freshman retention rate), the percentage of students admitted as freshmen who graduated (the graduation rate), determining both the predicted graduation rate and the actual graduation rate, the percentage of freshman applicants who were accepted (the acceptance rate), and the percentage of alumni who made donations to the university after graduation (the alumni giving rate) for this class.

School	Freshman Retention	Predicted Graduation Rate	Actual Graduation Rate	Acceptance Rate	Alumni Giving Rate
University of California – Berkeley	95	81	82	27	18
Johns Hopkins University	95	89	87	33	28
Northwestern University	96	85	92	32	29
Columbia University	97	89	91	14	32
University of Chicago	94	89	81	48	35
Cornell University	96	87	91	33	36
Stanford University	98	91	90	15	37
Washington University in St. Louis	96	83	86	34	37
Rice University	95	91	88	27	39
Emory University	92	83	86	42	39
University of Pennsylvania	96	87	90	26	40
California Institute of Technology	92	96	82	18	41
Brown University	97	87	93	17	43
Massachusetts Inst. of Technology	97	94	91	19	43
Duke University	97	89	92	28	45
Harvard University	96	94	97	11	47
University of Notre Dame	98	80	95	35	48
Yale University	98	93	94	16	49
Dartmouth College	96	90	94	21	52
Princeton University	99	92	96	11	66