An Introduction to Statistics

~ Part 2 ~

An Introduction to Statistics

~ Part 2 ~

Discussed in Part 1

- What a statistician does
- What *is/are* statistics
- Data and information in data tables
- Some of our terminology

An Introduction to Statistics

~ Part 2 ~

To be considered in Part 2

- Variables
- Level of measurement
- Units of measure

An Introduction to Statistics

~ Part 2 ~

To be considered in Part 2

- Variables
- Level of measurement
- Units of measure

as well as

- How to "do" Statistics "right"

Variables

Variables

- Categorical or Qualitative

- Quantitative

Variables

- Categorical or Qualitative variable
  - is a classification, categorization, or representation

- Quantitative variable
  - is a measurement, amount, or count

Variables

- Categorical or Qualitative variable
  - is a classification, categorization, or representation
    - May be indicated by words such as "type", words, codes, or symbols
- Quantitative variable
  - is a measurement, amount, or count
    - May be indicated by words "amount" or "number of"

Variables

- Categorical or Qualitative variable
  - is a classification or categorization
    - Something that is a name (sub-classification – Nominal Variable), word, or code
  - Examples:
    - Name of country
    - Type of bill in your wallet
    - Brand of shoe
    - Class year
    - Area code
    - Social Security Number

Variables

- Quantitative variable
  - is a measurement, amount, or count
    - Can be indicated by words "amount" or "number of"
  - Examples:
    - Amount of money in your wallet
    - Height of building
    - Amount of gasoline in your gas tank
    - Amount of time spent studying
    - Number of bills in your wallet
    - Distance traveled from home to FSU

Variables

- CAUTION
  - A thing is not a variable

Variables

- CAUTION
  - A thing is not a variable

- *Time* is not the same as
  - *Amount of time*
  - *The time at this moment*

Variables

- CAUTION
  - A thing is not a variable

- *Time* is a thing.

- *Amount of time* is a measurement or a count.
- *The time at this moment* is a reading on a clock (i.e. a representation).

Variables

- CAUTION
  - A thing is not a variable

- *Time* is a thing.
  - *Time* is not a variable.
- *Amount of time* is a quantitative variable.
- *The time at this moment* is a qualitative variable.

Variables

- CAUTION
  - A thing is not a variable

- The *bills in your wallet* are things.

- The *number of bills in your wallet* is a count.
- The *type of the bills in your wallet* is a classification.

Variables

- **CAUTION**
  - A thing is not a variable

- The *bills in your wallet* are things.
  - The *bills in your wallet* is not a variable.
- The *number of bills in your wallet* is a quantitative variable.
- The *type of the bills in your wallet* is a qualitative variable.

Variables

- CAUTION
  - A thing is not a variable
- Shoes are things.
- The *number of shoes* is a count.
- The *type of shoes* is a classification.

Variables

- **CAUTION**
  - *A thing is not a variable*

- **Shoes are things.**
  - *Shoes* **are not a variable.**
- **The *number of shoes* is a quantitative variable.**
- **The *type of shoes* is a qualitative variable.**

Level of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

Level of Measurement

- Nominal – the values of the variable name, label or categorize

Nominal Level of Measurement

- Examples:
  - Name of country
  - Brand of clothing
  - Class year
  - Gender
  - Color*

*Caution: For an artist, color is a thing.

Level of Measurement

- Ordinal – the values of the variable are can be arranged in a ranked or specific order
  - A ranking or rating

Level of Measurement

- Ordinal – the values of the variable are can be arranged in a ranked or specific order
  - A ranking or rating

- CAUTION: Numbers have order, however, not all numbers correspond to *rankings* or *ratings*.

Ordinal Level of Measurement

- Examples:
  - J.D. Powers customer service rating for cellular phone service
  - National Highway Traffic Safety Administration rating for car safety
  - Course grade
  - Customer service rating for BestBuy Geek Squad Electronic Service

Ordinal Level of Measurement

- Examples:
  - J.D. Powers customer service rating for cellular phone service
  - National Highway Traffic Safety Administration rating for car safety
  - Course grade
  - Customer service rating for BestBuy Geek Squad Electronic Service

Note: These variables could be qualitative or quantitative.

Level of Measurement

- Interval –
  - the values of variable have properties of ordinal level of measurement
  - differences in values of variable have meaning
  - the value zero does not mean the absence of the quantity
  - operations (addition and subtraction) can be performed on values of the variable

Level of Measurement

- Interval –
  - the values of variable are ordered
  - differences in values of variable have meaning
  - the value zero does not mean the absence of the quantity
  - operations (addition and subtraction) can be performed on values of the variable

Interval Level of Measurement

- Examples:
  - Altitude
  - Temperature
  - Time of day measured/represented in military time
  - Year (i.e. birth year, year of graduation, or year college/university founded)

Level of Measurement

- Ratio –
  - the values of the variable have properties of interval level of measurement
  - ratios of the values of the variable are meaningful
  - the value zero means the absence of the quantity
  - operations (multiplication and division) can be performed on values of variable

Level of Measurement

- Ratio –
  - the values of the variable are ordered, and sums and differences of the values of variable are meaningful
  - ratios of the values of the variable are meaningful
  - the value zero means the absence of the quantity
  - operations (multiplication and division) can be performed on values of variable

Ratio Level of Measurement

- Examples:
  - Distance from home to FSU
  - Amount of coffee in a cup
  - Number of students in the session
  - Amount of money in your wallet
  - Speed of car

Quantitative Variables have *Units of Measure*

Units of Measure for Quantitative Variables

- *Units of measure* a.k.a. *units*
- *Units* are quantities or amounts adopted as standard measurements

Units of Measure for Quantitative Variables

- *Units* are quantities or amounts adopted as standard measurements

- Examples
  - Dollars
  - Hours
  - Years
  - Feet

Quantitative Variables

- Quantitative variables can be
  - Discrete
  - Continuous

Quantitative Variables

- *Continuous* variable – a quantitative variable that has an infinite number of possible values that are not countable

Quantitative Variables

- *Continuous* variable – a quantitative variable that has an infinite number of possible values that are not countable
  - Note: Suppose a continuous variable can take on all values between 0 and 1, inclusive. Then, the values 0.9, 0.99, 0.999, …, 0.9999999999999, etc., are all possible values for the variable

Quantitative Variables

- *Discrete* variable – a quantitative variable that has either a finite number of possible values or a *countable* number of possible values

Quantitative Variables

- *Discrete* variable – a quantitative variable that has either a finite number of possible values or a *countable* number of possible values
  - Note: We say that the values for a variable are countable *if you can count all the possibilities* ...

Quantitative Variables

- *Discrete* variable – a quantitative variable that has either a finite number of possible values or a *countable* number of possible values
  - Note: We say that the values for a variable are countable *if you can count all the possibilities*, for example, 1, 2, 3, 4, 5, ….

Quantitative Variables

- *Discrete* variable – a quantitative variable that has either a finite number of possible values or a *countable* number of possible values
  - Note: We say that the values for a variable are countable *if you can count all the possibilities*, for example, 1, 2, 3, 4, 5, …. We can say that there are 0 values as well if there are no possible values for the variable.

Quantitative Variables

- *A quantitative variable is discrete if the variable involves a count*
  - Examples:
    - The number of students
    - The number of bills in your wallet
    - The amount of money in your wallet
    - The number of minutes
    - The number of shoes

Quantitative Variables

- *A quantitative variable is continuous if the variable involves a measurement*
  - Examples:
    - Height
    - Distance traveled
    - The amount of cereal in the box
    - Amount of gasoline in your gas tank
    - The weight of a box

Cautions

- Important to consider what values *mean* and how they are being *used*

Cautions

- Important to consider what values *mean* and how they are being *used*
  - Do not treat all numerical data as quantitative

Cautions

- Important to consider what values *mean* and how they are being *used*
  - Do not treat all numerical data as quantitative
    - *A zip code* is a number but is neither a measurement nor a count
    - *A* phone number is neither a measurement nor a count
    - *Hair color* according to a five-point scale is neither a measurement nor a count
    - *A* social security number is neither a measurement nor a count

- For the following two examples,
  - Determine the variables;
  - Classify each variable as Qualitative or Quantitative;
  - Determine the level of measurement for each variable;
  - Classify each Quantitative variable as continuous or discrete;

and

  - State the units of measure for each Quantitative variable.

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/listofoecdmembercountries-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

Cautions

- Use meaningful identifiers for
  - Variables in data table
    - May not remember what heading represents later
    - Keep a code book for data for software that does not allow many characters for column names
    - Make sure that you have the context for data
- Be skeptical
  - Information, context, and testing questions can be slanted

To "do" Statistics "right"

- **Think** about the scenario that you are examining
- **Show** through calculating statistics and making necessary displays (graphs and tables)
- **Tell** what you have learned through your analysis so that ***anyone can understand*** (meaningful interpretation)

Caution

- In statistics, there can be *more than one "right answer"* since a statistician *interprets* the data and the statistics.
- Important for "right answer"
  - Analysis
  - Explanation
  - Tools
    - Graphs/tables
    - Calculations