Summary for

Intro to Statistical Research Methods

Meaning of Variable

A characteristic, number, or quantity that increases or decreases over time, or takes different values in different situations.

Types of variables

Variables can be divided for administrative purposes to:

Independent variables:

They are responsible for the creation of variables change in the phenomenon.

Dependent Variables:

That change as a result of the change in the independent variables.

What is a Research Study?

A research study consists in executing an analytical study in order to prove or disprove a hypothesis, or answer a specific question.

A research study must:

Ask a question

Identify a research target

Describe a research approach

Test or measure data obtained

Analyze the results

Valid research study?

What are the most important things you need to consider in order to conduct a valid research study?

· A good sample size (how many people participated in the survey)

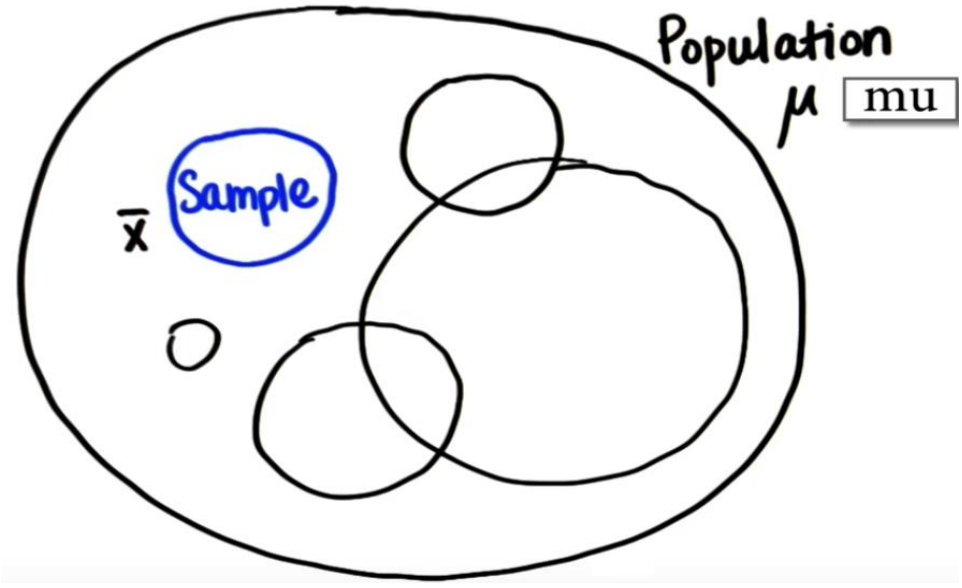
· A representative sample (who was surveyed)

· A sound methodology (methods used to survey)

Definitions:

Constructs are things for which defining a single correct way to measure them by is difficult (Ex: memory, happiness, guilt, intelligence, effort, hunger). Constructs can be measured by defining an **operational definition**.

Extraneous Factors (lurking variables) are unwanted variables that, if unaccounted for, could affect the data subsequently collected and affect the outcome of the experiment. Test results are more reliable if the same conditions are applied for all the test takers.



Population parameters (mu — see Notations and Symbols) are values that describe the entire population.
Sample statistics (X-bar — see Notations and Symbols) are values that describe our sample

We use statistics to estimate the population parameters by estimation. So, we use X-bar to estimated mu. The bigger the sample, the closer the population parameter.

Example: Determine what the key terms refer to in the following study. We want to know the average (mean) amount
Of money first Year College students spend at ABC College on school supplies that do not include books. We
Randomly surveyed 100 first year students at the college. Three of those students spent $150, $200, and $225, respectively.

Solution:
The **population** is all first year students attending ABC College this term.
The **sample** could be all students enrolled in one section of a beginning statistics course at ABC College (although this sample may not represent the entire population).
The **parameter** is the average (mean) amount of money spent (excluding books) by first year college students at ABC College this term.
The **statistic** is the average (mean) amount of money spent (excluding books) by first year college students in the sample.
The **variable** could be the amount of money spent (excluding books) by one first year student. Let X = the amount of money spent (excluding books) by one first year student attending ABC College.
The **data** are the dollar amounts spent by the first year students. Examples of the data are $150, $200, and $225.

Example of a research study:

Measuring memory

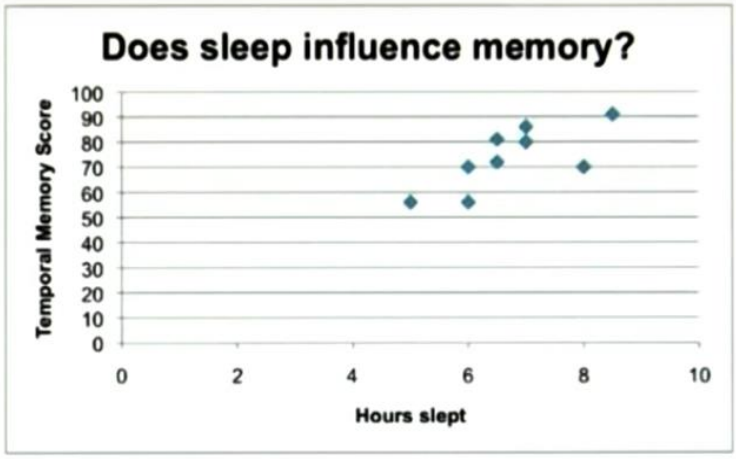
In my opinion, measuring memory first depends on which type of memory you want to measure: short-term memory or long-term memory, usually by giving an individual something to memorize and testing him after a period delta time Δt to see how much (and which parts) of that information his brain has retained.

You could measure memory and draw conclusions by taking into account the difficulty and size of the information given, the delta time before quizzing the individual, the percentage of correct and incorrect answers given by the individual as well as other external factors (age, language barriers, time of day, time spent to take the test, etc).

BBC came up with a test to measure memory by how well the test taker can remember faces. Users were shown 24 photos in 2 parts with a 5 minutes break in between and in the 3rd step they were asked to pick and identify from a sample of 48 photos the ones they saw in the 1st and 2nd part respectively. After the test, memory was measured in two ways: one by computing a “recognition score” — the percentage of correctly identified seen photos and a “temporal memory score” — the percentage of correctly identified photos placed in their correct part.

The operational definition for the BBC test is the percent of faces correctly recognized and placed from parts I and II

Hours Slept	Temporal Memory
7	86
8	70
6	56
5	56
6	70
7	80
6.5	72
8.5	91
6.5	81
7	86



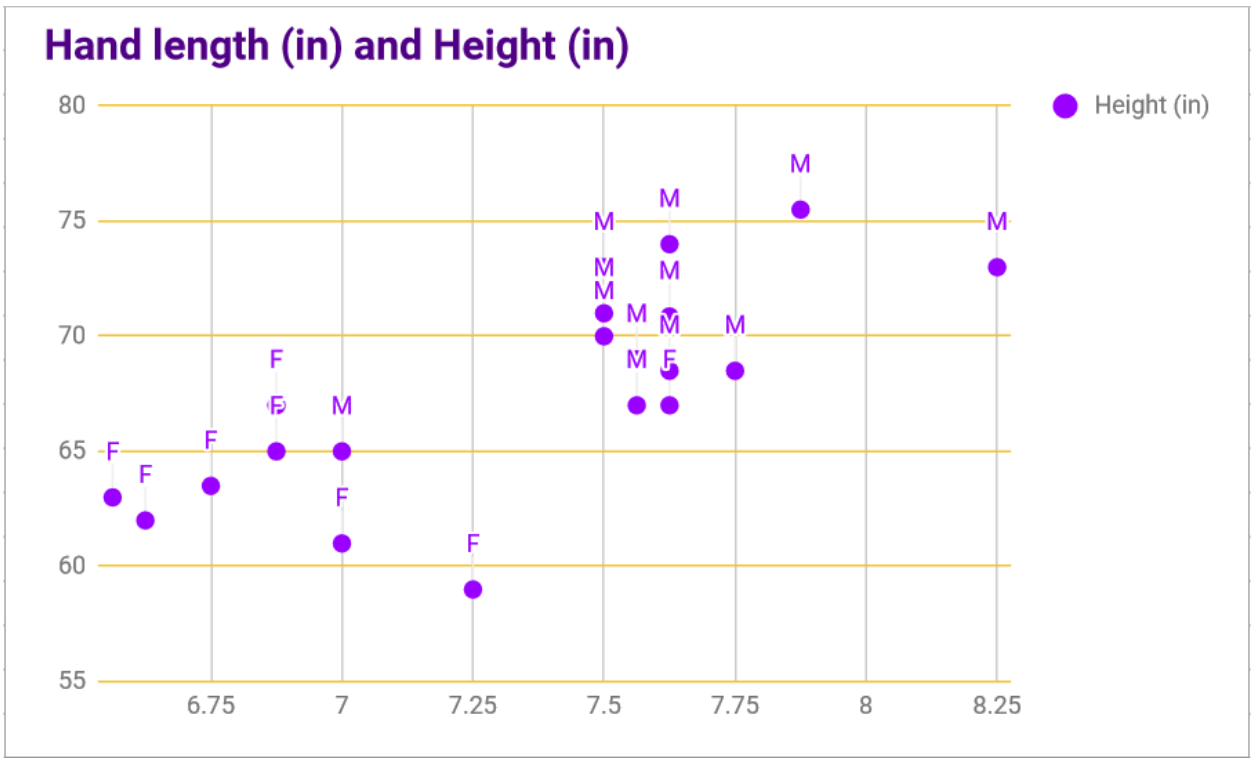
Outcomes gathered from the BBC research concluded that people who sleep more tend to have a better memory, the time of day the tester took the test influenced the results and age is also an important factor in memory.

Randomness:

Perhaps the best typical example that we can give of the scientific meaning of random distribution is afforded by the arrangement of the drops of rain in a shower. No one can give a guess whereabouts at any instant that a drop will fall, but we know that if we put out a sheet of paper it will gradually become uniformly spotted over; and that if we were to mark any two equal areas on the paper these would gradually tend to be struck equally often.

-John Venn

Random samples are more likely to approximate a reality of the population.



Hand length vs Height, plot done in Google Spreadsheets

For the final quiz, to check our plotting knowledge, the instructor challenged us to find out her possible height, given a set of information regarding the correlation between hand length and height, all measured in inches. According to the samples and observing the trend, the instructor's height should be about ~63.5 inches and people with longer hands tend to be taller.

Notations and Symbols:

μ (mu) = Population average

$\bar{x}$ = Sample average

$\mu - \bar{x}$ = sampling error

Conclusions and final thoughts

This was an excellent introduction to research methods, research influencers and the basic principles data should be gathered by.

This is the first step everybody looking to set foot into the data science field should take before delving deeper.

I am very much looking forward to the next level and what they teach us next!