Sept. 19

1) Jupyter server → start program... → new r studio.

Statistics

Descriptive Statistics
- Describing some population.

Inferential Statistics
- want make inference about a population using a sample

population parameters:

- μ: population mean.
- p: population proportion
- σ: population standard deviation
- σ^2: population variance

population: refer to all objects in a given set.

Ex:
- All university students

sample: refers to a subset of all objects in a given set.

- A sample of randomly selected U of C students

Sample Statistics

- $\bar{x}$: sample mean
- $\hat{p}$: sample proportion
- s: sample standard deviation.
- s^2: sample variance

Random Variable (r.v.'s): A variable with a set of outcomes, that have associated probabilites, and varies randomly.

Numeric / Quantitative r.v.'s

Ex:
- # of 1's that appear when rolling 3 dice
- # of children Khaled saves in a given week

Discrete
- takes on a 'countable infinite' number of values.

Ex: # of job applications Niko makes in a given week.

Continuous
- takes on a 'non-countable' infinite number of values.

ex:
- Height of children Khaled saves.

Categorial / Qualitative r.v.'s

Nominal
- No meaningful ranking system
- colour.
- type

Ordinal
- Categorical variables that can be meaningfully ranked/ordered.

Ex:
- Race placement.
- Likhart scales
- Things that involve preference.

How do we gather data in a dataset

'Tidy Data': Refers to organizing data such that observations form rows, and random variables form columns.

Ex: Pokémon Dataset