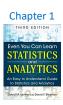
IMGD 2905

Fundamentals of Statistics



Why Do We Need Statistics?



Aggregate data into meaningful information.

 $\overline{x} = \dots$

Ok, but what are statistics?

Key Words

- Population all members of group pertaining to study
 - e.g., every person in IMGD 2905 in D-term
- e.g., every Heroes of the Storm player in the world
- In many cases, *impossible* to survey a population!
 - Typical for game analytics → want to understand/improve game for all

So ... what to do?



Key Words

- Sample part of population selected for analysis
 - e.g., all League of Legends players at WPI
 - e.g., students in first row in IMGD 2905



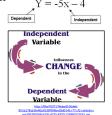
- Often hope sample is representative of population. ...
- (e.g., poll: "did you finish chart for part 3?")
- But Is it? → the method to obtain a sampling is important! (We won't talk much about this right now, however.)

Key Words

- Variable characteristic of individuals in population
 - e.g., time spent in competitive mode in Starcraft 2
 e.g., vehicle choice in Grand Theft Auto (GTA)
- independent variable is inherent in population, versus dependent variable that want to asses

RANCOMASSICAMENT

Study.com



Key Words

- Observation all variable values for sample
 - e.g., League of Legends competitive hours/week and Champion most played could be (2 observations)

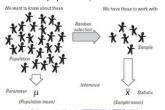
"Player A: Leona, 2 hours"

- "Player B: Teemo, 7.5 hours"
- Can be continuous (time) or discrete (Champions)
- Often, data in grid
 - Observation in rows
 - Variables in columns
- Player Hours Champ A 2 Leona B 7.5 Teemo
- Consider our project 1 → FIFA 18 data!

Key Words

- Parameter measure of dependent variable for population
 e.g., average crashes in *Mario Cart* level for everyone

 - Usually what we want to know, but can't get easily
 - Statistic measure of dependent variable in sample
 - e.g., average crashes in Mario Cart level for IMGD 2905 class
- Statistics set of numerical methods for getting information about population based on data from sample, usually to get information about population parameters



"Statistics - a branch of mathematics dealing with the collection, analysis, interpretation and presentation of masses of numerical data." Merriam-Webster dictionary

Sources of Data

- Published generally made available from those that collected it
 - e.g., Riot's League of Legends data
 - e.g., Metacritic reviews and ratings
 - e.g., Kaggle dataset on FIFA 18
- Experiments multiple trials to collect data
 - Can be in laboratory or "real world" setting
 - e.g., play shooter, add lag and play again
- Survey ask people to answer questions
 - e.g., self-rating as gamer, difficulty with level, \ldots
- Ethical issues with stress and use of data ightarrow Institute Review Board (IRB) for approval with human subjects



Sampling Concepts

- $\label{eq:sampling-process} \textbf{Sampling-process by which members of population are selected for sample}$
 - e.g., choose ½ class based on spacing, or choose ½ class based on alphabet
- Probability sampling sampling considering likelihood of selection
- e.g., survey for intended Champ, ask ½ class, but when tournament starts, result different. Why?

 sample didn't consider League players! (e.g., often similar analogy for voter polls)
- e.g., voluntary polls/surveys
- Use probability sampling whenever possible, but sometimes it is not (cost) or not known
- Sampling with replacement once sample, put back in pool
 - e.g., die roll to see which attack boss makes
- Sampling without replacement once sample, won't sample again
 - e.g., user survey don't allow to submit twice
 - E.g., deck of 52 cards for blackjack

Using Sample Data

- · Word "sample" comes from same root word as "example"
 - Similarly, one sample does not prove a theory, but rather is an example
- Basically, in general, definite statement cannot be made about characteristics of all systems
- Instead, make probabilistic statement about range of most systems
- → That's where statistics come in!