Lecture 08
Sampling

CHOONG-WAN WOO | COCOAN lab | http://cocoanlab.github.io

Review: Key Points

Chapter 9 and 10: Regression wisdom, re-expressing data

- Bends, subgroups, outliers

- Cautious about extrapolation, causation, lurking variables, summary stats

- Tukey's ladder

Randomness

- Good videos:

- https://www.youtube.com/watch?v=9rIy0xY99a0

- https://www.youtube.com/watch?v=sMb00lz-IfE

What is Random?
Vsauce · 5.5M views · 3 years ago
There's more over on Veritasium! "What is NOT Random?": https://www.youtube.com/watch?v=sMb00lz-IfE SOURCES AND ...
CC
11:12

What is NOT Random?
Veritasium · 3M views · 3 years ago
Is the future of the universe already determined? **Vsauce** tackles "What is Random?": https://youtu.be/9rIy0xY99a0 Special Thanks ...
CC
10:00

Sampling

- Idea 1: Examine a part of the whole

    - **Population**: the entire group of individuals

    - **Sample**: a smaller group of individuals, selected from the population

    - difficult to ensure the sample **represents** the population

    - **Bias:** over- or under-represent some characteristics of the population (e.g., Literary Digest Poll, they sampled response using phone calls when telephone was a luxury)
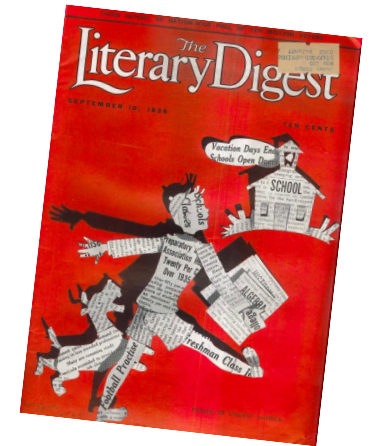
Sampling

- Idea 1: Examine a part of the whole

- Idea 2: Randomize

  - To make sure the sampling is not biased

Sampling

- Idea 1: Examine a part of the whole

- Idea 2: Randomize

- Idea 3: Sample size matters

    - Important thing is not the fraction of the population, but the sample size.

    - *Census*: examine the entire population

        - It is difficult to complete a census

        - More noisy

        - They do not stand still

Populations and parameters

- **Population parameters:** parameters to model for a population

- **Sample statistics (or statistics):** summaries of sample data to estimate the population parameters

Name	Statistic	Parameter
Mean	$\bar{y}$	μ (mu, pronounced "meeoo," not "moo")
Standard deviation	s	σ (sigma)
Correlation	r	ρ (rho)
Regression coefficient	b	β (beta, pronounced "baytah"[7])
Proportion	$\hat{p}$	p (pronounced "pee"[8])

Sampling methods

Simple Random Sampling

- Each person has an equal chance of being selected

Stratified Random Sampling

- **Strata:** the population is first divided into homogeneous groups (e.g., male and female)

- Then, simple random sampling within each stratum before the results are combined.

  - E.g., 60% men and 40% women in the campus: randomly sample 60 men and 40 women for 100 subjects

Sampling methods (cont'd)

Cluster Sampling

- Splitting the population into *representative* clusters, first.

- Then, select one or a few clusters at random, and do census.

- What's different from stratified sampling? Clusters can be heterogeneous!

Multistage samples

- combine several sampling methods

Systematic Samples

- select samples systematically (e.g., survey every 10th person on the list)

Common Mistakes in sampling

Mistake 1: Sample Volunteers: voluntary response bias

Mistake 2: Sample Conveniently: may not be representative of the population

Mistake 3: Use a Bad Sampling Frame: incomplete sampling frame introduces bias

Mistake 4: Undercoverage: some portion of the population is not sampled at all

Mistake 5: Nonresponse Bias: those who don't respond may differ from those who do

Mistake 6: Response Bias: anything survey design that influences the response

E.g., Sampling bias in pain research

https://ppw.kuleuven.be/ogp/anderepdf/efic-2017.pdf

Key Points

Chapter 11 and 12: Randomness, Sampling

- Population parameters vs. sample statistics

- Sampling methods:

  - Simple sampling, stratified sampling, cluster sampling, multistage sampling, systematic sampling

- Common mistakes in sampling...