SKKU Biostats and Big data

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

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

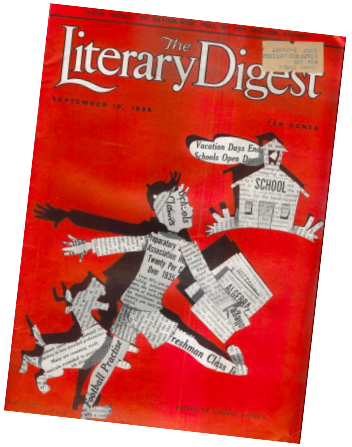
- Good videos:
- https://www.youtube.com/watch?v=9rIy0xY99a0
- https://www.youtube.com/watch?v=sMb00lz-IfE

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

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

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])

Quiz 08-1

https://forms.gle/HZY465DKcvQYEZyMA

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 10^{th} person on the list)

Quiz 08-2

https://forms.gle/KZQ9bE6uVC5DsJSs9

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

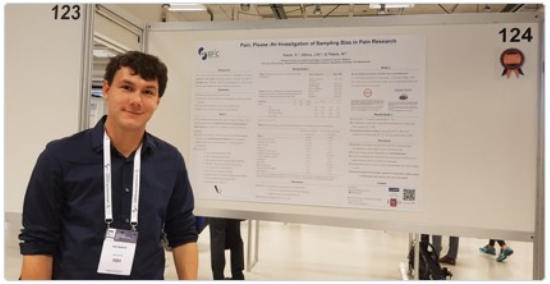
Mistake 6: Response Bias: anything survey design that influences the response

E.g., Sampling bias in pain research

Choong-Wan Woo 님이 리트윗했습니다

Kai Karos @KaiKaros 5d

Why on earth would people participate in #pain research? Poster 124 or download pdf at ppw.kuleuven.be/ogp/anderepdf/... #efic2017 #efic @EFIC_org



14 35

Pain, Please: An Investigation of Sampling Bias in Pain Research

Karos, K.[1], Alleva, J.M.[2], & Peters, M.[2]

1 Research Group on Health Psychology, University of Leuven, Belgium
2 Faculty of Psychology, Department of Clinical Psychological Science, Maastricht University, The Netherlands

Background

Pain research often relies on the recruitment of **volunteers** and involves **unpleasant and painful** sensations, making it especially susceptible to **sampling bias:** Volunteers differ from nonvolunteers in relevant ways, affecting generalizability and external validity of pain research.

Hypotheses

Participation in pain research is associated with:

1) Lower levels of fear of pain, pain catastrophizing and illness and injury sensitivity
2) Lower levels of depression, anxiety and body appreciation
3) Higher levels of sensation seeking and social desirability

Study 1

N = 275 healthy participants (63 male, *M*age = 20.48 years, *SD*age = 2.18 years) were asked about the **likelihood that they would participate in pain research** in an online survey in addition to several questionnaires.

Procedure

- Likelihood to participate in different kinds of research: Brain Imaging, computer tasks, **painful/unpleasant stimuli**, food consumption, physical activity, surveys, and medication (1 = *extremely unlikely* to 7 = *extremely likely*)
- Questionnaires:
  - Fear of Pain Questionnaire (FPQ)
  - Pain Catastrophizing Scale (PCS)
  - Illness-Injury Sensitivity Inventory-Revised (IISI-R)
  - PROMIS – Depression and Anxiety Short Forms
  - Body Appreciation Scale-2 (BAS-2)
  - Brief Sensation Seeking Scale (BSSS)
  - Balanced Inventory of Desirable Responding-6 (BIDR-6)
  - Personal Attributes Questionnaire (PAQ)
  - Demographics (sex, age, etc.)

Results Study 1

Table 1 Likelihood Estimates for the Different Types of Research

Type of Research	Mean (SD)
Painful / unpleasant sensations	2.87 (1.43)
Computer tasks	4.85 (1.10)
Brain imaging	4.25 (1.45)
Food consumption	4.60 (1.32)
Physical activity	3.79 (1.35)
Surveys	5.24 (0.93)
Medication	2.45 (1.45)

Table 2 Summary of Backwards Multiple Regression Analysis for the Final Variables Predicting Perceived Likelihood of Participating in Pain Research

Predictors	R^2	ΔR^2	$F(\Delta R^2)$	*B*	SE *B*	β	*p*
Model 1	.101	.088	7.586**				
Age				.091	.038	.138*	.017
Sensation seeking				.281	.117	.139*	.017
Pain catastrophizing				-.018	.009	-.128*	.044
Fear of pain				-.012	.005	-.155*	.015

Note: * p < .05. ** p < .01.

Table 3 Self-reported Reasons Provided for Participation / Non-participation in Pain Research

	More likely to participate *n* = 84 (30.5%)	More unlikely to participate *n* = 191 (69.5%)
Reason	***n* (%)**	***n* (%)**
Positive prior experiences	34 (40.5)	3 (1.6)
Personal growth / curiosity	19 (22.6)	0 (0)
Financial reward	11 (13.1)	1 (.5)
Indifference	8 (9.5)	2 (1)
Availability of other research	5 (6)	4 (2.1)
Avoidance of (unnecessary) harm	3 (3.6)	118 (61.8)
Low pain sensitivity / tolerance	2 (2.4)	5 (2.6)
Societal gain	1 (1.2)	2 (1)
Fear of pain	1 (1.2)	38 (19.9)
Bad prior experiences	0 (0)	13 (6.8)
Preexisting medical condition	0 (0)	3 (1.6)
Social situation	0 (0)	2 (1)

Note: Groups are created based on likelihood score. Participants scoring 4 or above are categorized as more likely to participate, whereas participants scoring 3 or less are categorized as more unlikely to participate in pain research.

Discussion

- Likelihood to participate in pain research is associated with:
  - Lower fear of pain and lower pain catastrophizing
  - Higher sensation seeking and older age
- Possibility for sampling bias in pain research

Study 2

Do the likelihood estimates translate into actual behaviour?

N = 87 healthy participants (11 male, *M*age = 21.10 years, *SD*age = 8.85 years) chose between two identical studies, one involving painful stimuli and the other involving neutral stimuli.

Results Study 2

- Pain group (*n* = 36) and no-pain group (*n* = 51)
- No differences on age, fear of pain, pain catastrophizing or illness and injury sensitivity ($p > .230$)
- **Increased sensation seeking** in pain group (M=3.5, SD=0.53) compared to no-pain group (M=3.21, SD=0.6) ($t(85)$=2.349, p=.021)

Discussion

- **Intention** to participate associated with lower fear of pain and pain catastrophizing as well as higher sensation seeking and age
- Sensation seeking was the only predictor of **actual behaviour**
- Sensation seeking has been associated with reduced pain sensitivity, better coping with pain, and increased placebo responding
- Sampling bias in pain research can affect generalizability and external validity of results

Volunteers in pain research might represent an especially resilient subset of individuals, who actively seek out novel stimuli

Contact

Kai Karos
Kai.Karos@kuleuven.be
@KaiKaros

Research Group on Health Psychology
Tiensestraat 102
3000 Leuven, Belgium

Scan for PDF

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

Quiz 08-3

https://forms.gle/Rh8FxcRobCWMzZv38

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…