

Topic 6: Human Evaluation

Survey Sampling

What is it?

- The proper way to select your target audience
- Selection of participants for a **population**
- **Census**: Surveying all members of a population
- In most cases, this is **not possible**

Key features

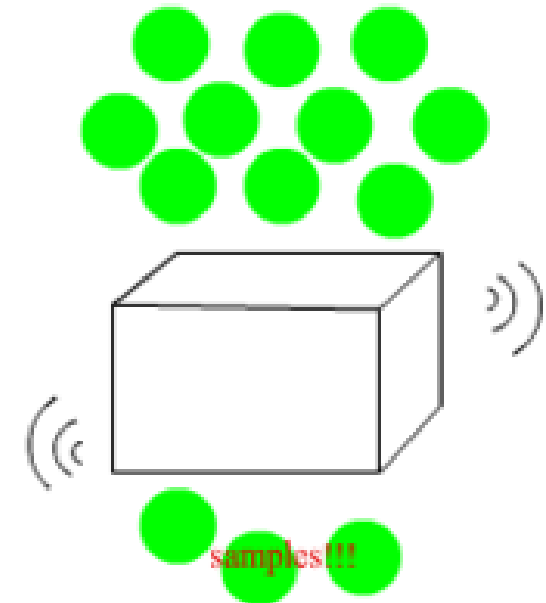
- Representativeness
- Flexibility
- Effectiveness
- Consistency
- Diversity
- Transparency
- Probabilistic or non-probabilistic?

Probabilistic Sampling

- Every individual of the population has a non-zero chance of being selected
- Ensures representativeness
- Three sub-categories:
 - Random
 - Systematic
 - Stratified

Random Sampling

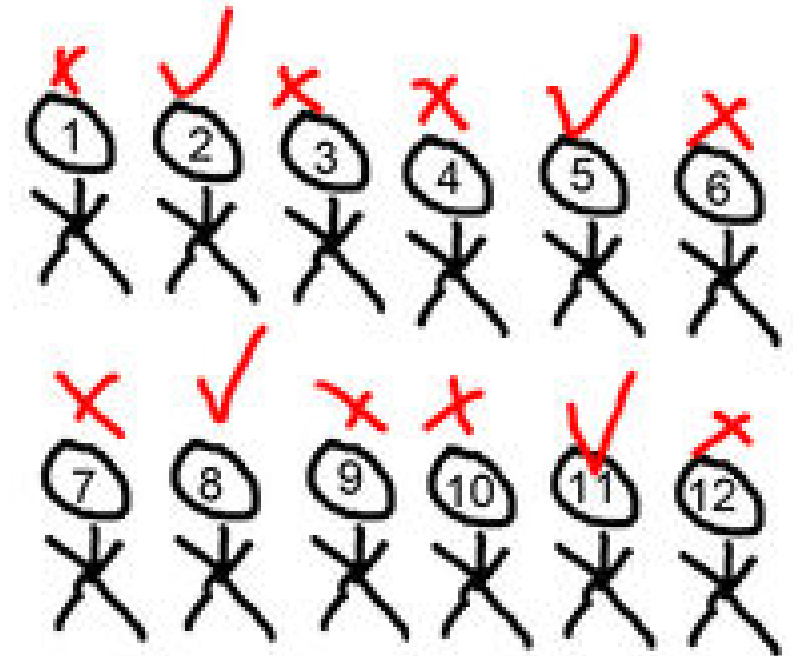
- Used when identifying characteristics is (almost) impossible
- Sample selected independently of others
- Equal chance of being selected as subject when sampling
- **Advantage:** Reduce bias
- **Issue:** It may be biased already without you knowing! (e.g. phone polling)



https://alevel-sociology.fandom.com/wiki/Simple_random_sample

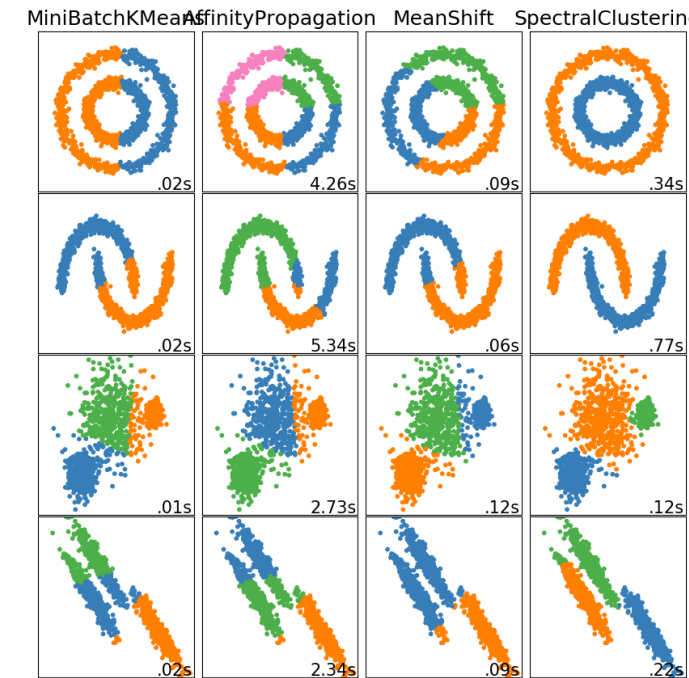
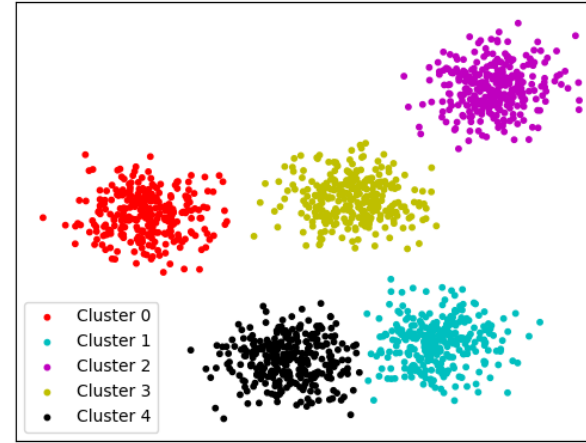
Systematic Sampling

- The proposed location is logically homogenous
- First decide sample size, then arrange elements to select members at regular intervals
- Good as random sampling if there is no hidden order
- **Issue:** Periodicity tends to create patterns
- **Solution:** Randomise before sampling



Stratified Sampling

- You divide the population into groups of characteristics (depending on focus)
- Then you sample within each category and select randomly
- **Issue:** More complex
- **Solution:** Machine learning!



Non-Probabilistic Sampling

- Samples are collected with no specific structure in mind
- Ensures practicality
- Three sub-categories:
 - Convenience
 - Snowball
 - Quota

Convenience Sampling

- Samples are selected based on availability and accessibility (to the test)
- Created rapidly without extra load
- **Issues:** Poor representativeness, **should only be used as an approach!**

Snowball Sampling

- Select an individual, and then this leads you to someone else.
- Used in academic research (not as you may think!)
- **Advantage:** Low cost, high relevance
- **Issues:** Homogeneity



Quota Sampling

- You need your sample to be of certain features
- (In a way) equivalent to stratified
- Superior to the previous two, but doesn't have any statistical insight

So which one is this?!



Hunger games, in case you didn't know

Planning Surveys

How to start?

- State a question, pose a problem
 - State the hypothesis
- Plan your study
 - What do you want to find?
 - Which is the basis of comparison?
 - How will you measure?
 - Which are the users (sample)
 - Which tool to use?
- You will need to ... results
 - Summarise
 - Analyse
 - Visualise
 - Interpret
 - Discuss

The central principle

- A good evaluation captures **qualitative and quantitative** data
- It establishes a proper **scale** for each!
- Which is best?
 1. The proper one to get the most reliable data
 2. Mix approaches

Scales

- The measurements that you choose influence on the participants' responses
- Types of scales:
 - Nominal: Used for tally
 - Membership, characteristics, etc
 - Ordinal: Same, but with an order
 - Never – Sometimes – Always
 - Interval: There is an unclear scaling
 - Never 1 2 3 4 5 6 7 Always
 - Ratio/Continuous: Score
- Retrofit?



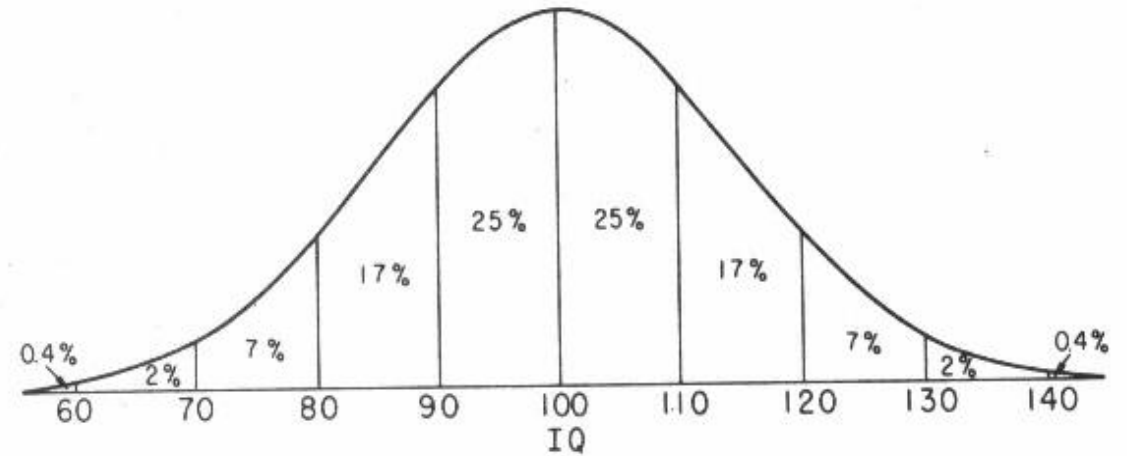
Statistics

Reporting Counts

- Nominal and ordinal counts have to be summarised
 - Tally each response
 - Average frequency of each category
 - You can discuss this!
- Interval
 - Summarised
 - Compared between levels (as if it were ordinal)
 - Stats are limited
- Ratio/Continuous
 - Summarised
 - Stats can be applied

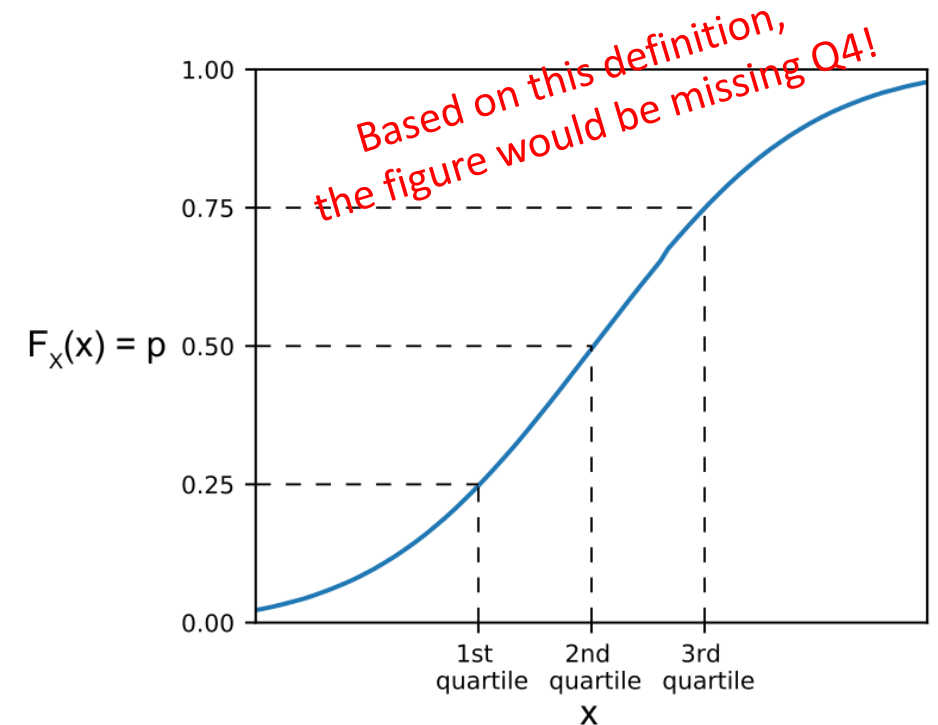
Counting

- Simplest form to measure things → frequency analysis
- Can be applied to different types of data
- You can convert to a **percentage**



Quartiles

- A type of quantile/percentile (i.e. way to split counts) which divides data points into four (more or less) equal parts
- Q1 would be the top 25%, Q2 the following 25% and so on...
- Used by the Journal Citation Reports ([JCR](#)) to see which are “the best” journals



Example from JCR

PATTERN RECOGNITION

ISSN: 0031-3203
eISSN: 1873-5142
ELSEVIER SCI LTD
THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND
ENGLAND

[Go to Journal Table of Contents](#) [Go to Ulrich's](#) [Printable Version](#)

TITLES
ISO: Pattern Recognit.
JCR Abbrev: PATTERN RECOGN

CATEGORIES
ENGINEERING, ELECTRICAL &
ELECTRONIC -- SCIE
COMPUTER SCIENCE, ARTIFICIAL
INTELLIGENCE -- SCIE

LANGUAGES
English

PUBLICATION FREQUENCY
12 issues/year

Journal Impact Factor Calculation

$$\text{2018 Journal Impact Factor} = \frac{4,276}{725} = 5.898$$

How is Journal Impact Factor Calculated?

$$\text{JIF} = \frac{\text{Citations in 2018 to items published in 2016 (1,839) + 2017 (2,437)}}{\text{Number of citable items in 2016 (320) + 2017 (405)}} = \frac{4,276}{725}$$

Example from JCR

Rank



JCR Impact Factor



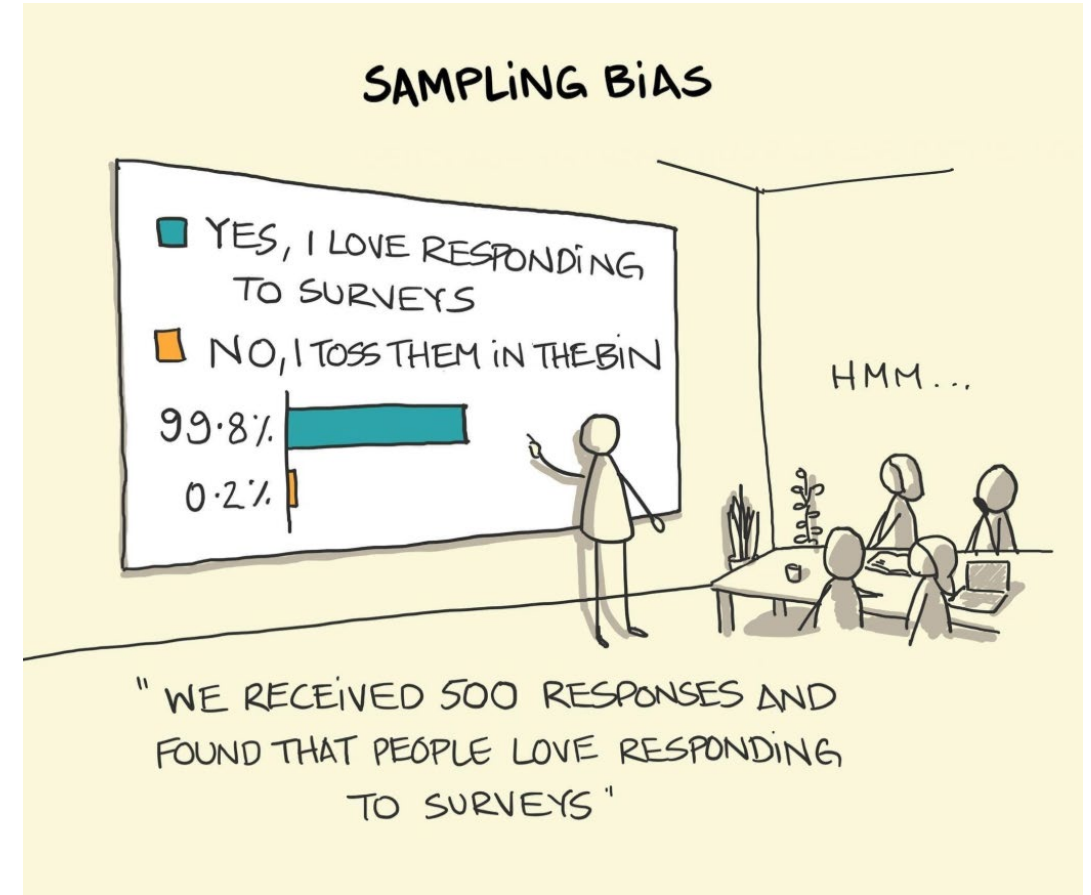
JCR Year ↕	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE			ENGINEERING, ELECTRICAL & ELECTRONIC		
	Rank	Quartile	JIF Percentile	Rank	Quartile	JIF Percentile
2018	14/134	Q1	89.925	25/266	Q1	90.789
2017	16/132	Q1	88.258	37/260	Q1	85.962
2016	15/133	Q1	89.098	23/262	Q1	91.412
2015	15/130	Q1	88.846	20/257	Q1	92.412
2014	15/123	Q1	88.211	20/249	Q1	92.169

Measures of Central Tendency

- Arithmetic mean
- Variance & Standard Deviation
- Median
- Mode

Final Tips

- Do I have sufficient data to extract the measures?
- Is my data of any distribution
- Is the data sufficiently skewed so that these are representative?



Demos (Lab): Statistics with Python & Excel
