SAMPLING TECHNIQUES IN SURVEY CONSTRUCTION AND DATA ANALYSIS: STRENGTHENING VALIDITY AND GENERALIZABILITY IN STATISTICAL PROCEDURES

DONALD L. GILSTRAP, PHD

DEAN AND PROFESSOR OF LIBRARIES

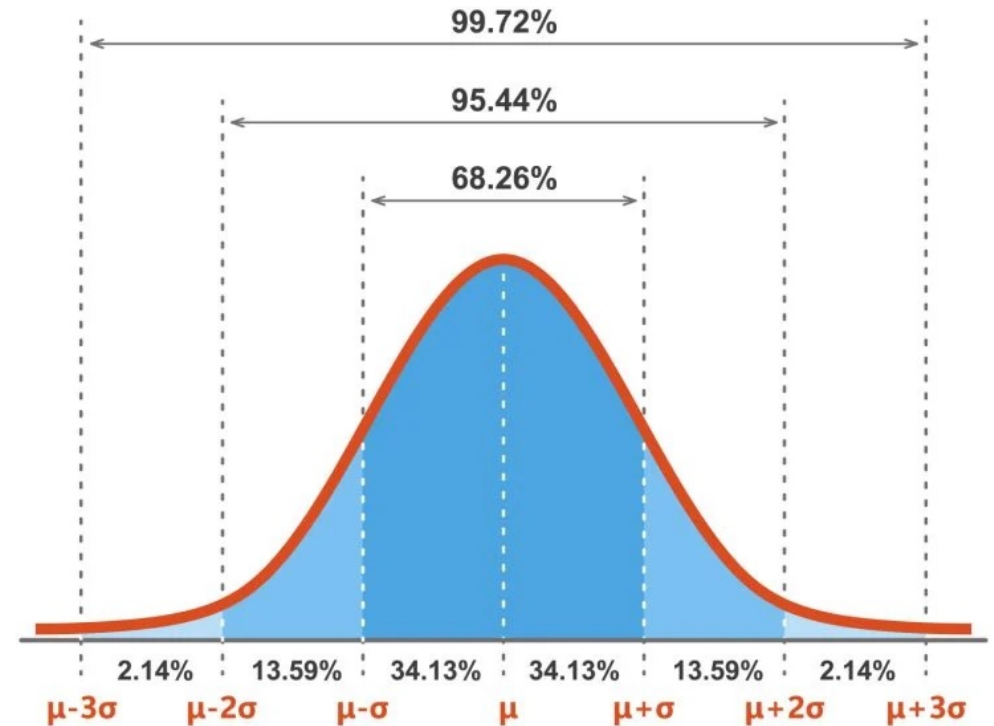PROFESSOR OF EDUCATIONAL LEADERSHIP, POLICY, AND TECHNOLOGY STUDIES

UNIVERSITY OF ALABAMA

DIFFERENT TYPES OF SAMPLING

All are reflections of the population to varying degrees

- Convenience Sampling

- Random Sampling

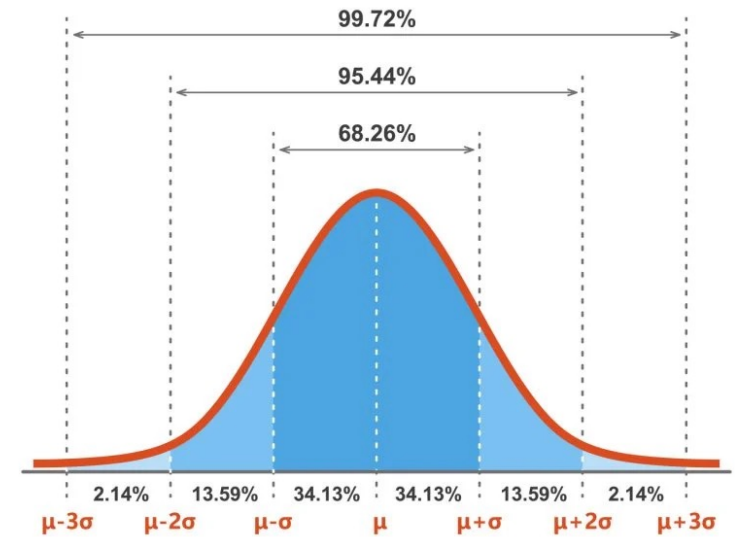- Stratified Sampling

- Clustered Sampling

CONVENIENCE SAMPLING

- Responses come from anyone willing to participate

- Pros
  - Simple
  - Easier to administer
  - Faster to receive responses
  - The most commonly used survey technique

- Cons
  - Confirmation Bias is pervasive - Responses reflect the tails of a normal curve more often
  - Need a large sample (>50%)
  - Considered "non-probable"
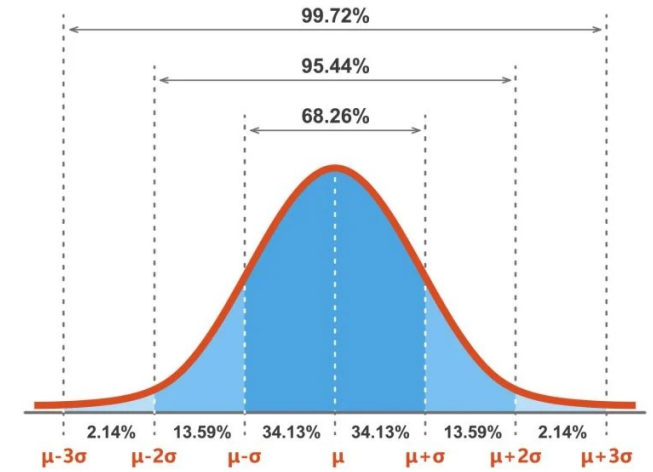 - Findings cannot be generalized to the population
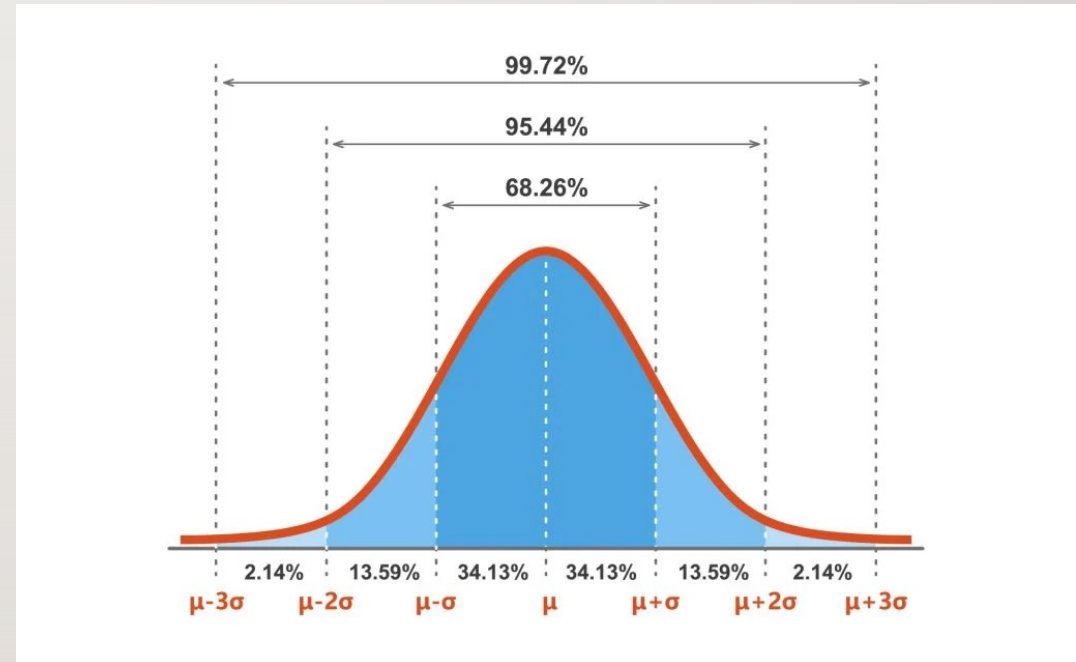
RANDOM SAMPLING

- Every *n*th person is asked to participate

- This sampling technique is the closest to a representation of the population

- Pros

  - Tests of significance tend to have have stronger confidence levels and confidence intervals

  - Smaller sample size (*n*) is needed, and this is computed *a priori* from the population (*N*)

  - Findings are strongest in survey analysis

- Cons

  - More difficult to administer

  - Takes more time to get people to respond -,Follow ups are often required
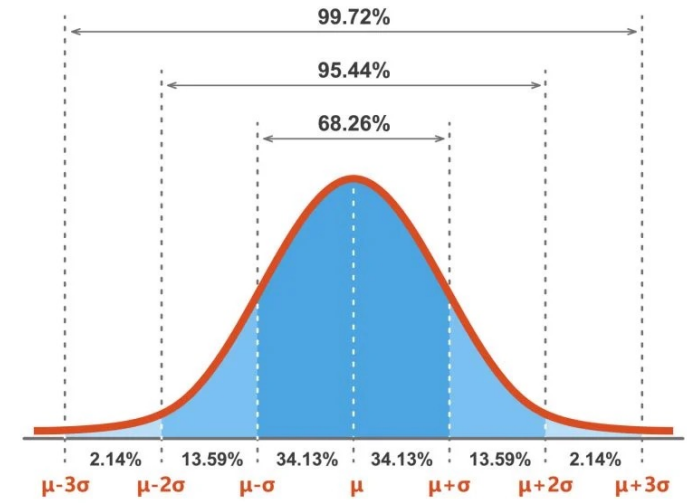
  - If response rate is not met, survey becomes invalid

STRATIFIED SAMPLING

- Every *n*th person is asked to participate from predefined, homogenous groups

- This sampling technique ensures respondents from all groups under focus are included

- Pros
  - This sampling technique is similar to random sampling but isolates groups based on certain characteristics
  - Smaller sample size (*n*) is needed, and this is computed *a priori* from the population (*N*)

- Cons
  - More difficult to administer
  - Takes more time to get people to respond
  - Becomes invalid if representative sample from homogenous group under focus does not respond

CLUSTER SAMPLING

- All persons are asked to participate from predefined, heterogeneous groups that are randomly sampled from the population

- This sampling technique is the closest to a representation of sub-populations within the population based on a non-homogenous characteristic

- Pros
  - A robust representation of the entire population
  - Differentiates by predefined groups

- Cons
  - More difficult to administer
  - Takes more time to get people to respond – follow ups are often required
  - Survey becomes invalid if response rate is not meant for each heterogeneous group

COMPUTING YOUR RANDOM SAMPLE SIZE

- https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=1000&x=62&y=16

- https://tinyurl.com/yu9u8uj3

DETERMINING YOUR CONFIDENCE INTERVAL WITH MARGIN OF ERROR

- https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=5&pp=50&ps=1000&x=62&y=16

- https://tinyurl.com/yu9u8uj3

INTER-RATER RELIABILITY

- Used to measure how closely different raters score items in assessing a qualitative response

- Raters score independently of each other

- Without two (or more) raters, scoring is very subjective

- A simple correlation coefficient is computed between the two raters' scores

- Example: *How well do you think your content knowledge improved after your library instruction class?*
  - *I felt like I came a way with a much better understanding of how databases work and how Boolean logic works where I can apply it in my searches*

- Scale = 1-5
  - Rater one = 4   Rater two = 5

CALCULATING CORRELATION COEFFICIENT

- https://www.statskingdom.com/correlation-calculator.html

USING EXCEL TO CACULATE IRR
USE THE CORREL FUNCTION

- https://support.microsoft.com/en-au/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92

- Click on a cell and type "=CORREL("

- Highlight the first column of data range and type ","

- Highlight the second column of data range and type ")"
  to end the function and hit Enter

- =CORREL(A2:A7,B2:B11)

- Correlation Coefficient r = .996783067

	A	B
1	X	Y
2	1	2
3	3	4
4	2	3
5	7	8
6	9	10
7	11	11

REFERENCES:

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). New York: Routledge.

- Stratton, S. (2021). Population Research: Convenience Sampling Strategies. *Prehospital and Disaster Medicine, 36*(4), 373-374. doi:10.1017/S1049023X21000649