

Pitfalls of Various Sampling Approaches & Statistical Analysis and Interpretation of Results

Troy D. Thomas, Ph.D

Dean
Faculty of Natural Sciences
University of Guyana

April 1, 2022

- 1 Outline
- 2 Revision of Sampling Techniques Seminar by Prof. A Ansari
- 3 Purpose of Sampling
 - Assuring Representativeness
 - Can Humans Select at Random?
- 4 Bias
 - Sources of Bias
 - The Problem of Bias
 - Addressing Researcher Bias: Pseudo-Random Number Generators
 - Bias in Various Sampling Approaches
- 5 Statistical Analysis & Interpretation of Results
 - Data Analysis Considerations
 - Interpretation of Results

Sampling Techniques

Prof. Ansari Addressed

- Sample
- Why sample?
- Sample frame
- Target population
- Probability and non-probability samples

Purpose of Sampling

- Support inferences about a population/ target population
 - ▶ Representativeness is key
 - ▶ Sampling procedure affects representativeness
 - ▶ Sample size? useful but not a good way to address representativeness
 - ▶ **A sample cannot speak for a group it does not represent!**
- Piloting
 - ▶ Test instruments, data collection procedures
 - ▶ Compute variance. Usually needed for calculation of sample size
 - ▶ **Sample size not an important consideration for piloting**

Assuring Representativeness

- Random selection is key
- Random
 - ▶ No discernable pattern in occurrence
 - ▶ lack of predictability
 - ▶ many variables in nature are random
- Why randomness?
 - ▶ Elimination of bias
 - ▶ Bias results in discernable patterns related to some extraneous variable/source
 - ▶ Inclusion under random selection is entirely determined by chance

Can Humans Select at Random?

- Can we select at random?
- Survey results on selection

Sources of Researcher Bias

- Conscious and unconscious sources of bias
- What might account for my “random” *wink wink*, selection of five students from the class?
- Random vs miscallenous Facebook photos

The Problem of Bias

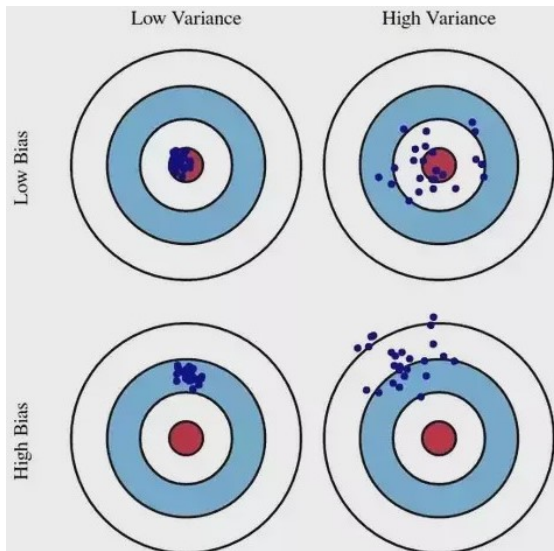


Image: <https://towardsdatascience.com/bias-variance-tradeoff-e8995c42b55b>

The Problem of Bias

- Research results do not have the expected meaning
- Statistics do not approximate the parameters of the target population
- The larger the bias the less valid the results
- A combination of sources of bias can affect the same set of results

Addressing Researcher Bias: Pseudo-Random Number Generators

- If you or someone else has to intervene to choose sample elements, randomness is affected.
- Use a psuedo-random number generator
 - ▶ Not truly random but behaves random (hence pseudo-random)
 - ▶ Implemented in many software including statistical software
 - ▶ Random number tables are not practical for large samples

Addressing Researcher Bias: Pseudo-Random Number Generators

Use MS Excel to select 20 items at random from a list of 1000.

First ensure that the Analysis Toolpack is loaded.

- Launch MS Excel
- Click File>>Options
- Select Add-ins
- In the manage box select Excel Add-ins>>Go
- In the Add-ins box check Analysis Toolpack then click OK

See: <https://support.microsoft.com/en-us/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4>

Examples of Sampling Approaches

What biases or sources of bias can be present in the following examples?

- Convenience Sample: Interview persons passing through the bus park for a sample to represent Guyanese.
- Convenience Sample: Select students passing a certain point on a catwalk on campus to represent university students.
- Snowball sample: Respondents recruit other respondents for a survey

Examples of Sampling Approaches

What biases or sources of bias can be present in the following examples?

- Data collected from the 2022 graduate students to represent graduate students
- Selected households decided which member responded to a household survey
- Volunteer sample
- Quota Sample: Moved on to the next household (or telephone number) if one refused.
- Three examples from audience

Address Bias Before Analysis

- Statistical analysis will not fix bias in the sample.
 - ▶ Address bias at the sampling and data collection stages
 - ▶ Data analysis is **garbage in garbage out!**
 - ▶ The technique applied does not indicate whether the results are interpretable

Appropriate Analysis for Variable Type

- Analysis must be appropriate for the variable type
 - ▶ Know whether parametric or non-parametric techniques are appropriate
 - ▶ If you have continuous dependent variables, parametric techniques might be applicable subject to satisfaction of model assumptions
 - ▶ If you have categorical variables, non-parametric techniques might be appropriate
 - ▶ Some rating scales might exist in a gray area

Use Latent Variable and Multivariate Analyses where Appropriate

- Analysis must be appropriate for the kind of constructs being estimated
 - ▶ Are your variables/constructs observed or latent?
 - ▶ Use latent variable models for latent variables.
 - ▶ Are the data multivariate or univariate?
 - ▶ Analysing multivariate data one variable at a time will not give appropriate results

Look out for Multilevel Structures in Data

- Analysis should account for structures in the data
 - ▶ There is more awareness of multilevel structures in data now
 - ▶ Eg. Data from students within several schools. Schools are one level and students on another.
 - ▶ Multilevel data require enough higher-level units for estimation.

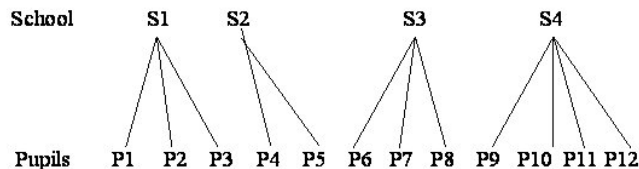


Image: Center for Multilevel Modelling, University of Bristol:

<http://www.bristol.ac.uk/cmm/learning/multilevel-models/data-structures.html>

Account for Sample Design During Analysis

- Analysis needs to be appropriate for the sampling applied (not often honoured)
 - ▶ This is not often done. Simple random sampling is most often assumed
 - ▶ Many software not have survey analysis modules that take the sample design into consideration.
 - ▶ Accounting for the sample design can lead to more precise estimates.

Interpret Results with Sampling in Mind

- Results interpretation should account for sampling
 - ▶ To make inferences, the population represented by the data must be known
 - ▶ Bias in the data affects representativeness
 - ▶ If bias is large, the results might not be interpretable.
 - ▶ If we know what bias exists and can estimate it, it might be possible to account for the bias in the analysis
 - ▶ E.g. unit or item nonresponse bias can be addressed by imputation or weighting

Interpret Results with Sampling in Mind

- Results interpretation should account for sampling
 - ▶ If the sample does not represent a particular population, do not project results onto that population. Use caution.
 - ▶ If non-probability sampling is used, you may have to refine the target population when interpreting the results
 - ▶ **Know and acknowledge the limitations of your data.**

Interpret Results with Analysis in Mind

- The analysis technique applied can imply some caveats for interpretation.
 - ▶ You must know your model assumptions
 - ▶ Check that model assumptions are satisfied.
 - ▶ Acknowledge model assumptions that are not satisfied. If possible find another appropriate model
 - ▶ You may need to transform or recode or reorganise your data to use the techniques that you can execute.
 - ▶ Understand the efficacy of your measurement instrument
 - ▶ **Know and acknowledge the limitations of your measurements and analysis.**

Do not Overreach

- Data analysis can be quite technical.
- Learn appropriate methods of analysis; otherwise using what you know can limit what you get out of your study.
- I see no problem with getting help with data analysis **but check with your school and follow its rules.**
- Do not overreach with your interpretation.
- Better to have results that are considered less than ideal than to have something not valid at all due to overreaching.

Questions