

Special Topic Lecture: Implementing simulation studies

Author: Nicholas G Reich, Jeff Goldsmith

*This material is part of the **statsTeachR** project*

Made available under the Creative Commons Attribution-ShareAlike 3.0 Unported License: http://creativecommons.org/licenses/by-sa/3.0/deed.en_US

What is a simulation study?

Why run a simulation study?

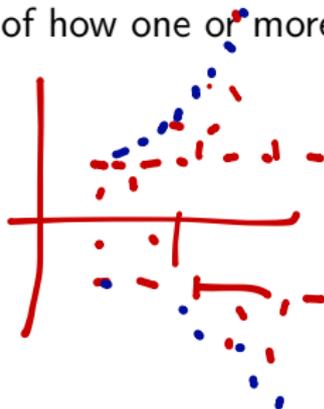
- to evaluate whether your statistical method works!
- to determine how much variability you might reasonably expect from your estimates
- to calculate power for a study

Especially useful when data model and/or statistical method are complex, and do not have tidy theoretical results.

What is a simulation study?

A statistician's laboratory

- tight control over the parameters of your data generating model
- systematic exploration of possible parameters
- careful evaluation of how one or more methods perform

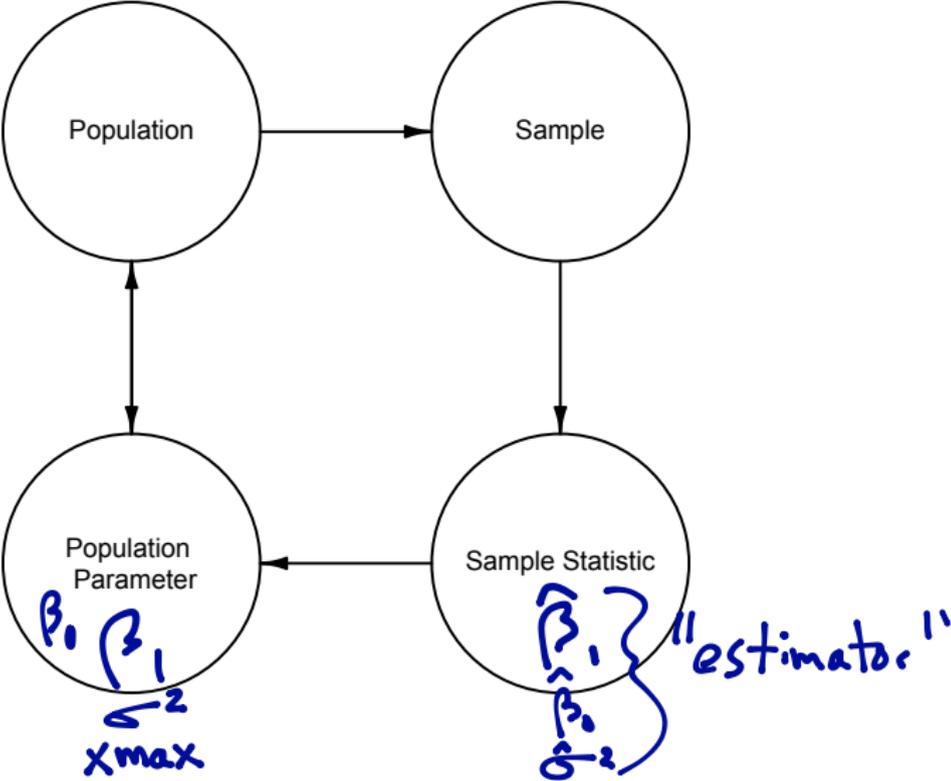


How to run a simulation study

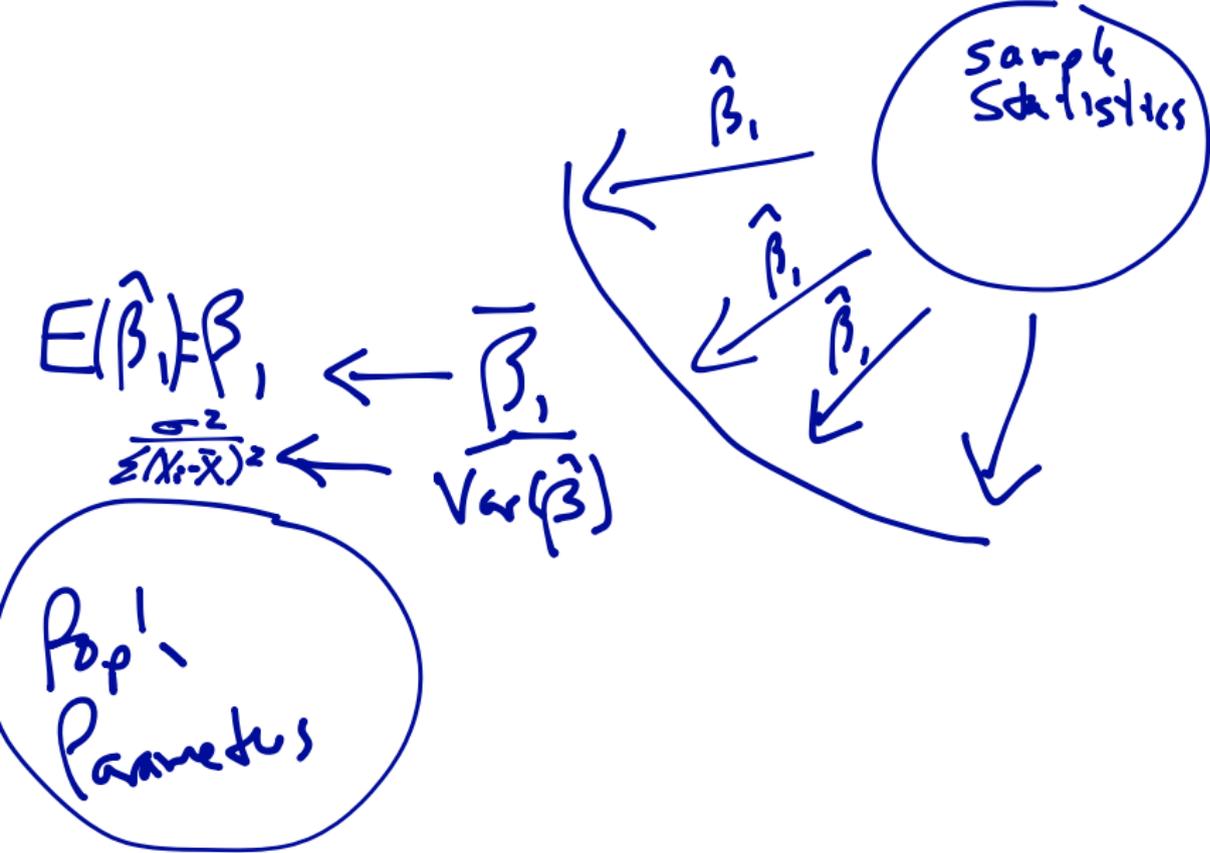
Key steps

- Identify a data generating model and its associated parameters
- Define the question and scope: which parameters do you want to investigate? what ranges?
- Write code to run the analysis that is easily replicated (maybe write a function?)
- For each distinct set of parameters, generate and analyze data, storing the results. (Note: try to minimize operations within your loops!)
- Summarize the results.

Circle of Life



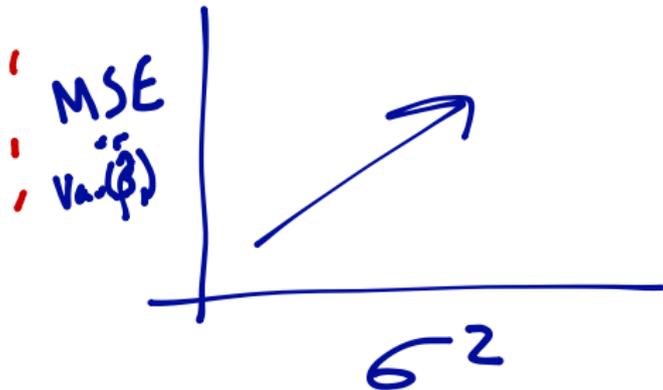
Circle of Life: a simulation study perspective

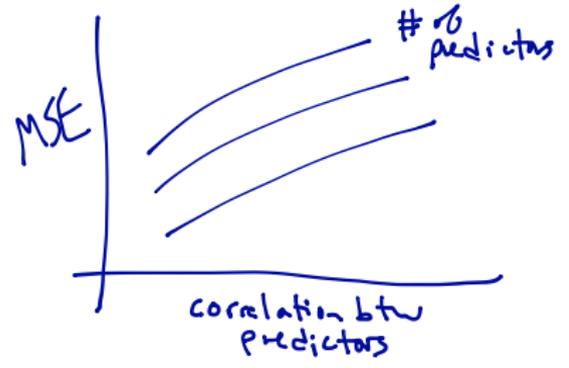
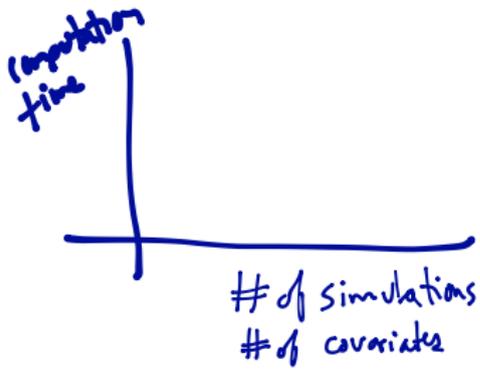
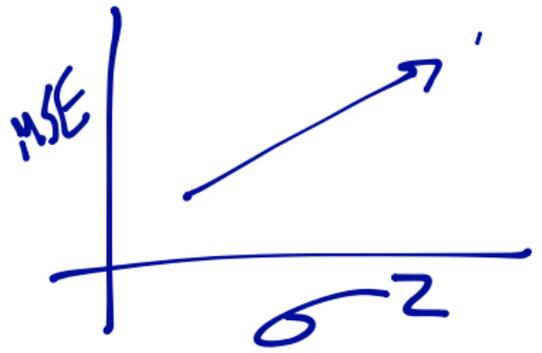
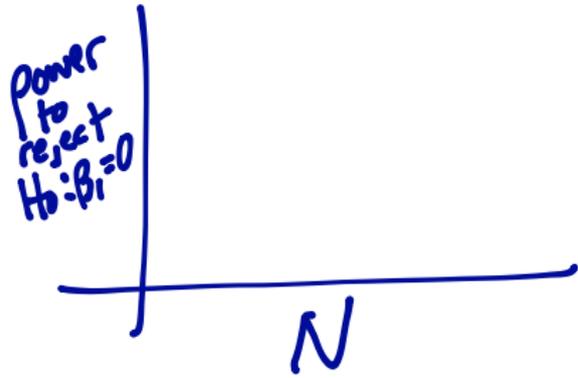


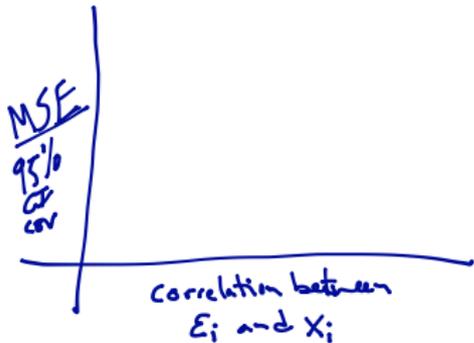
$E(\hat{\beta}_1)$ $\text{true } \sigma^2$ $\text{Var}(\hat{\beta}_1)$ $\bar{\beta}_1$ $\widehat{\text{Var}}(\hat{\beta}_1)$ MSE 95%

true
 value set
 in the
 laboratory

results
 from the
 simulation







- Spilled data
 - % age of obs to "spile"
 - mean of spilled obs

correlated errors

