

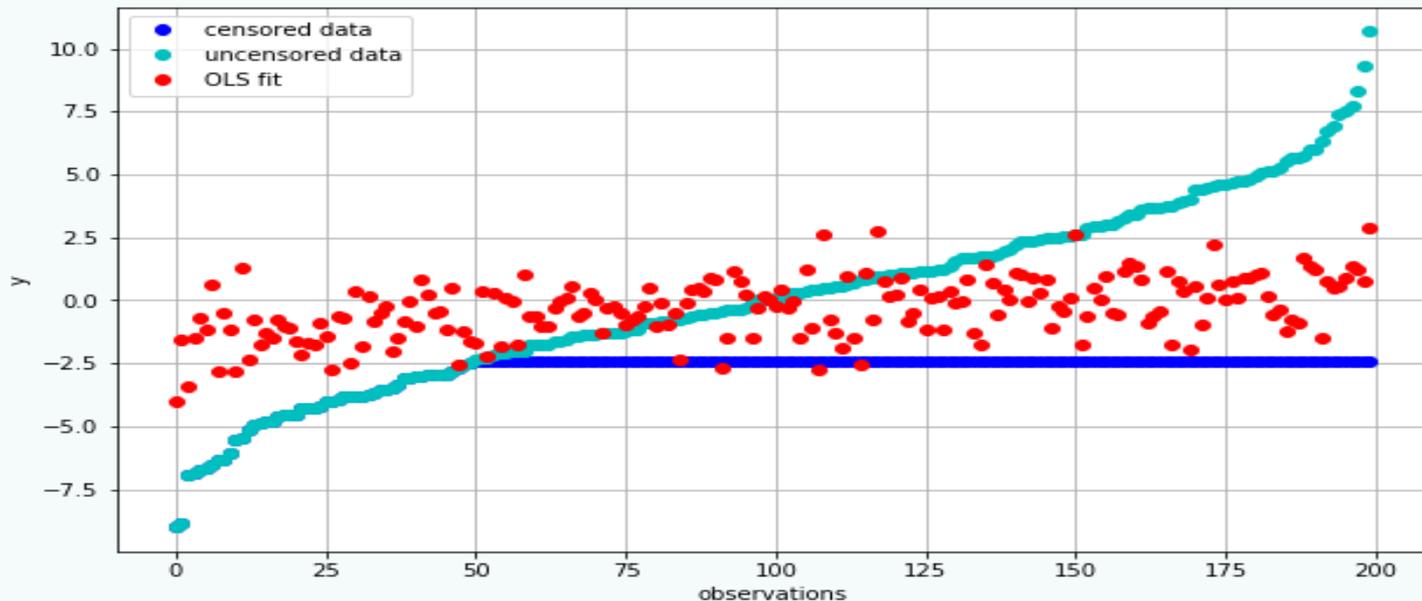
Lava is Hot! Assessing the Quality of Correlation Estimates Under Data Point Censoring

Fitsum A. Ayele & Kimberly A. Barchard

UNLV

Fall Research and Spring Proposal

- Fall censored data analysis project.
- *Lava* package in R.
- Here I will outline the spring project.



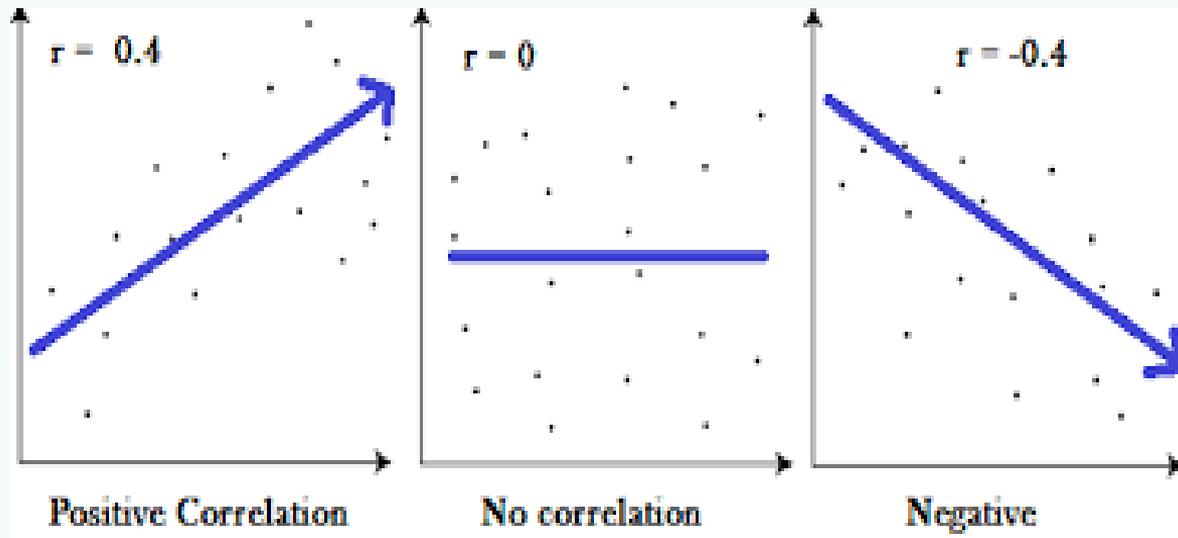
Introduction

- What is censoring?
- Left censoring: True value is less than or equal to recorded value (Barchard & Russell, 2020; Pesonen et al., 2015).
- Right censoring: The event of interest does not have enough time to occur (Gijbels, 2010).



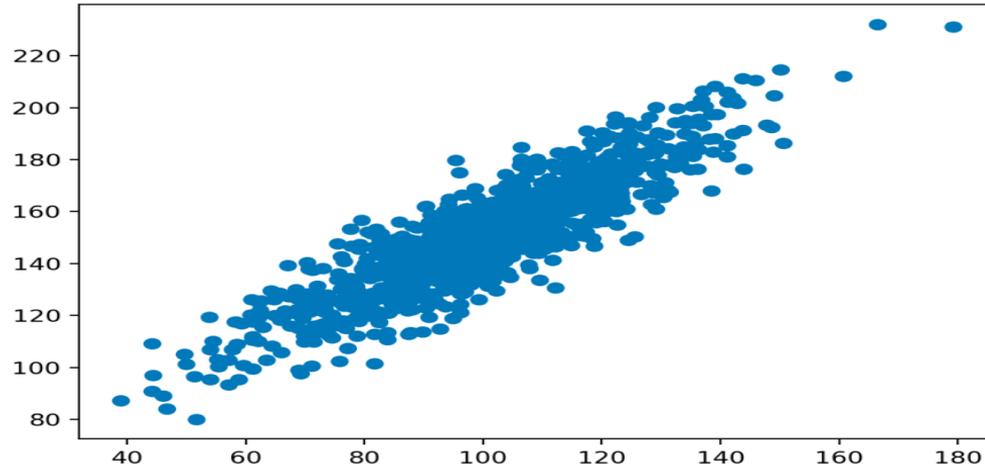
Introduction and Research Proposal

- Censoring can skew results and invalidate conclusions.
- Purpose of this study: *Lava* package in R (Holst et al. 2015).
- Two point estimates of ρ_{XY} : Correlation and Regression.
- Bias = $\hat{\rho}_{XY} - \rho_{XY}$.



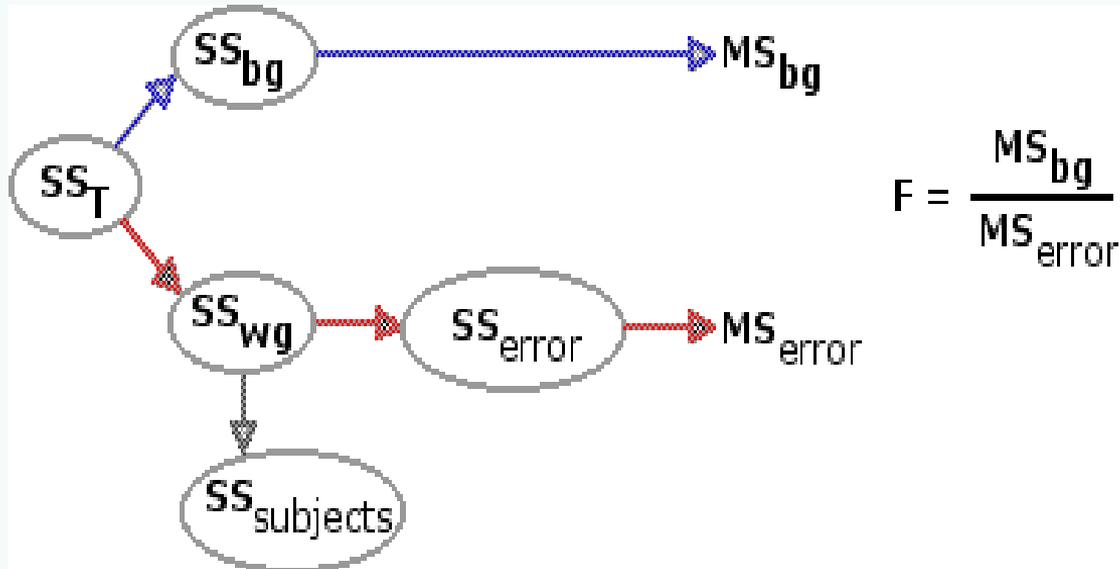
Proposed Research Question

- Research Question: Are estimates of ρ_{XY} more biased for the correlation model or regression model, when x and y are censored? And how does that bias vary for different values of ρ_{XY} , different sample sizes, and different patterns of censoring?
- In the fall, we specifically investigated the correlation model.



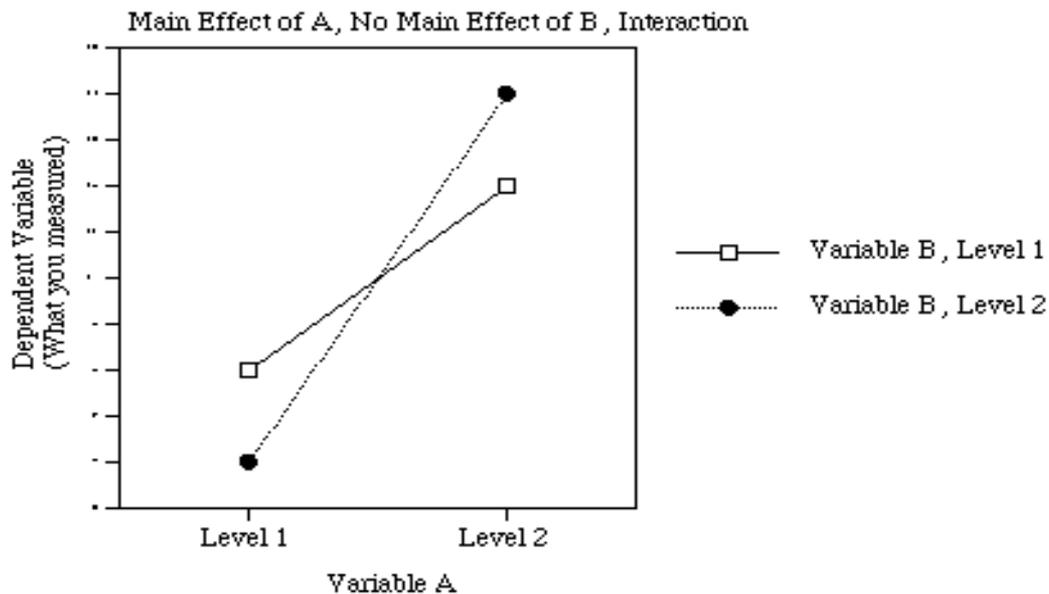
Proposed Study

- 100 cells. Each cell: unique combination of sample size, ρ_{XY} , and censoring pattern.
- 2 sample sizes; 10 values of ρ_{XY} ; 5 censoring patterns.
- Data analysis: Four-way Between-Within ANOVA.



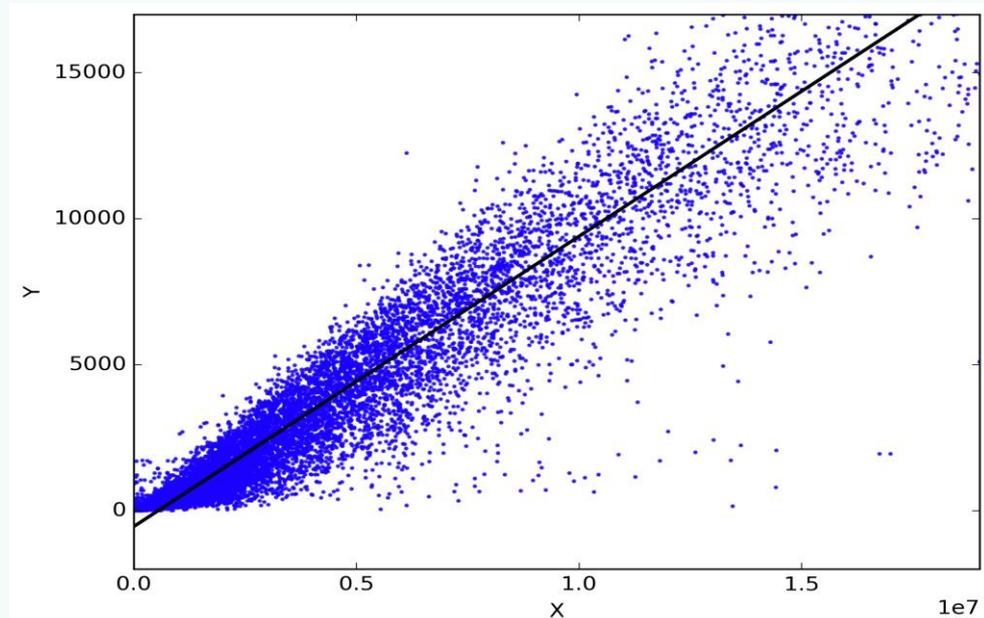
Results

- We expect a significant censoring pattern * ρ_{XY} * sample size interaction.
- No effect of model: Constraining the estimate/solution with regression will produce marginally better results.
- Theoretically may be better; Will not be better in terms of bias.



Discussion

- Regression model may converge on estimates more frequently.
- Therefore, we will recommend that researchers use the regression model.
- Implications.



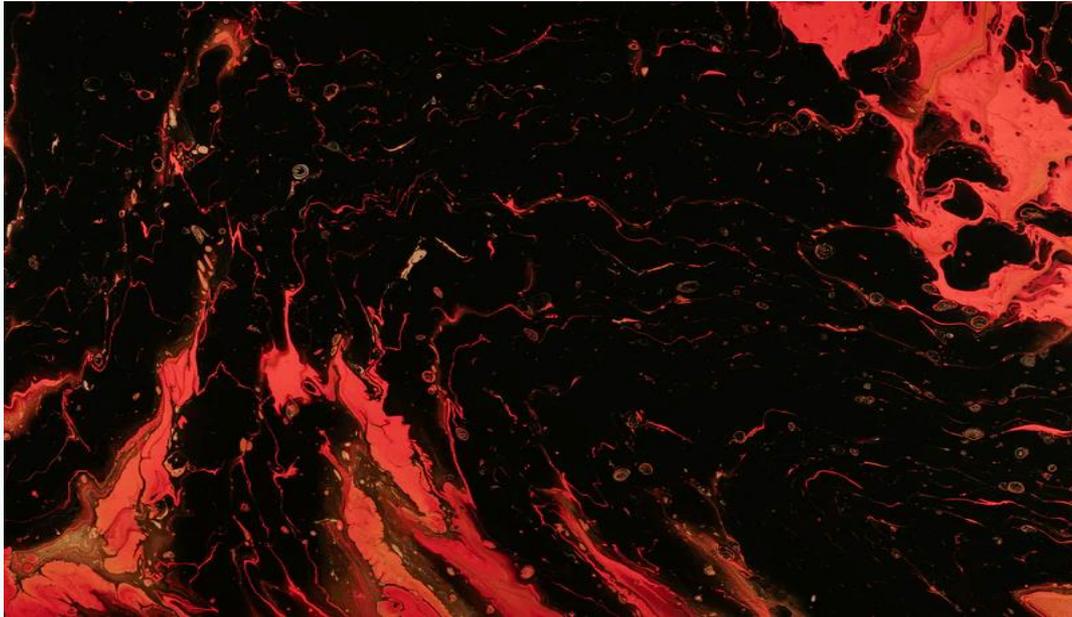
Future Research

- Investigate more complex models.
 - Encouraging results from Holst et al. (2015).
 - Binary outcome variable, a latent variable with four indicators, and a continuous covariate.
- Investigate *lava* estimates for non-normal distributions.



Conclusion

- Censored data are typically unrecognized in research.
- *Lava* provides a good way to conduct censored data analysis.
- We hope future researchers can use *lava* in their own investigations.



References

- Barchard, K. A., & Russell, J. A. (2020). *Modelling and Correcting the Effect of Data Point Censoring on Correlations* [Unpublished manuscript]. Department of Psychology, University of Nevada, Las Vegas.
- Gijbels, I. (2010). Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), 178-188. <https://doi.org/10.1002/wics.80>
- Holst, K. K., Budtz-Jørgensen, E., & Knudsen, G. M. (2015). *A latent variable model with mixed binary and continuous response variables*. Available at https://www.researchgate.net/publication/279864661_A_latent_variable_model_with_mixed_binary_and_continuous_response_variables
- Pesonen, M., Pesonen, H., & Nevailanen, J. (2015). Covariance matrix estimation for left-censored data. *Computational Statistics and Data Analysis*, 92, 13-25. <https://doi.org/10.1016/j.csda.2015.06.005>

Any Questions?

