UNLV

Censored Data: Evaluating Two Methods of Estimating Correlations within R package *Lava*

Cassandra K. Hoffman, Eden K. Thiess, Fitsum A. Ayele, and Kimberly A. Barchard
University of Nevada, Las Vegas
Contact Information: Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, kim.barchard@unlv.edu

Introduction

- Censored data occur when only partial information about the value of a variable is known. If researchers do not correct for censoring, their statistical results may be distorted and conclusions invalid.
- Left and right censoring occur when values fall below or above a given threshold, respectively. If both variables are censored in the same direction (both left or both right), bias is greatest for negative correlations (Barchard & Russell, 2020).
- The present study compares the correlation and regression models within Holst et al.'s R package *lava* to determine which produces the least biased estimates.

Method

- Monte Carlo simulations were used to examine the effects of various patterns of left censoring for different combinations of correlation values and sample sizes. We calculated the mean bias (difference between *lava's* estimates and the true values) and the proportion of trials for which *lava* failed to converge on an estimate.
- Additionally, we used ANOVA and graphed significant interactions.

Results

- *Lava* always converged on an estimate, regardless of model, and bias was minimal unless both variables had 80% censoring and the correlation was large and negative.
- Bias decreased when sample size increased from 500 to 1000.
- Differences between models were trivial; however, the regression model is preferred as it always produces estimates between -1 and 1.

Discussion

- Although our study used left-censored data for both variables, we expect research findings to be similar to ours when using right-censored data for both variables. Future research could examine this supposition.
- Our data were normally distributed and did not contain outliers. Future research could assess *lava's* accuracy for non-normally distributed data and the presence of outliers.
- We did not assess the effects of a small sample. Future research should assess *lava's* performance when working with smaller sample sizes.
- *Lava* appears to be a useful method of reducing bias for censored data.

Lava produces **unbiased correlation estimates** for **censored data** when samples are large and censoring is low to moderate.

Figure 1

Effect of Population Correlation, Censoring Pattern, and Sample Size on Bias (Sample Size set to 500)

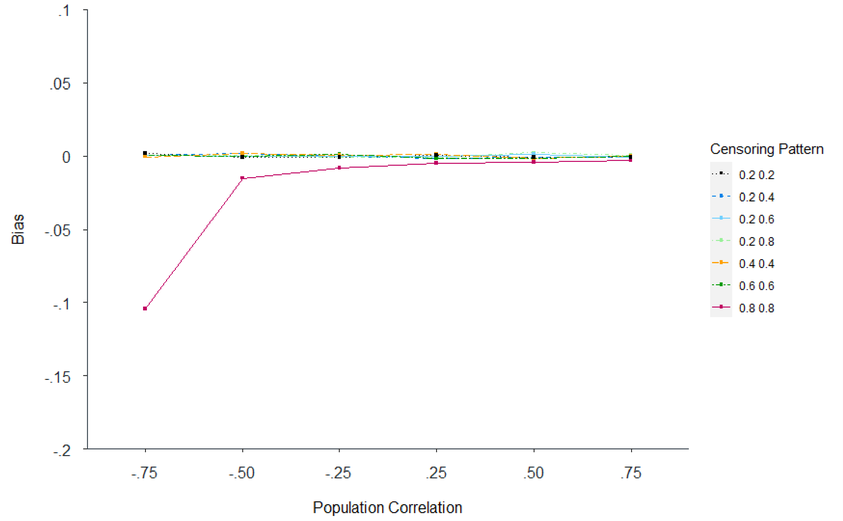
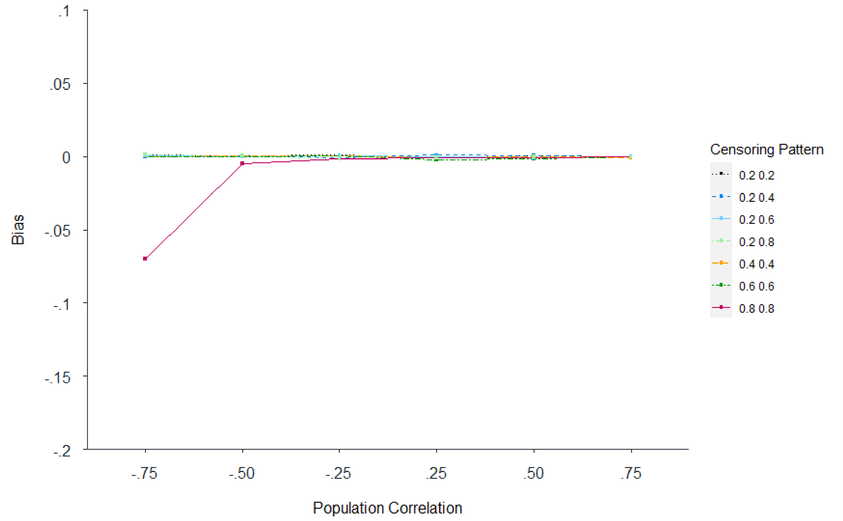


Figure 2

Effect of Population Correlation, Censoring Pattern, and Sample Size on Bias (Sample Size set to 1000)