INTRODUCTION

DATA CENSORING

- Data censoring is highly pervasive in research within various fields of study.
- Censoring is a condition in which the value of a measurement is partially observed.
- Distorted analysis may occur if censored data is not addressed.
- Right censoring is a common occurrence in survival and longitudinal studies where the time-to-event is being recorded but does not occur within the specified duration of the study for certain participants (Gijbels, 2010).
- Left censoring occurs when the event of interest occurs prior to the commencement of the study, or because the limit of detection which the measurement device lacks the sensitivity to capture all data points (Gijbels, 2010).

PSILOCYBIN STUDIES

- **Left Censoring**
  - In a study by Carhart-Harris et al. (2017), researchers measured serotonin levels (indicates decreased depressive symptoms) in 19 participants after they were given a dosage of psilocybin. Serotonin was not recorded in three of the participants which could be due to lower levels in which the fMRI could not detect. This would lead to left censoring.
- **Right Censoring**
  - In a study by Roseman et al. (2018), 20 volunteers with treatment resistant depression were administered 2 dosages of psilocybin (10mg and 35mg) one week apart. It is uncertain if the volunteers took more than the amount given by the researchers during the 5-week study timeframe. If additional psilocybin was taken by any of the participants, outside of what was given by the researcher, right censoring would have occurred.

RESEARCH QUESTION

- We will evaluate R package *lava's* performance on estimating the correlation between uncensored variables given data from censored variables.
- *Lava* can be used when there is censoring on more than one variable

METHOD

- To evaluate the accuracy of R package *lava* in estimating the correlation between uncensored variables X and Y based on data from censored variables x and y, we varied the patterns of censoring for x and y along with ρ_{XY} ; sample size fixed at 500.
- We ran 30 cells where each cell represents a unique combination of ρ_{XY}, censoring pattern, and sample size.
- For each cell, we ran 1000 trials and generated a random set of data for which X and Y had a normal distribution.
- Patterns of censoring on x and y included: 10% censoring on both, 50% censoring on both, and 95% censoring on both, and of 20% censoring on x and 80% censoring on y.
- ρ_{XY} values of -.95, .95, -.5, .5, -.05, and .25 were used.
- We provided the x and y data to *lava* and asked it to estimate the correlation between X and Y (ρ_{XY}).
- To assess *lava*'s performance, we calculated bias – mean difference between actual values of ρ_{XY} and *lava* estimates of ρ_{XY}.

Lava is all you need:

R package reduces bias for correlations among censored variables

Jerlyn Malasig*, Monica Cordova-Medina*, LaShawn Tith*, Fitsum A. Ayele, and Kimberly A. Barchard

University of Nevada, Las Vegas

*These authors contributed equally to this poster.

RESULTS

- See Table 1. The table shows the bias for *lava* estimates of ρ_{XY} for each unique combination of censoring pattern, ρ_{XY}, and sample size.
- The *lava* estimates of ρ_{XY} for all cells with censoring patterns of 10% on x and y, 50% on x and y, and 20% on x and 80% on y were unbiased.
- The *lava* estimates of ρ_{XY} were biased when the censoring pattern was 95% on x and y; bias was largest when the initial correlation was negative.

DISCUSSION

- The results from our Monte Carlo study showed that *lava* estimates of ρ_{XY} for all unique combinations of sample size, ρ_{XY}, and censoring patterns of up to 50% were unbiased. Mixed censoring on X and Y (20% and 80%) produced similar unbiased results.
- When the censoring pattern was high (95% on both x and y) and the correlation was negative, there was substantial bias.
- If there was same direction censoring on both variables (either both left or both right), *lava* produces severe biased estimates for certain negative correlations (low and moderate).
- If we had high left censoring on one variable and high right censoring on the other, we would have found that *lava* produced bias estimates for positive correlations.

IMPLICATIONS

- We are confident of R package *lava's* performance when degrees of censoring are low to moderate on at least one of the variables.
- We recommend R package *lava* to other researchers in their studies with low to moderate censoring and encourage them to minimize censoring to avoid biased estimates.

LIMITATIONS AND FUTURE RECOMMENDATIONS

- The R-package *lava* assumes a normal distribution for the values of X and Y. If data sets reflect a distribution that is not normal (i.e. skewed), *lava* estimates could differ and potentially be even more bias. Therefore, it is recommended that future researchers investigate how different distribution patterns affect censored data analysis.
- This study looked into the effect *lava* holds on estimating for correlation; future studies should assess regression models that will allow for estimations of future values and the regression of current data points, even when censoring occurs.

Table 1

Bias for lava Estimates of ρ_{XY} for Different Values of ρ_{XY} Under Different Patterns of Censoring

ρ_{XY}	Patterns of Censoring			
	10% x 10% y	50% x 50% y	20% x 80% y	95% x 95% y
-.95	.00	.00	.00	.06
-.50	.00	.00	.00	-.35
-.05	.00	.00	.00	-.24
.25	.00	.00	.00	-.04
.50	.00	.00	.00	-.02
.95	.00	.00	.00	.00

Note. Sample size was fixed at 500.