To view annotations

The Effect of Data Point Censoring on Correlations Between Skewed Variables

Kimberly A. Barchard
University of Nevada, Las Vegas
kim.barchard@unlv.edu
James A. Russell
Boston College

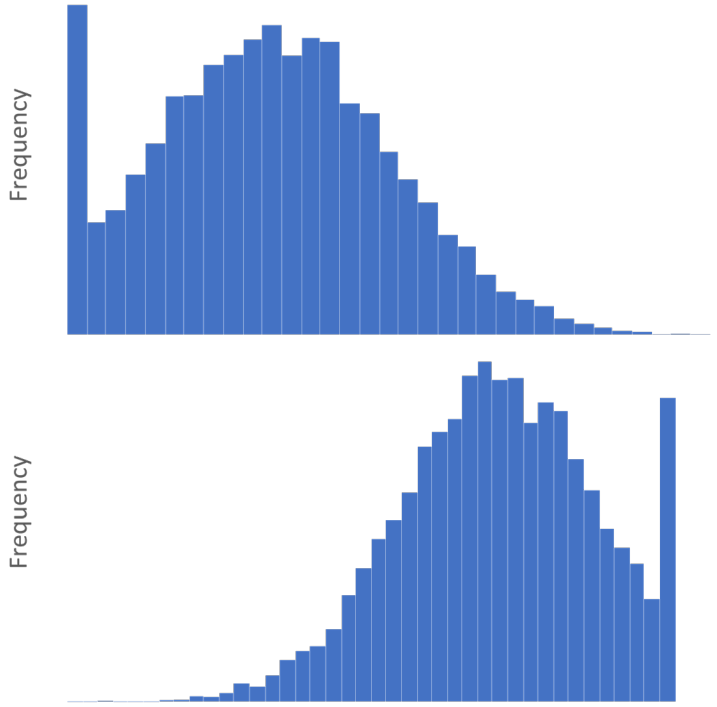
February 19, 2022

#SPSP2022

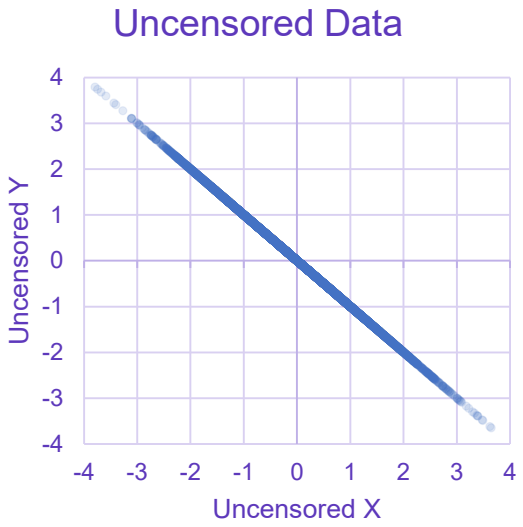
Data Point Censoring

Data point censoring occurs when a researcher knows a value is at least as large as (or no larger than) a certain value

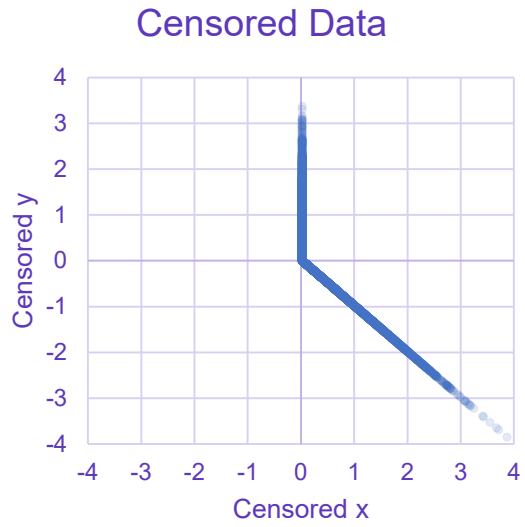
- Left censoring
  - Value is too small to be detected
  - Floor effects
- Right censoring
  - Length of time to event or frequency of event is underestimated
  - Ceiling effects



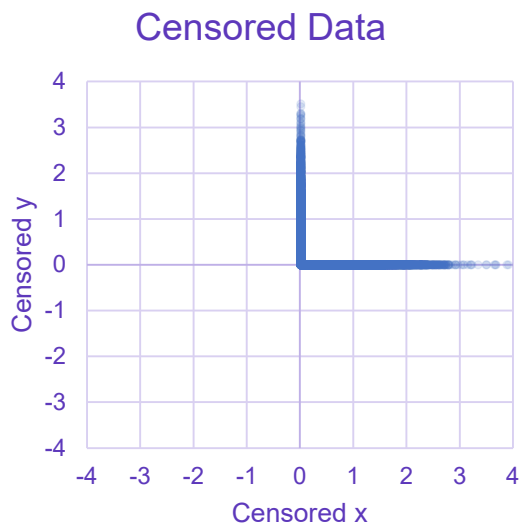
Censoring Distorts Correlations



X and Y have bivariate normal distribution with $\rho_{XY} = -1$



Half the x values are censored.
If X < 0, x = 0.



Half the y values are also censored.
$\rho_{xy} = -.467$

(Russell & Carroll, 1999)

We Examined the Effect of Left Censoring on Correlations for Skewed Data

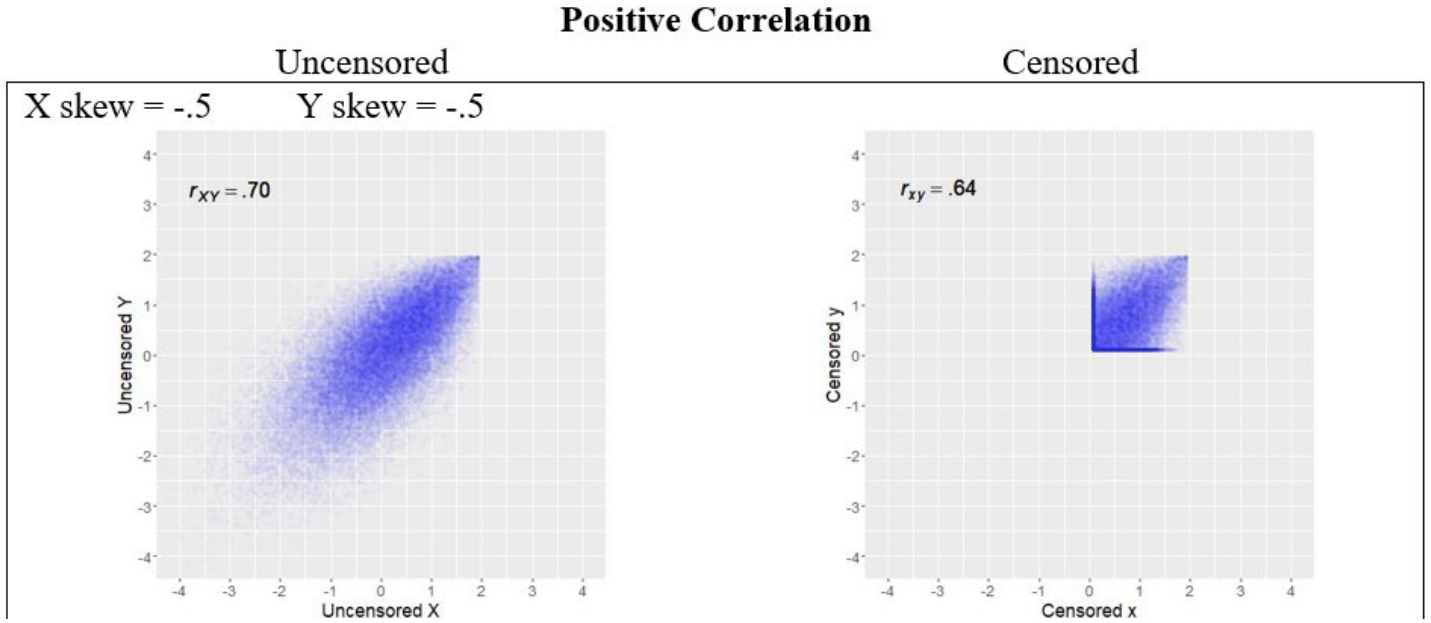
24 datasets with 50,000 cases

- Correlations: .7 and -.7
- Skew: .5 and -.5
- Left censoring: 30%, 50%, 70%

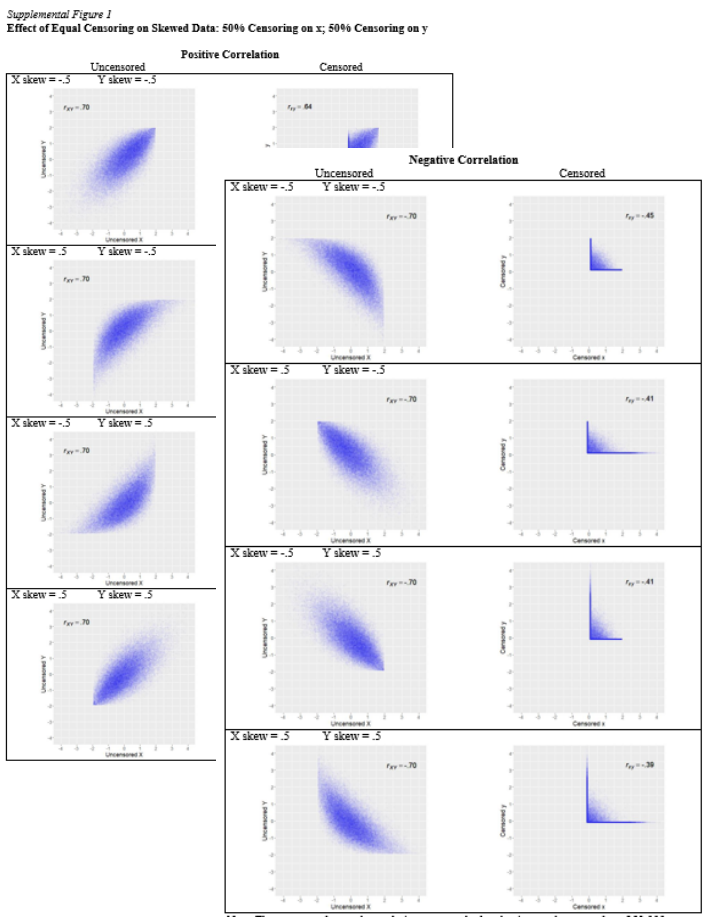
One of our datasets:

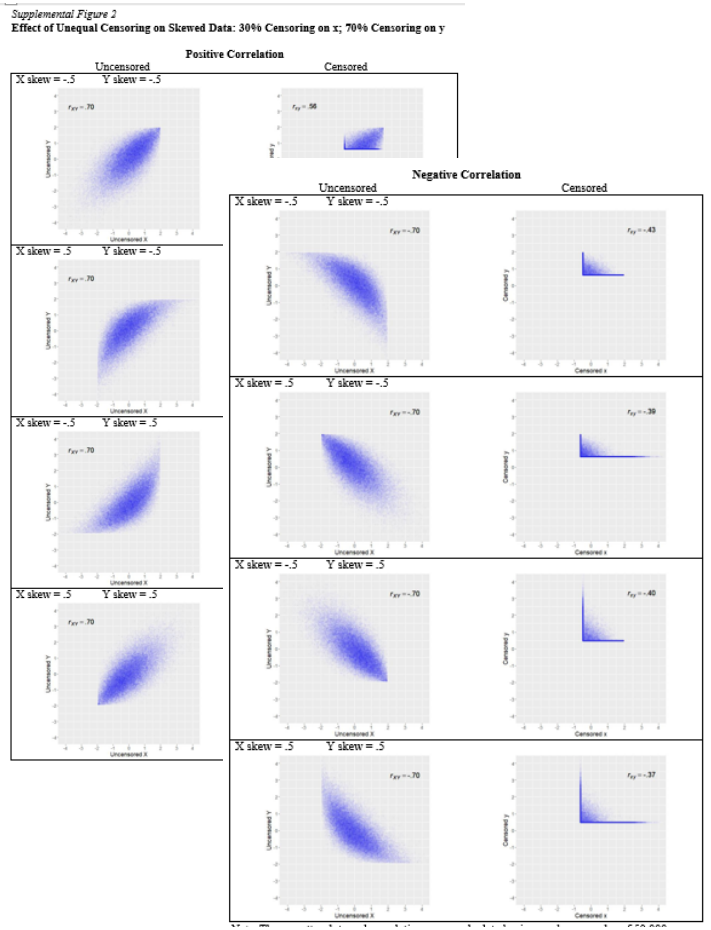
50% censoring on x
50% censoring on y

$\rho_{XY} = .70$
$\rho_{xy} = .64$



Our 24 Datasets





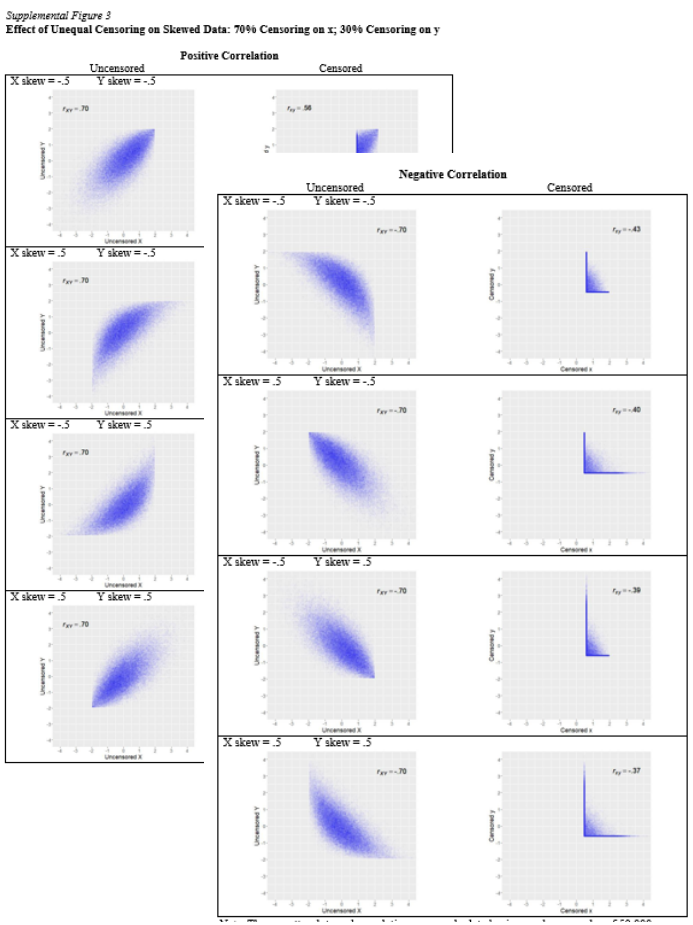
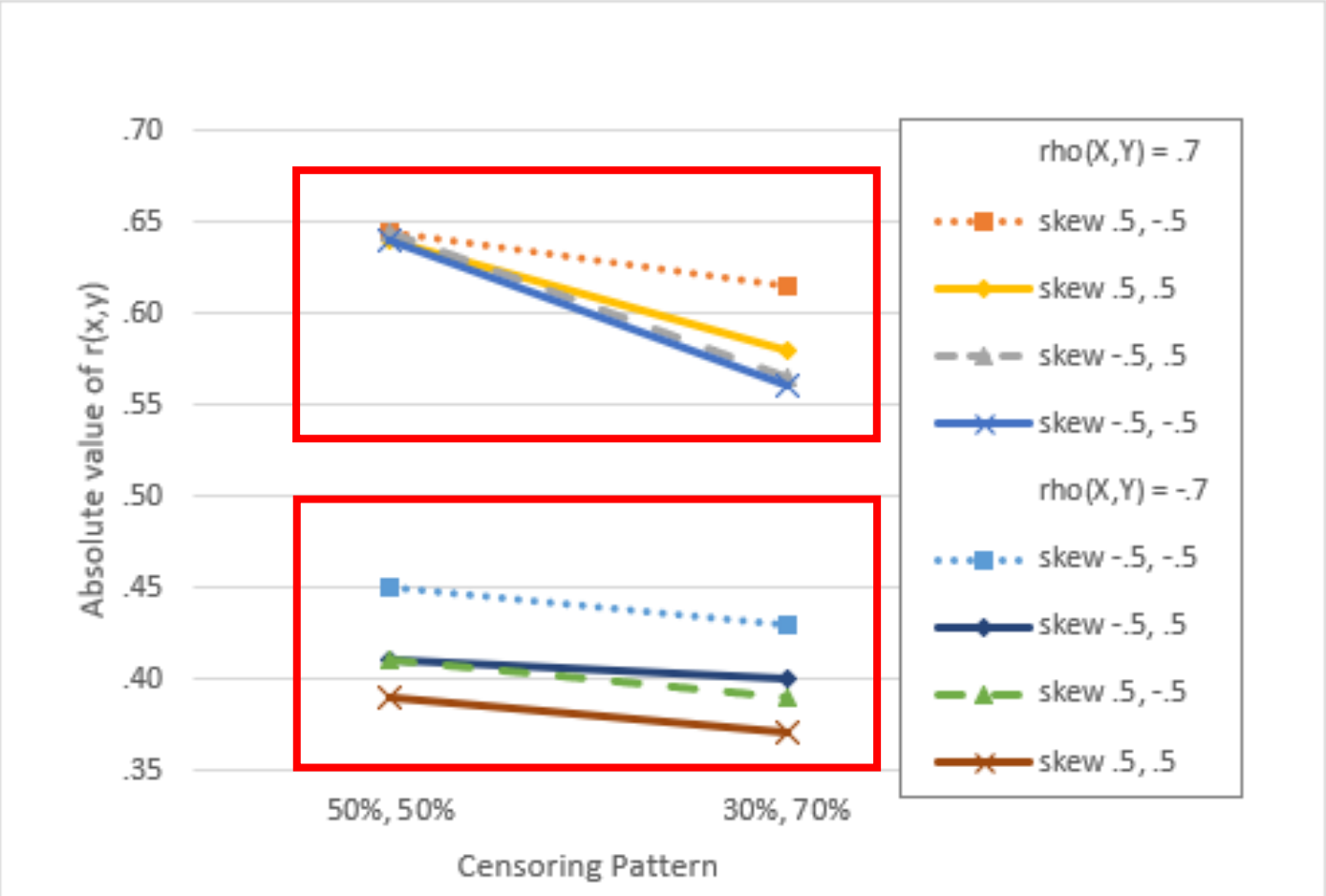


Figure 1

Three-way Interaction of rho(X,Y), Censoring Pattern, and Skew on r(x,y)

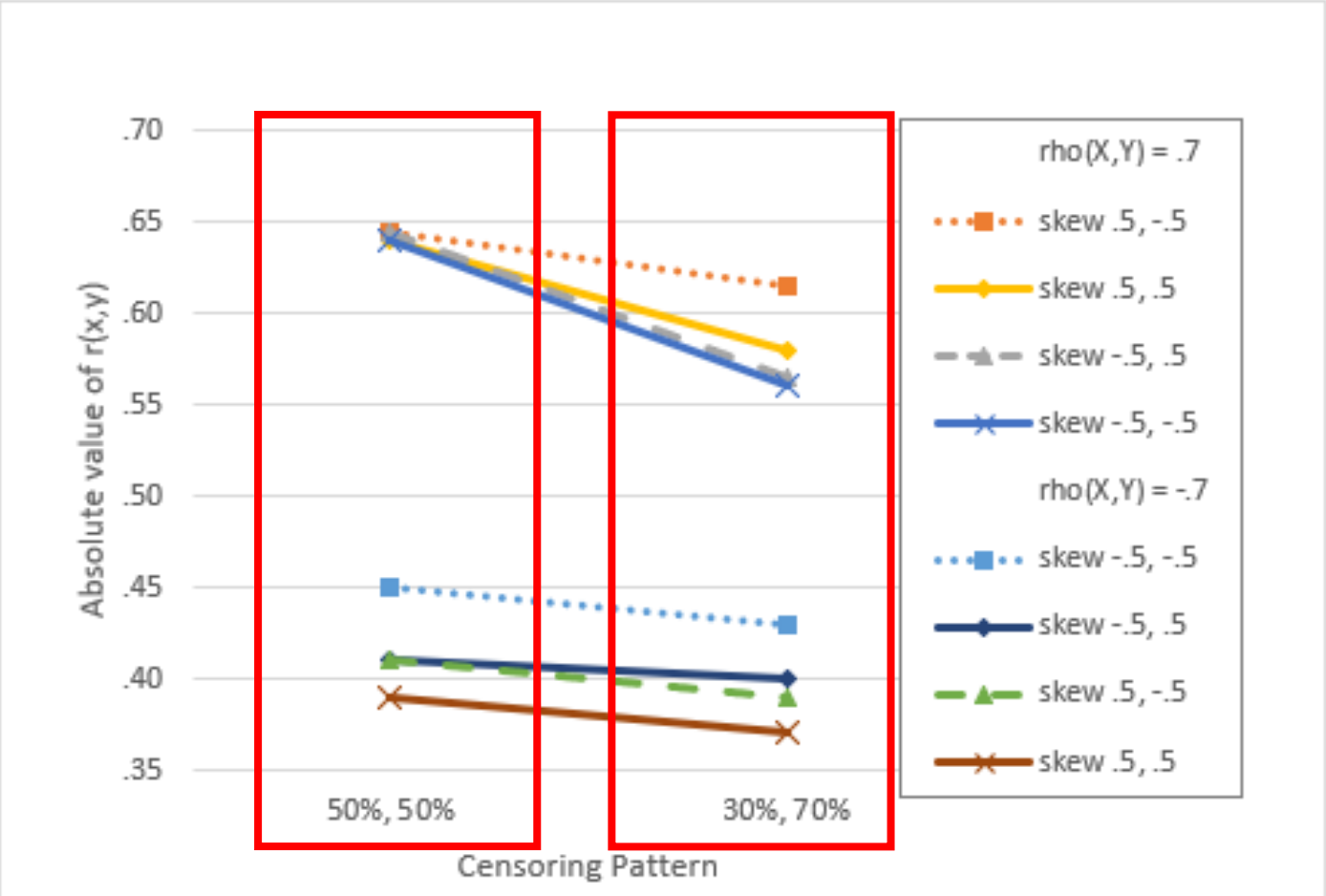


Main Effect for Correlation

Left censoring affects negative correlations more than positive correlations.

Figure 1

Three-way Interaction of rho(X,Y), Censoring Pattern, and Skew on r(x,y)

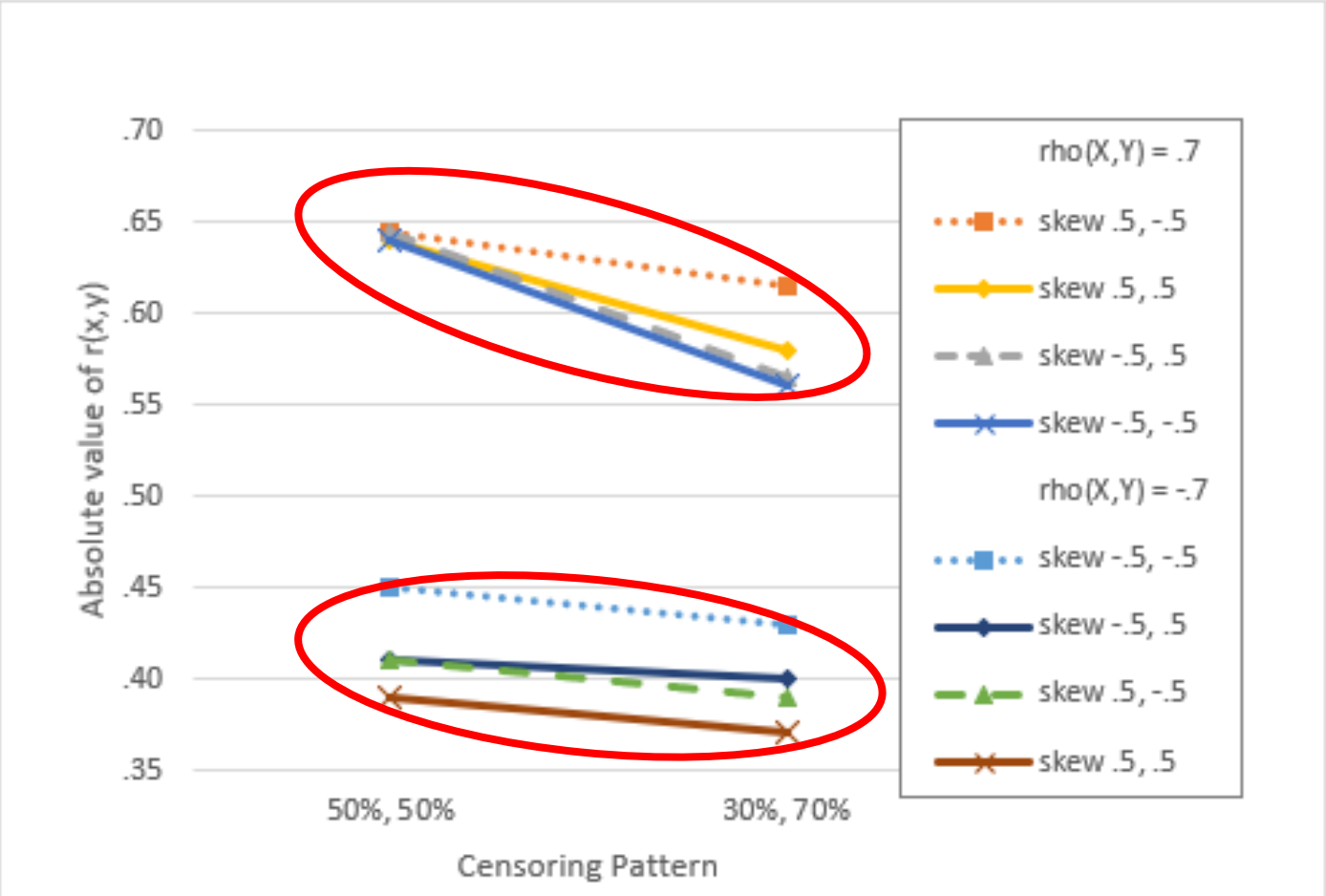


Main Effect for Censoring Pattern

Unequal censoring has a greater effect than equal censoring.

Figure 1

Three-way Interaction of rho(X,Y), Censoring Pattern, and Skew on r(x,y)

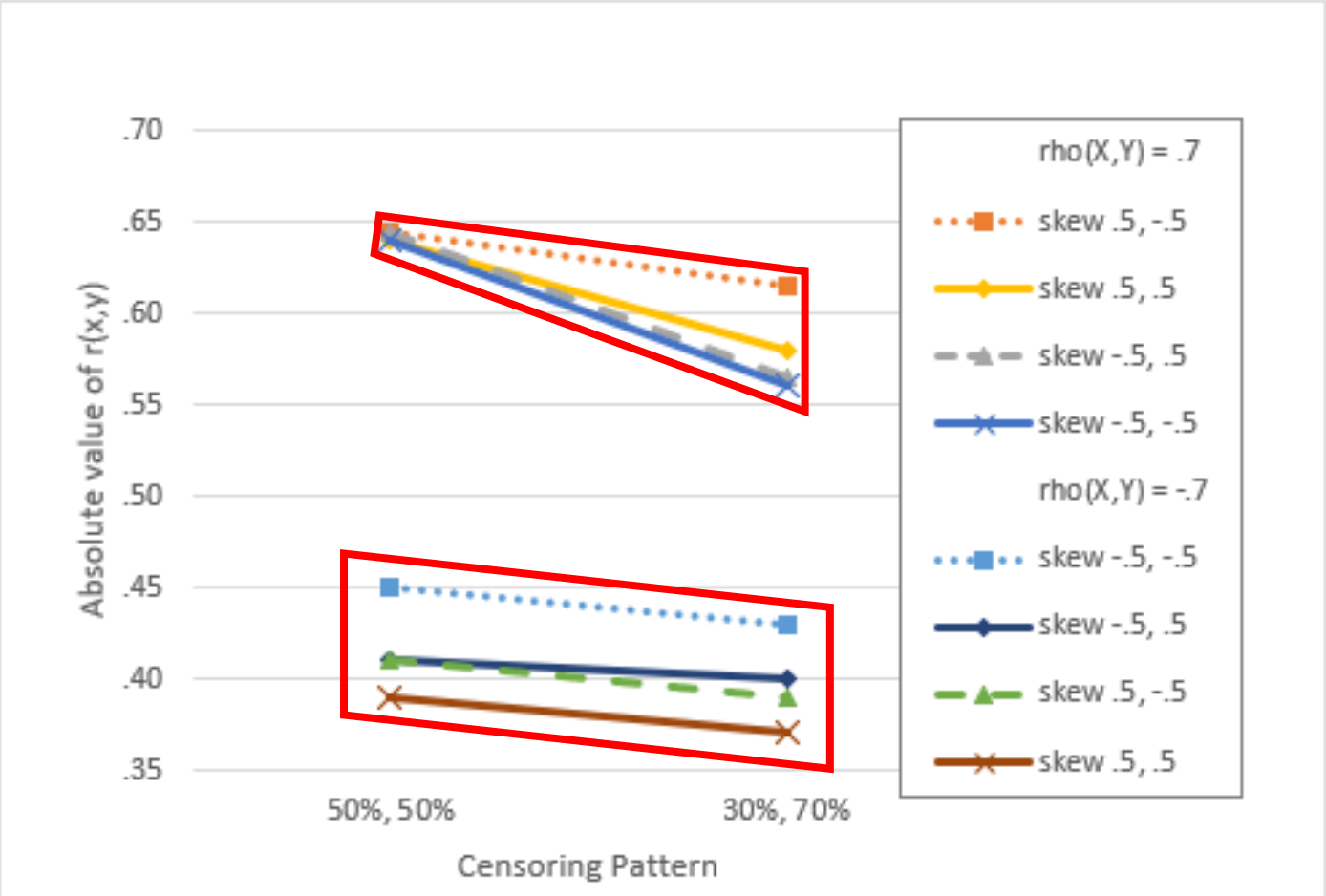


Interaction of Censoring Pattern and Correlation

Unequal left censoring has bigger effect on positive correlations than negative ones.

Figure 1

Three-way Interaction of rho(X,Y), Censoring Pattern, and Skew on r(x,y)



Three Way Interaction

For left censoring, censoring pattern and skew interact more for positive correlations than negative ones.

Discussion

- Right censoring is the mirror image of left censoring. Therefore,
  - If both variables have right censoring
    - Unequal censoring has a greater effect than equal censoring.
    - Censoring affects negative correlations more than positive ones.
    - Censoring pattern and skew interact more for positive correlations than negative.
  - If one variable has left censoring and one has right censoring
    - Unequal censoring has a greater effect than equal censoring.
    - Censoring affects positive correlations more than negative ones.
    - Censoring pattern and skew interact more for negative correlations than positive.
- A variety of methods (e.g., Mplus, R package lava) use data from censored variables to estimate correlations for uncensored variables.
  - Future research: How well do they work with skewed data?