Instructions

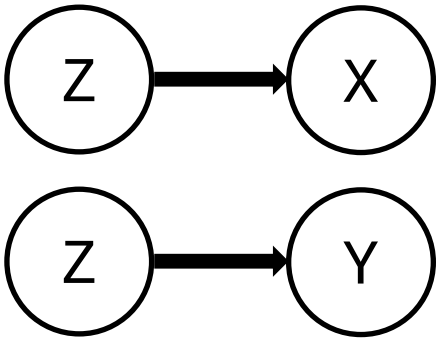
Complete the following simulation and upload your results to your course Github repository. You may submit your results as an R file, an R markdown/notebook file, or as a pdf. Please include all relevant R code and outputs from running the code in your answers. Make sure to comment your code.

Simulation

1. Use the rnorm() function to create two random variables in R with 20 observations each. Then, calculate the correlation between the two variables. Repeat this process many times. Plot the distribution of the correlation coefficients and report the standard deviation. On average, what would we expect the correlation between the two variables to be? What does this distribution tell us about sample estimates of population parameters?

2. Repeat the previous step with a sample size of 1,000 and provide a substantive interpretation of how the results differ.

3. Create three random variables in R that have the following causal relationship:



That is, Z causes both X and Y, but X and Y have no causal relationship. Plot X and Y on a scatter plot and report their correlation. What does this tell us about interpreting correlations?

Hint: Start by generating Z as a random variable, then create X and Y as some function of Z plus random noise.