University of Koblenz - AG Softlang

Data Science

Assignment 7: Causal Inference

For the submission deadline see OLAT

December 15, 2021

Question 7.1: Simulating Causal Influence

The lecture introduces a series of simulations, that produce synthetic data that comes close to real scenarios examining causation.

Select a question from the reference solution to assignment 1 [1] and write a simulation that produces synthetic data that corresponds to this question (data that you make up). Please just select causal questions (from the reference solution) that at least involve three variables (e.g., the variables *time to travel to work*, *job satisfaction* and *salary*). Write a simulation that produces (reasonable) data for the three variables. Define some variables as functions of other variables simulating a causal relationship (e.g., *job satisfaction* as a function of *time to travel to work* and *salary*; and *time to travel to work* as a function of *salary*). Provide the simulation code for producing the data and example data as CSV file.

You are also allowed to just implement a simulation for the example on *time to travel to work*, *job satisfaction* and *salary* that we have listed here as a reference to help you.

Question 7.2: Examining Causal Influence

Implement a multiple regression model in ULAM or STAN that examines your simulated data. In particular, try to recover how the variables relate to each other (e.g., how *job satisfaction* relates to the *time to travel to work*, and *salary*). Run the model on the simulated data, and report on the posterior for the relevant parameters of you model.

If you have done everything right, the model parameters will come close to the parameters that you have used to simulate the data. Provide us with the model code and plots of the posterior with respect to the parameters.

[1] https://videoakademie.ko-ld.de/Panopto/Pages/Viewer.aspx?id=cd6c4dd8-0c3a-480f-89bb-ade600adc75e