Assignment 3: Causal Inference

Deadline for hand-in: 26/03-2019
Github: https://github.com/saraoe/assignment_3_causal_inference

First part

Q1.1) Test if schizophrenia involve higher altercentric intrusion
To test if people with schizophrenia have higher altercentric intrusion, we have made a model that predicts the score for altercentric intrusion from the diagnosis (control or schizophrenia)

Altercentric Intrusion ~ 0 + Diagnosis

We define the same priors for both estimates - for schizophrenics and controls - which is a normal distribution, as altercentric intrusion is a continuous variable, with a mean of 4 and a SD of 1. By having the same prior, our prior belief is that there is no difference between the two groups. The prior for sigma we defined as a normal distribution with a mean of 1 and a SD of 2.

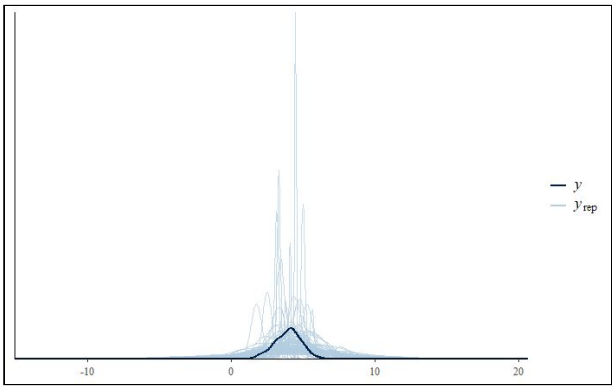


Figure 1: Prior predictive check

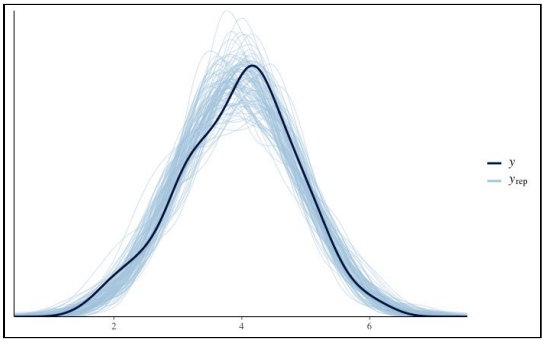


Figure 2: Posterior predictive check

The model indicates a credible difference in altercentric intrusion in the two groups supporting our hypothesis (b = 0.36, CIs = 0.16, 0.57, ER = 1332). Controls showed on average an altercentric intrusion effect of 3.86 (CIs 3.74, 3.98), and schizophrenia of 4.22 (CIs = 4.01, 4.43). The difference between controls and schizophrenics can also be seen visually in figure 3 and 4.

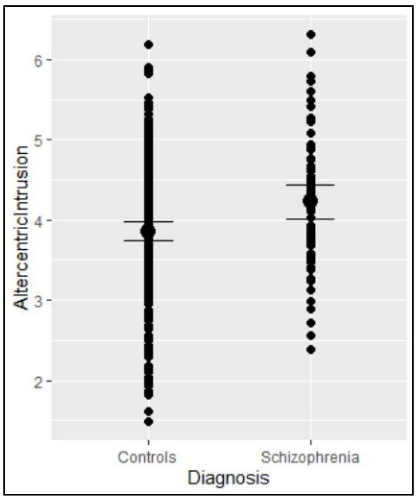


Figure 3: Plot of actual data

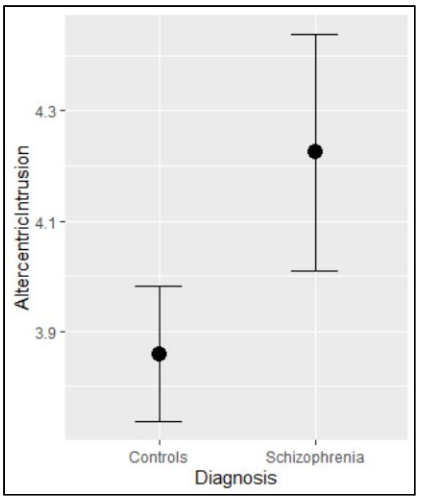


Figure 4: Difference between controls and Schizophrenics

Q1.2) Is altercentric intrusion related to specific symptoms in the patients?

To test if altercentric intrusion is related to Voice Hearing, Mind reading and Apathy we made the following models:

$$Altercentric\ Intrusion \sim 1 + Voice\ Hearing$$
$$Altercentric\ Intrusion \sim 1 + Mind\ Reading$$
$$Altercentric\ Intrusion \sim 1 + Apathy$$

We have mean centered all predictor variables for better interpretation of the intercept of the models. For all models we have a prior posterior which is normally distributed with a mean of 4 and a SD of 1 and a sigma, which is also normally distributed, with a mean of 1 and a SD of 2.
For the voice hearing and mind reading models, we have made a normally distributed prior for the beta, which has a mean of 0.5 (as we expect a positive slope) and a SD of 1. For the apathy model, we have also made a normally distributed prior for the beta, with a mean of - 0.5 (as we expect a negative slope) and a SD of 1.

The model for voice hearing showed a small positive effect of voice hearing on altercentric intrusion (b = 0.08, CIs = -0.19, 0.35). The credibility interval crosses zero. Therefore, there is a probability that voice hearing actually has no effect on altercentric intrusion.

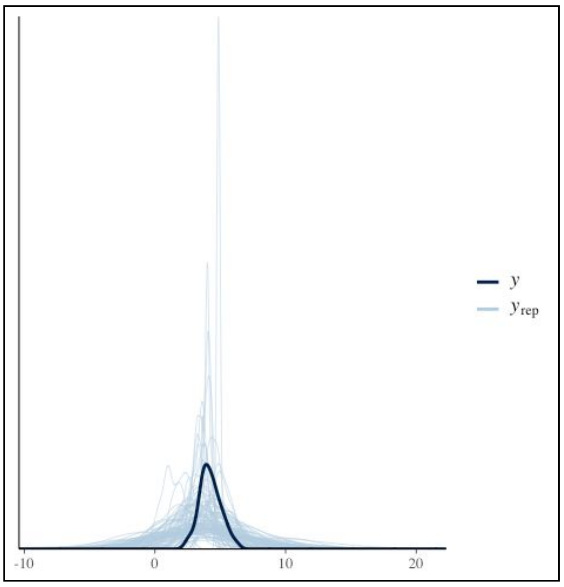


Figure 5: Prior predictive check for model with Voice Reading

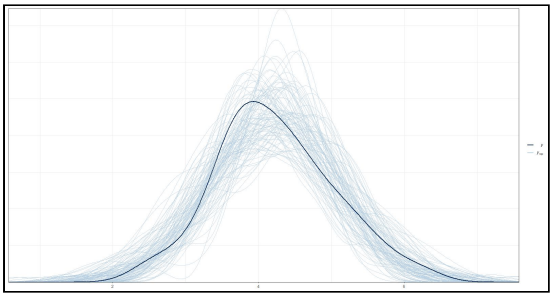
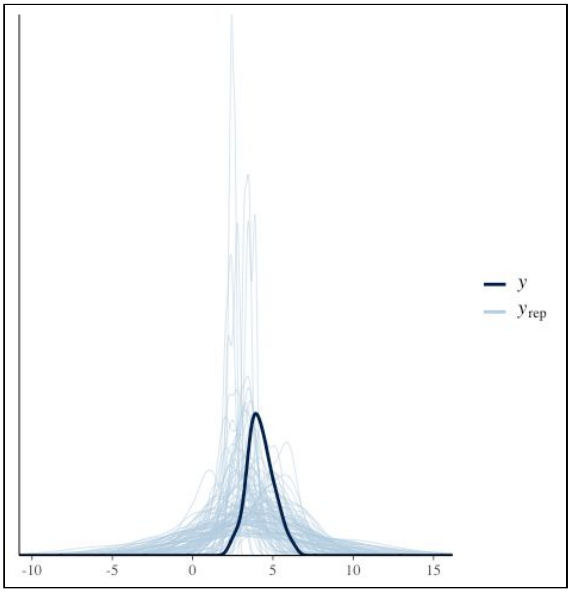


Figure 6: Posterior predictive check for model with Voice Reading

The model for mind reading showed a small positive effect of mind reading on altercentric intrusion (b = 0.09, CIs = -0.14, 0.30). The credibility interval crosses zero, there is a probability that mind reading actually has no effect on altercentric intrusion.



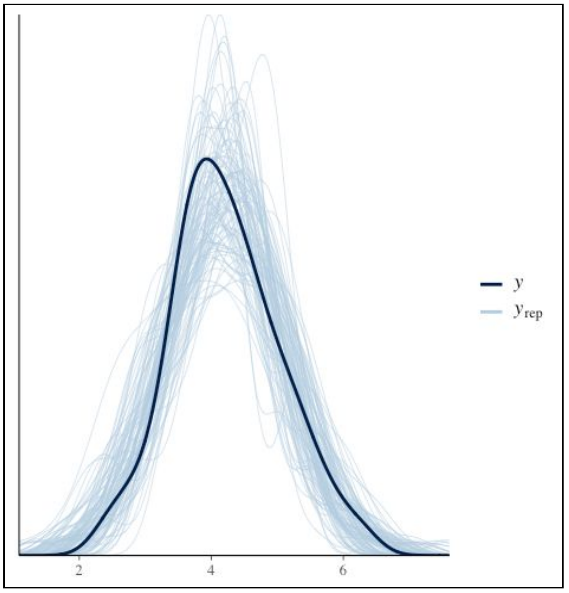


Figure 7 (left): Prior predictive check for the model with Mind Reading
Figure 8 (right): Posterior predictive check for the model with Mind Reading

The model for apathy showed a small negative effect of apathy on Altercentric Intrusion (b = - 0.23, CI-95% = -0.49-0.03). The credibility interval crosses zero, there is a probability that apathy actually has no effect on altercentric intrusion.

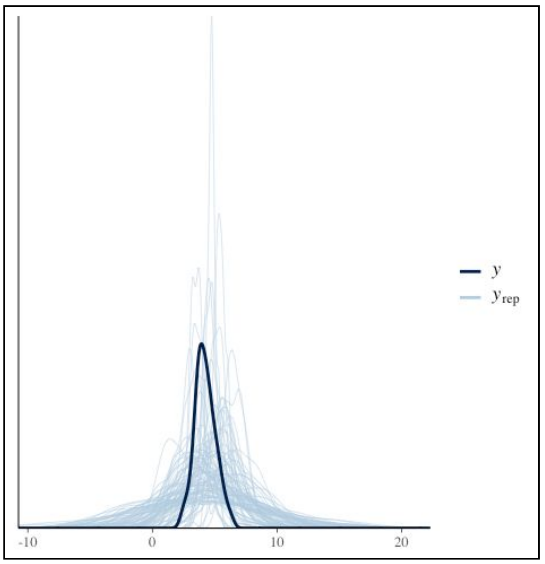


Figure 7: Prior predictive check for the model with Apathy

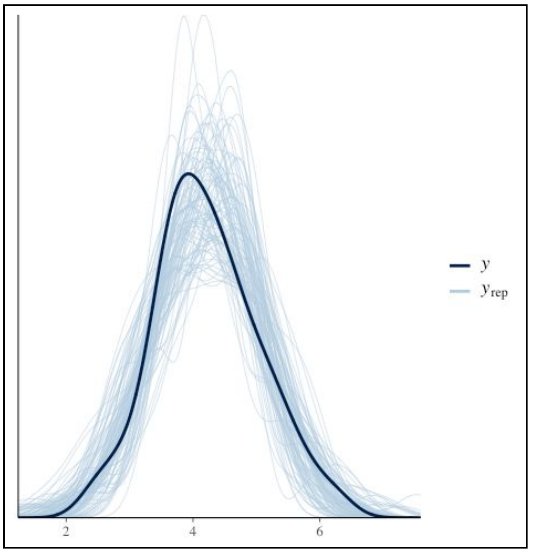


Figure 8: Posterior predictive check for the model with Apathy

Multiple Regression Model

As none of the symptoms seem to predict altercentric intrusion especially well on their own, we suspect that there might be a masking effect of the other variables. To unmask this effect, we have chosen to include all three symptoms in our final multiple regression model.

Altercentric Intrusion ~ 1 + Voice Hearing + Mind Reading + Apathy

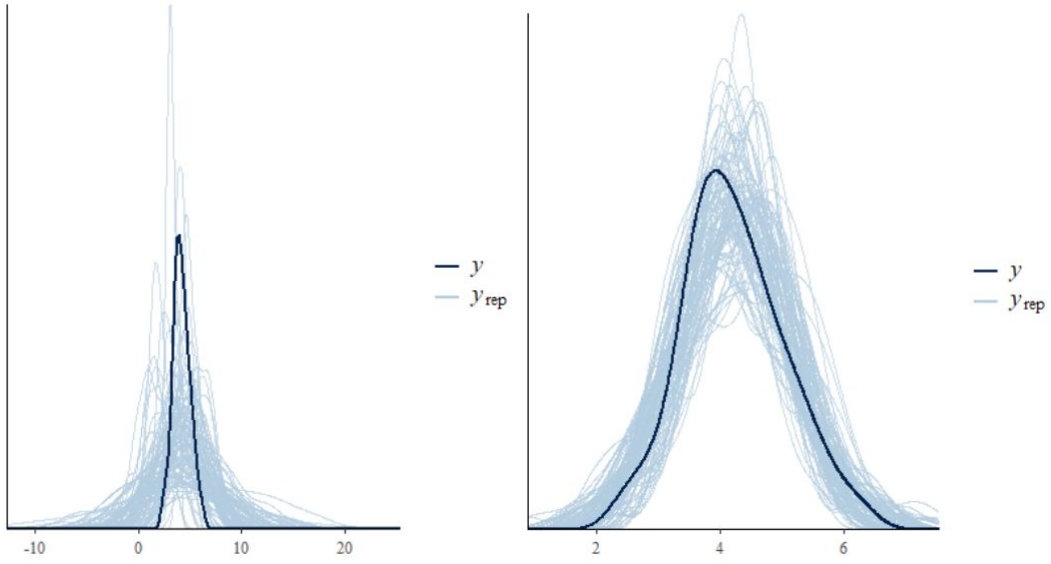


Figure 9 (left): Prior predictive check for our multiple regression model
Figure 10 (right): Posterior predictive check our multiple regression model

The effects for the multiple regression model is very similar to the effects of the symptoms in the models above: There is a small positive effect for both voice hearing (b = 0.05, CIs = -0.26, 0.36) and mind reading (b = 0.03, CIs = -0.25, 0.30) on altercentric intrusion, while there is a small negative effect of apathy (b = -0.21, CIs = -0.53, 0.11). However, all with credibility intervals crossing zero, making the estimated effects unlikely (see figure 12).

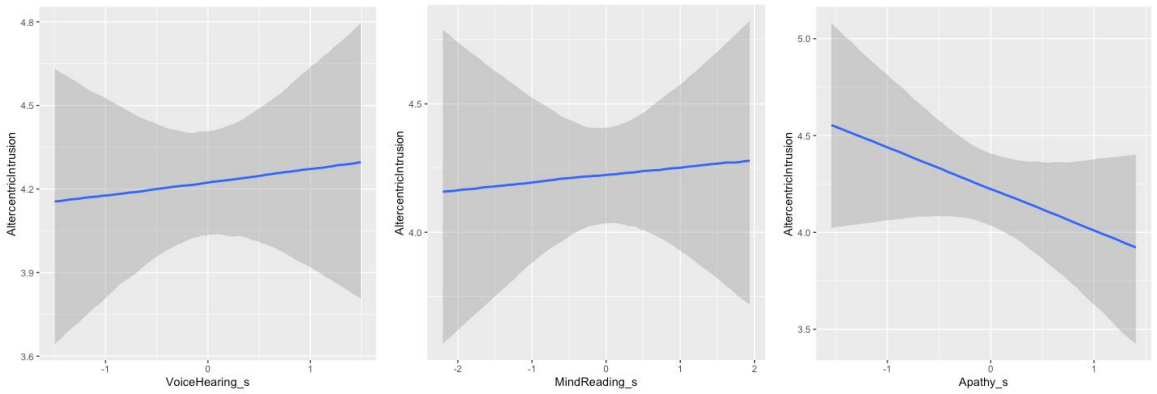


Figure 11: Plotting the effect of the three predictors (left = Voice Hearing, middle = Mind Reading, right = Apathy), CI = 95%.

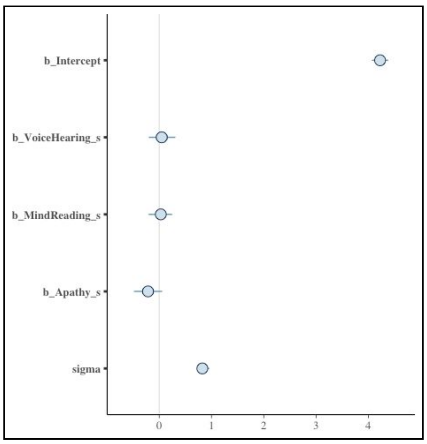


Figure 12: coefficient plot, all parameters overlap with 0, CI = 95%. As such, the plot shows how bad the model is.

Second part

Q2.1) Draw a causal graph (Directed Acyclical Graph) of the variables. Discuss which biases you might have introduced

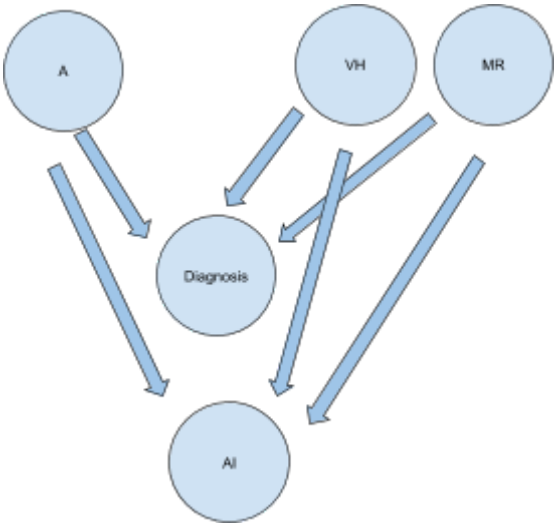


Figure 13: DAG plot showing causal inferences. AI = Altercentric Intrusion, VH = Voice Hearing, MR = Mind Reading, A = Apathy

The multi regression model we made above was made on a subset of the data, where we only included the schizophrenics. However, subsetting the data like this introduced a bias in our model (see figure 13). Both altercentric intrusion and the symptoms that we used as predictors cause schizophrenia - i.e. the higher score one has in the different symptoms, the more likely the person is to have schizophrenia, and also the higher score one have in altercentric intrusion, the more likely the person is to have schizophrenia.

Q2.2.) Redesign your analysis following the graph and report how the results change
To redesign the study, we refrain from conditioning on diagnosis. In other words, we run the same model, but on data for both controls and people with schizophrenia. Thereby, we have closed the backdoor so the same variance will not flow to altercentric intrusion more than once.

By running the model on the complete dataset the coefficients change. There is a positive effect for both voice hearing (b = 0.15, CIs = 0.05, 0.25) and mind reading (b = 0.16, CIs = 0.05, 0.27), and the credible intervals no longer overlap with zero, which indicates a small probability that the effect is zero - i.e. no effect (see figure 14). However, even though we hypothesized that apathy would have a negative impact on altercentric intrusion - it most likely has not. There is a small positive effect of apathy (b = 0.01, CIs = -0.10, 0.13), though, very uncertain. Furthermore, we made a model without apathy predicting altercentric intrusion from voice hearing and mind reading alone. This is illustrated with the pseudo code underneath. The results can be seen in the right plot of figure 14. The estimates of the other predictors do not change markedly.

Altercentric Intrusion ~ 1 + Voice Hearing + Mind Reading

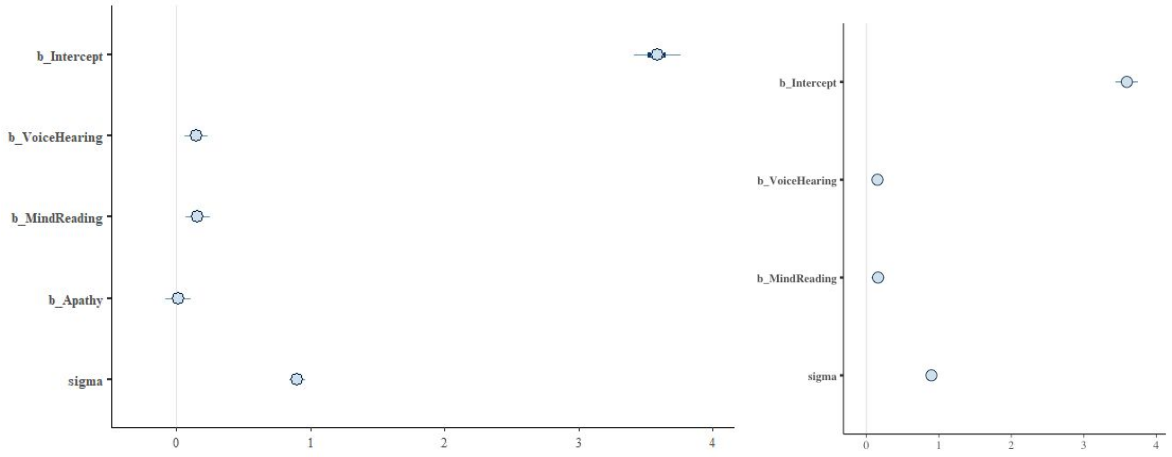


Figure 14: Coefficient plots using the entire dataset (left = multi regression model including Apathy, right = multi regression model excluding Apathy), CI = 95%

Third part

Q3.1) Look through the code and identify whether the results you have match the underlying truth. (or if the direction at least is the same). Discuss what you have learned.
It turns out that apathy did not predict any variance in altercentric intrusion, as we thought it would have in our model. The true model is shown in figure 15 and only differs by the missing causal link between apathy and altercentric intrusion, that we anticipated (see figure 13).

When conditioning on diagnosis, we did not find the effects of mind reading and voice hearing. When we ran the same model on the entire data set with both schizophrenics and controls, we found a positive effect of both mind reading and voice hearing (as one would according to the true model) and no effect of apathy, even though we added this as a predictor.
The beta estimates were not precisely 0.2. They were 0.15 and 0.16 so very close to the true effect of the symptoms. This could be due to the priors of the betas. We defined the mean of the prior as 0.1 for both betas and this might have 'drawn' the estimate away from 0.2 and closer to 0.1, leaving out estimates a bit low compared to the 'true' estimates. However, the true effect lies within the CIs and one would always expect some noise.

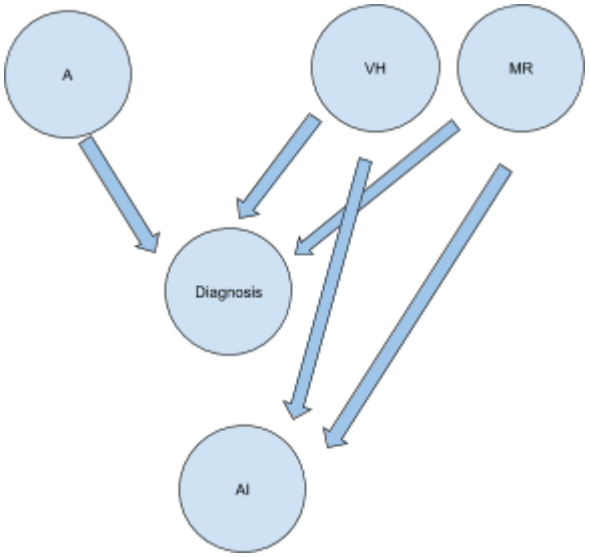


Figure 15: The "true" model