

Causal Inference Demo

Alex van Vorstenbosch

15 April, 2022

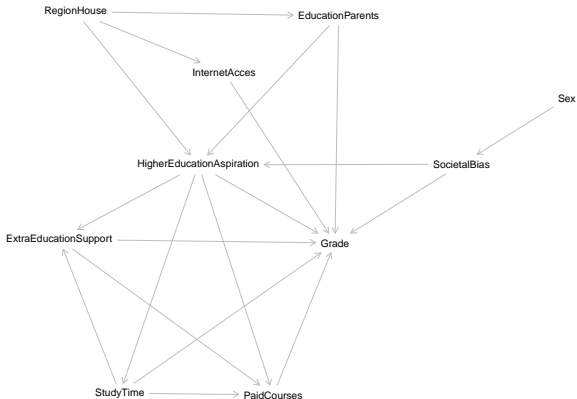
What are we going to see

- Defining the problem
- DAG
- Simulating the data
- Can we use BRMS to retrieve the parameters
- **Real World Data...**
- Is our dag valid?
- Results

Defining the Problem

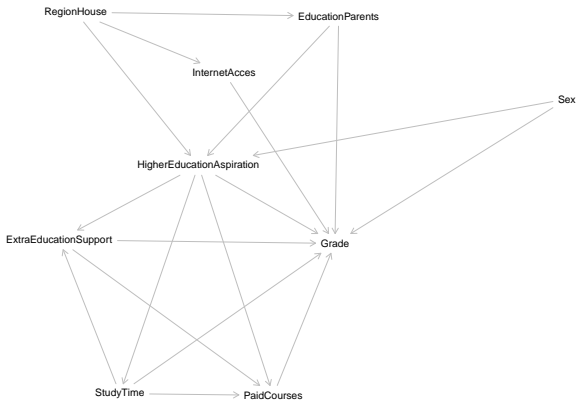
- Mathematics is one of THE fundamental school subjects. . .
- . . . yet, it is stereotypically seen as boring and difficult.
- In this demo we explore what influences school performance in Math
- . . . and possible intervention paths we could use.
- As outcome variable we take the end of year grade in Mathematics for high school students.

DAG - The Full Causal Pentagram



The unmeasured variable SocietalBias explains how the student's Sex effects Grade and HigherEducationAspiration. However, since Sex is only a 'Mechanism', we can remove it from the DAG and simplify it.

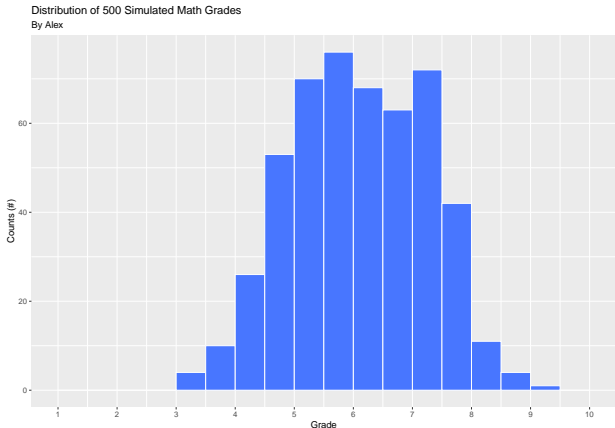
DAG - The (slightly simplified) Causal Pentagram



This DAG is slightly simpler, but the 'Causal Pentagram' looks like it might cause problems. Time to find out!

Simulate data

Simulated grades distribution

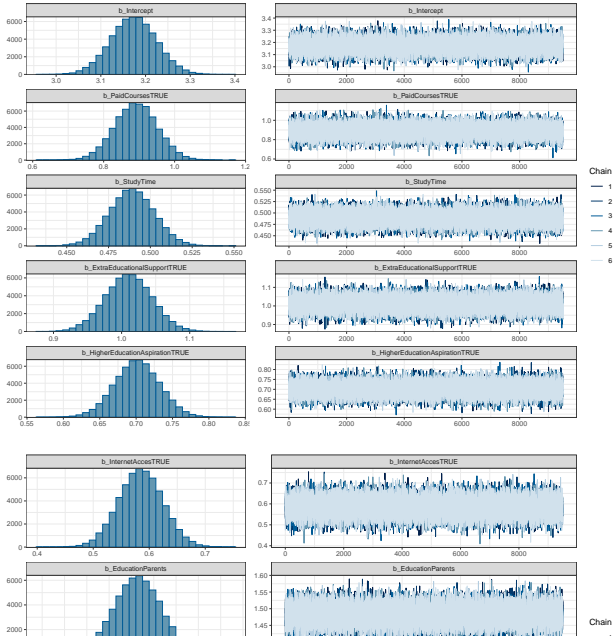


In our simulation, Most students score between a 5.5 and a 6. This seems reasonable. Also, very high scores are rare, as well as extremely low scores. The distribution is roughly normal, but not quite.

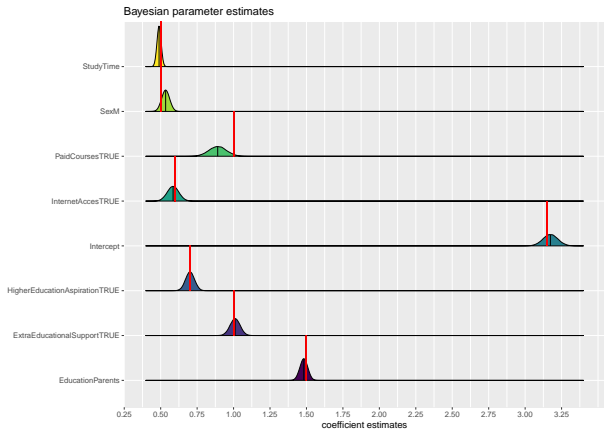
Retrieve these direct effects using BRMS

```
## Running MCMC with 6 parallel chains...  
##  
## Chain 3 finished in 4.6 seconds.  
## Chain 1 finished in 5.1 seconds.  
## Chain 4 finished in 4.9 seconds.  
## Chain 2 finished in 6.6 seconds.  
## Chain 6 finished in 7.0 seconds.  
## Chain 5 finished in 8.6 seconds.  
##  
## All 6 chains finished successfully.  
## Mean chain execution time: 6.1 seconds.  
## Total execution time: 9.2 seconds.
```

Checking our posterior and chains!

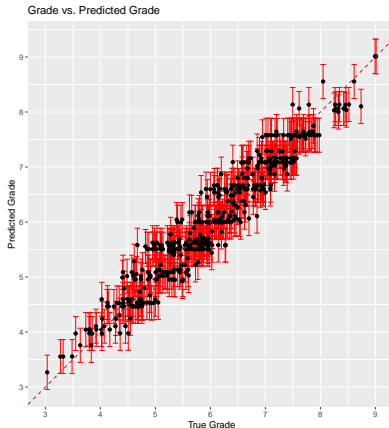


Bayesian Parameter Estimates



Almost all Parameter values are perfectly retrieved! PaidCourses is slightly underestimated, And Sex is slightly overestimated. But these results are very promising

How does the model perform in terms of predictions?



- This looks good! Most estimates overlap the true value within 1 sigma deviation.
- Estimates seem consistent across the whole domain