Exercises week 6

Hello :)

Today's class has three components:

1. Brief follow-up on the open task in the portfolio

2. Conceptual exercise on confounders

3. An exercise on coming up with DAGS and analyzing them

The first half of the class. is for steps 1 and 2, the second half for step 3 (roughly).

Below are the exercises for 2 and 3.

Enjoy :)

# Exercise 1: Conceptual Exercise on Confounders

This is an **'explain-to-your-group' exercise**. That means that each group member gets a different theme and has 10 minutes to prepare. Then each group member takes 10 minutes to present to the others (and to discuss etc).

The point of the presentation is ***not*** to do a completely perfect presentation of the content; it's somewhere to start a discussion, and, importantly, often you find out where you don't understand completely when you have to explain it.

For this exercise group members will be responsible for presenting on the **different kinds of biases / confounders** that is explained in chapter 6 in the book.

One member presents on **Multicollinearity (Section 6.1 in the book).**

One member presents on Post-Treatment Bias **(Section 6.2 in the book).**

* Is is normal to worry about removing variables from a model. What is not very normal is to worry about adding new variables that might be causal to other variables: This is the post-treatment bias.
* Blindly tossing variables into the causal salad is never a good idea, no matter how the data were collected.
  + Example
    - You’re growing plant in a green house and you want to measure what affects its growth. You end up with four parameters:
    - The problem is that fungus is mostly a consequence of treatment.

One member presents on Collider Bias **(Section 6.3.1 in the book).**

The fourth member (if there is one) explains the Haunted Dag example **(Section 6.3.2 in the book).**

If there are more than four members, they team up with others.

So to sum up:

10 minutes looking through that bit of the book. Make sure you understand **the example, the DAG** that is used, the **plots**, **the linear regressions** being used, and **how they result in bad inferences.** Stuff like how he chooses the priors, and how it is coded in R, is less important.

Then take turns one at a time explaining the example and talking about it. Spend 10 minutes: if you're done early, go on to the next person. Ask questions if necessary.

Peter will keep an eye on how everyone are doing :)

# Exercise 2: DAGS and statistics

**Step 1: The DAG**

- **Come up with an** incredibly interesting and scientifically important made-up **example** for a phenomenon to investigate. Decide on two variables (an outcome and a predictor) that you would like to investigate the relation between. If in doubt, you **can be inspired by Peter's amazing example** on the next page.

- **Make a DAG** for the phenomenon. Make it medium complicated, or so that there are interesting answers to the following steps. Code it in daggity, draw it somehow (on paper, in R, laser engraved in diamond).

- Find **elemental forms of variable relations** in the DAG. (See overview and procedure from the book on the last page).

- Find out **what variables to include (and not include)** in a multiple linear regression to avoid 'back door' (AKA non-causal) paths. Do this first with your eyes and your mind, and then by using daggity. If there is nothing in the DAG, maybe make a more interesting one.

- Find out which **conditional independencies** the DAG implies. First mind, then daggity. If there are none in the DAG, maybe make a more interesting one.

- Find the full list of **Markov equivalent** DAGS. Use daggity. It's for your own good.

**Step 2: The data**

- **Simulate some data that fits the DAG.** There are many ways to do this. Ask if in doubt. A simple trick is just to sample one variable from a normal distribution which has another variable as mean. McElreath does this in the book a few times.

**Step 3: Statistics**

- Run **multiple linear regression**s to **test the conditional independencies** **implied by your DAG**. Make sure to avoid backdoor paths. See that the linear model shows the conditional independencies implied by your DAG (if they don't the data and the DAG doesn't fit).

**Step 4: Messing it up**

- Try and **deliberately have an open back door path** and see if you can get wrong inference.

- Try and deliberately **simulate some data that doesn't fit the DAG**, or **create a new DAG that doesn't fit the data**.

- Use the same approach as above to **show that the DAG is wrong** (by showing that conditional independencies don't exist in the data, for example).

**Peter's perfectly optimal and extremely interesting example**

*In a galaxy far, far away...*

*It is a period of civil wars in the galaxy. A brave alliance of underground freedom fighters has challenged the tyranny and oppression of the awesome GALACTIC EMPIRE.*

*To crush the rebellion once and for all, the EMPIRE is constructing a sinister new battle station. Powerful enough to destroy an entire planet, its completion spells certain doom for the champions of freedom.*

*The evil Emperor has figured out, however, that neither the battle station nor the Force can help him avoid that more solar systems join the rebellion. He has therefore hired a CogSci student to use causal modelling and multiple linear regressions to investigate how the activity of the Death Star and other factors affects the probability that a given solar system will join the rebellion (this allows him to more optimally suppress freedom in the Galaxy).*

*You are that student.*

We assume that  
The probability of a solar system joining the rebellion depends on  
- how many rebellion sympathizers there is in the system (more rebels -> higher probability of joining the rebellion)  
- how scared people are in the system (more scared -> lower probability of joining the rebellion).

How many rebellion sympathizers there is in a system depends on

- crime levels (less crime -> less rebellion sympathizers)

- number of planets recently destroyed by the Death Star (more planets destroyed -> more rebellion sympathizers)

- number of Jedis in the system (more Jedis -> more rebellion sympathizers)

How scared people are depends on

- whether or not the Death Star is nearby (nearby -> more scared)

- how many jedis are in the system (more Jedis -> less scared)

- how much time Darth Vader has spent in the system recently (more time -> more scared)

crime levels depend on

- number of planets recently destroyed by the Death Star (more planets destroyed -> less crime)

And so on....

