More DAGs; models with interactions

ISTA 410 / INFO 510: Bayesian Modeling and Inference

U. of Arizona School of Information

March 31, 2021

Outline

Previously:

- Causal DAGs and various forms of confounding
- The backdoor criterion
- Intro to interactions

Today:

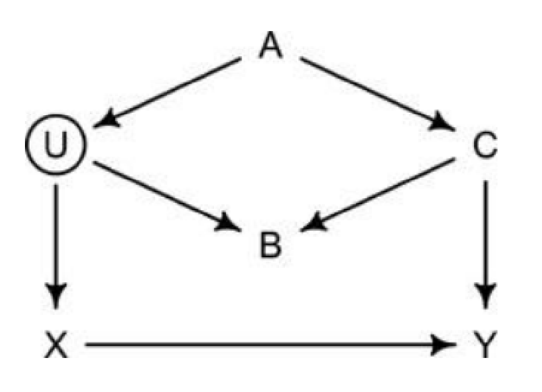
- Unobserved variables and their consequences
- Linear models with interactions

PSA: talk by Andrew Gelman tomorrow 3PM MST

Unobserved variables

Unobserved variables in DAGs

Last time we saw a DAG with an unobserved variable:



We can't control for unobserved variables

An unobserved fork

Example: stress and severity of COVID-19 disease

Imagine the following possible scenario:

- Want to estimate whether stress markers predict the severity of symptoms
- Suspected relationship: high stress increases probability of severe disease
- Grab a data set of early confirmed positive tests, regress severity on stress

An unobserved fork

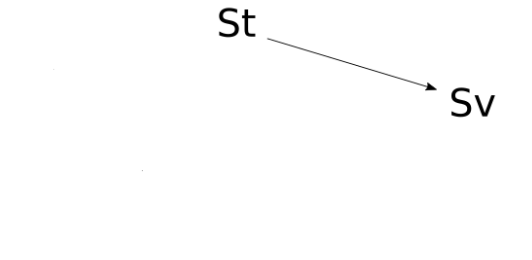
Example: stress and severity of COVID-19 disease

Imagine the following possible scenario:

- Want to estimate whether stress markers predict the severity of symptoms
- Suspected relationship: high stress increases probability of severe disease
- Grab a data set of early confirmed positive tests, regress severity on stress
- Surprising result: negative association

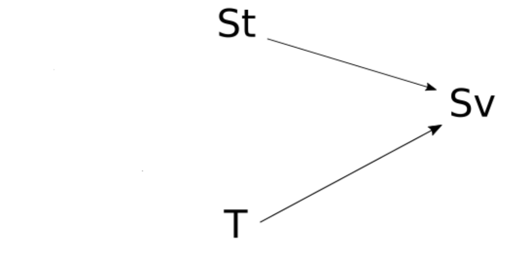
Why this result?

- Possibly, high stress decreases severity – but this would go against all we know about stress and immune response
- Possibly, we have confounding:



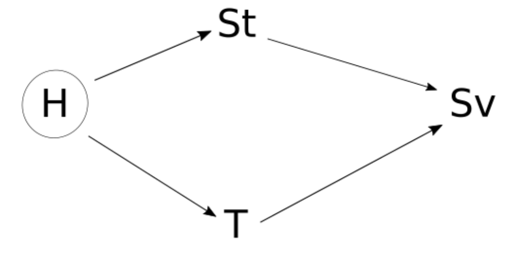
Why this result?

- Possibly, high stress decreases severity – but this would go against all we know about stress and immune response
- Possibly, we have confounding:



Why this result?

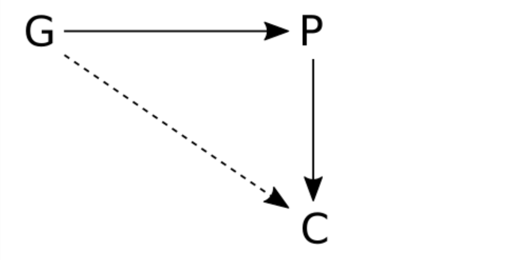
- Possibly, high stress decreases severity – but this would go against all we know about stress and immune response
- Possibly, we have confounding:



Collider bias from unobserved variables

Example from Statistical Rethinking: does the education level of grandparents influence the educational achievement of their grandchildren?

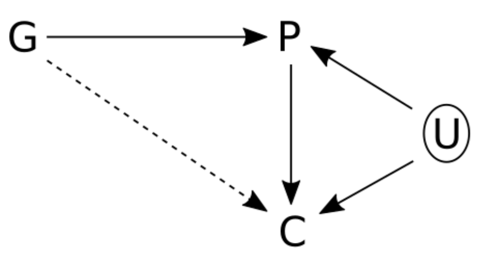
- Goal: infer direct effects of parents (P) and grandparents (G) on children (C)
- Reasonable to assume that parents influence children, grandparents influence parents; do grandparents also directly influence children?



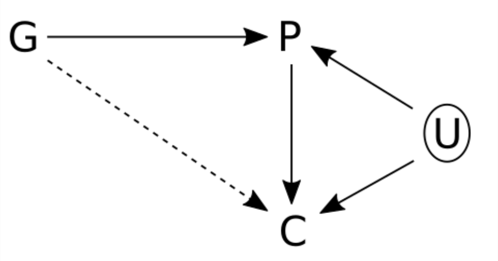
Collider bias from unobserved variables

Example from Statistical Rethinking: does the education level of grandparents influence the educational achievement of their grandchildren?

- Goal: infer direct effects of parents (P) and grandparents (G) on children (C)
- What if there is an unobserved variable (U) – e.g., neighborhood – that influences parents and children?

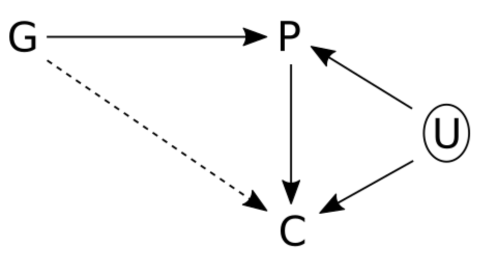


Collider bias from unobserved variables



- To isolate the direct effect $G \to C$, we must control for P
- Controlling for P opens the collider path through U

Collider bias from unobserved variables



- To isolate the direct effect $G \rightarrow C$, we must control for P
- Controlling for P opens the collider path through U
- We're stuck!

DAG summary

In summary:

- DAGs give a heuristic model for causal links between variables
    - Not a mechanistic model – doesn't tell you the functional effect
- Can provide a strategy for deconfounding models – or tell you such a strategy doesn't exist / requires different data

Modeling interactions

Interactions

Interaction effects:

- Interaction of one predictor is *conditional* on another
  - Effect of *water* on *plant growth* is conditional on *sunlight*
  - Effect of *gene* on *survival* is conditional on *environment*
  - Effect of *total traffic* on *bike traffic* is conditional on *bike lane*
- Interactions appear frequently in real systems

Interactions in DAGs

Here is what an interaction looks like in a DAG:

Interactions in DAGs

Here is what an interaction looks like in a DAG:

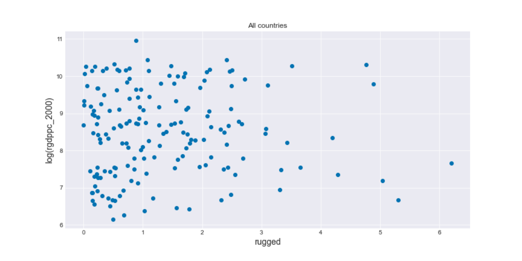
Any time two variables influence a third leaves the possibility for interaction!

Terrain "ruggedness" and economy

(Example from Statistical Rethinking Ch7.) What is the relationship between the geographic terrain in a nation and its economy?

Data: observations on many countries

- Outcome: log GDP (as of 2000, when data was collected)
- Predictor: terrain "ruggedness" index

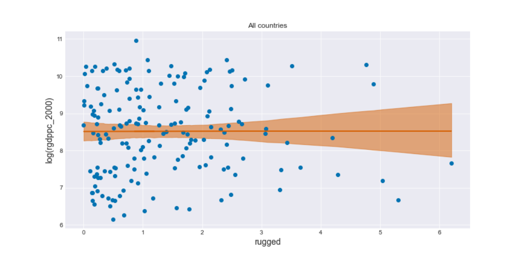


Terrain "ruggedness" and economy

(Example from Statistical Rethinking Ch7.) What is the relationship between the geographic terrain in a nation and its economy?

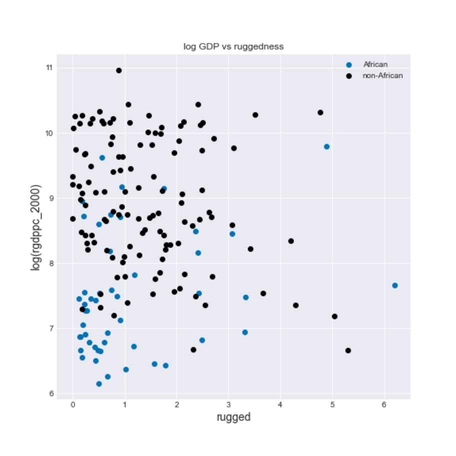
Data: observations on many countries

- Outcome: log GDP (as of 2000, when data was collected)
- Predictor: terrain "ruggedness" index



Terrain ”ruggedness” and economy

Closer examination of the data reveals an interesting phenomenon: the relationship is different for countries in Africa.



Terrain ”ruggedness” and economy

Conditionality:

- The effect of ruggedness on modern economy is *conditional* on continent
- African nations respond differently to ruggedness than non-African nations

Want to incorporate this effect into a model; ideally, a *single* model

- Pooling data gives better estimates of continent-independent parameters
-
- Model comparison techniques assume the same data

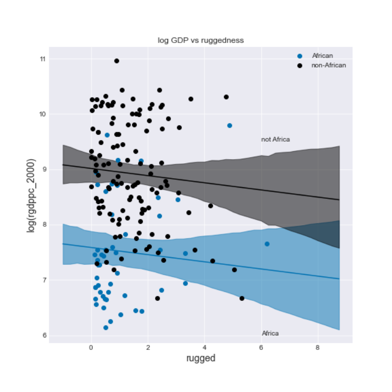
A simple approach that won't work

A simple approach that's not quite good enough: add an indicator variable for African countries, and do a bivariate regression:

$$
\begin{aligned}
\log GDP &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_R R_i + \beta_A A_i \\
\beta_R &\sim \text{Normal}(0, 1) \\
\beta_A &\sim \text{Normal}(0, 1) \\
\alpha_j &\sim \text{Normal}(9, 3) \\
\sigma &\sim \text{HalfCauchy}(5)
\end{aligned}
$$

A simple approach that won't work

Problem:



Allows for a shift, but not a change in slopes.

Can see this also with the fox problem from last week.

Allowing interactions

To add interactions:

$$\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_R R_i + \beta_A A_i + \beta_{AR} A_i R_i \\
\beta_R &\sim \mathrm{Normal}(0, 1) \\
\beta_A &\sim \mathrm{Normal}(0, 1) \\
\beta_{AR} &\sim \mathrm{Normal}(0, 1) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}$$

So we have a third slope, for the *product* of R and A.

(note: could also use indexed slopes!)

Why is this the approach?

Where this comes from: just model the slope β_R as being itself a linear function of A.

$$
\begin{aligned}
\log GDP &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \gamma_i R_i + \beta_A A_i \\
\gamma_i &= \beta_R + \beta_{AR} A_i \\
\beta_R &\sim \text{Normal}(0, 1) \\
\beta_A &\sim \text{Normal}(0, 1) \\
\beta_{AR} &\sim \text{Normal}(0, 1) \\
\alpha_j &\sim \text{Normal}(9, 3) \\
\sigma &\sim \text{HalfCauchy}(5)
\end{aligned}
$$

Why is this the approach?

Plug a linear equation into another linear equation:

$$\mu_i = \alpha + \gamma_i R_i + \beta_A A_i$$

$$\gamma_i = \beta_R + \beta_{AR} A_i$$

Why is this the approach?

Plug a linear equation into another linear equation:

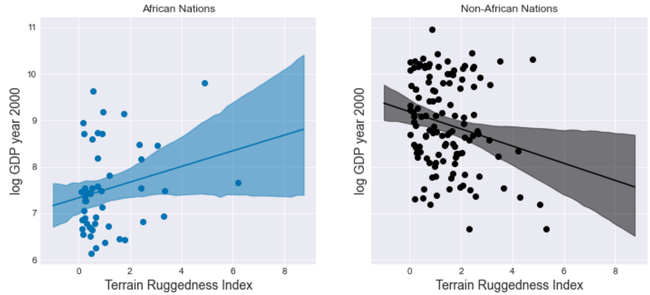
$$\mu_i = \alpha + \gamma_i R_i + \beta_A A_i$$

$$\gamma_i = \beta_R + \beta_{AR} A_i$$

$$\mu_i = \alpha + (\beta_R + \beta_{AR} A_i) R_i + \beta_A A_i$$

The result

Result from the interaction model:



Here, we can see the different slope. Ruggedness has a positive association with GDP for African nations, negative for others.

The Africa interaction as a multilevel model

When we did multilevel/hierarchical regression:

- Observations grouped into subpopulations
- Each subpopulation gets its own model parameters, drawn from a common hyperprior

Ruggedness model: different slope depending on continent

- We could set up the model in a multilevel way by including two slopes explicitly, drawing from a hyperprior

The Africa interaction as a multilevel model

Setting it up as a multilevel model:

$$
\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{C(i)} + \beta_{C(i)} R \\
\beta_j &\sim \mathrm{Normal}(0, \tau) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \\
\tau &\sim \mathrm{HalfCauchy}(5) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}
$$

Plate diagram for the multilevel version

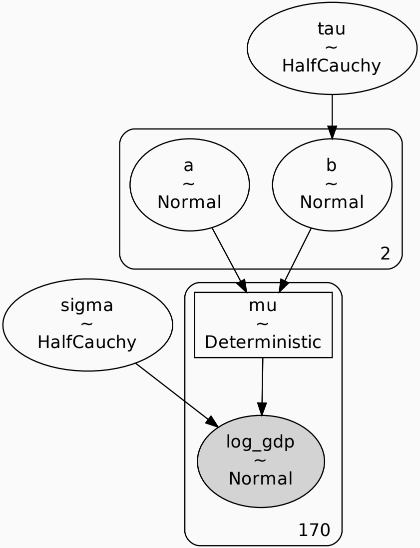
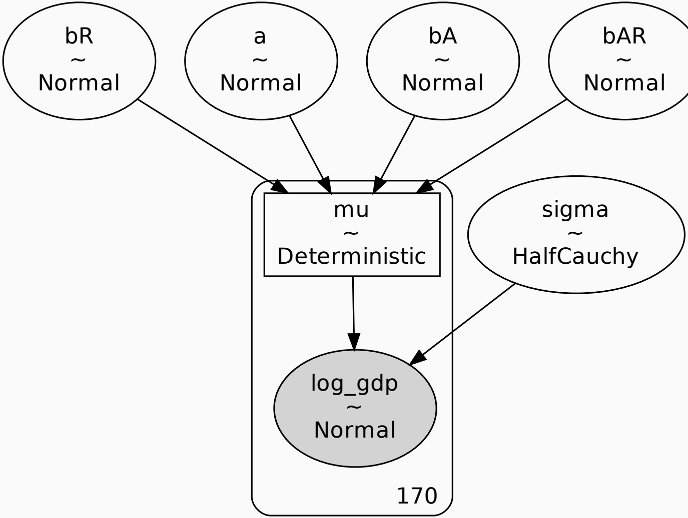
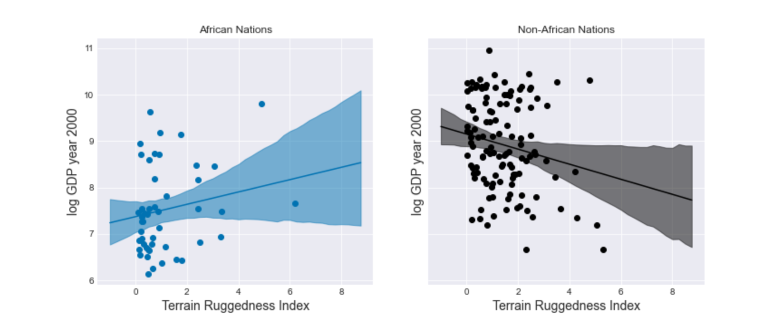


Plate diagram for the product version



The Africa interaction as a multilevel model

Result from the multilevel version:



Essentially the same result as the conventional interaction version.

Let's inspect the parameters.

Comparing parameters

- Parameters from the product-interaction model:

	mean	sd	hdi_3%	hdi_97%
a	9.182	0.144	8.913	9.442
bR	-0.184	0.078	-0.332	-0.040
bA	-1.842	0.223	-2.275	-1.445
bAR	0.350	0.130	0.131	0.618

- Parameters from the product-interaction model:

	mean	sd	hdi_3%	hdi_97%
a[0]	9.191	0.140	8.931	9.458
a[1]	7.308	0.179	6.979	7.639
b[0]	-0.178	0.079	-0.333	-0.037
b[1]	0.160	0.107	-0.050	0.354

Model comparison

Comparing all models: pooled, non-interacting, product, multilevel

	rank	loo	p_loo	d_loo	weight	se	dse
product	0	-234.729809	5.115108	0.000000	0.901376	7.301258	0.000000
multilevel	1	-234.975806	5.189719	0.245997	0.000000	7.223924	0.493125
noninteracting	2	-238.015613	4.111312	3.285805	0.098624	7.362950	3.013388
pooled	3	-269.811536	2.661599	35.081727	0.000000	6.489850	7.327695

Multilevel models as interaction machines

Multilevel models:

- Group observations into clusters
- Parameters for each cluster drawn from a common hyperprior

Multilevel models as interaction machines

Multilevel models:

- Group observations into clusters
- Parameters for each cluster drawn from a common hyperprior
- This is just a way of expressing that the parameter values are conditional on group membership
- Multilevel models give a way to encode interactions *and* regularize automatically across groups

Not surprising: remember we constructed the interaction by putting a linear model inside a linear model

Higher order interactions

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Pairwise interactions:

$$\begin{aligned}\mu_i = \alpha &+ \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &+ \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i}\end{aligned}$$

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Pairwise interactions:

$$\begin{aligned}\mu_i &= \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &\quad + \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i}\end{aligned}$$

Higher-order interactions:

$$\begin{aligned}\mu_i &= \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &\quad + \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i} \\ &\quad + \beta_{123} x_{1,i} x_{2,i} x_{3,i}\end{aligned}$$

The "Judgement of Princeton"

The Judgement of Princeton

- 9 judges, 20 wines
- Wines split between red and white, NJ or France
- Judges split between American or French/Belgium

Predictors:

- Wine color: red or white
- Wine origin: NJ or France
- Judge nationality: US or EU

Interactions

Potential for interactions between all predictors:

- Interaction between origin and judge: judge bias.
  Judge bias might depend upon color.
- Interaction between color and judge: taste preference.
  Taste preference might depend upon origin.
- Interaction between origin and color: relative advantage.
  Advantage might depend upon judge.

Summary

Today:

- DAG wrap up
- Interaction effects

Next week:

- GLMs
- covariance among parameters
- Multilevel linear models