

More DAGs; interactions

ISTA 410 / INFO 510 - Bayesian Modeling and Inference

University of Arizona School of Information

September 20, 2021

Last time:

- Multiple regression
- Causal DAGs

Today: more DAGs; modeling interactions

Chapter 6/8 of *Rethinking* (we'll do 7 after)

HW2 available now; due Friday 10/1

DAG recap

Graphical example

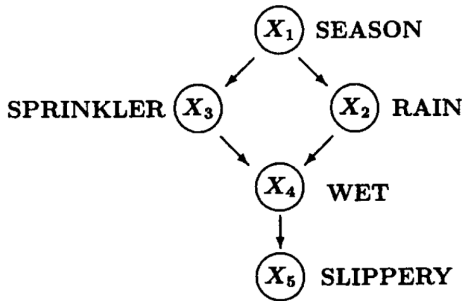


Figure from *Causality* by Judea Pearl

Graphical example

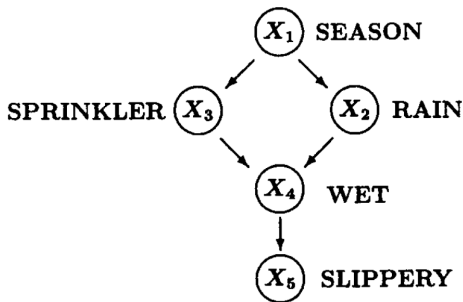


Figure from *Causality* by Judea Pearl

$$P(x_1, \dots, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)$$

Functional model of a DAG

Functional model: each variable X_i satisfies an equation in the graph:

$$x_i = f_i((pa_i), u_i)$$

where

- f_i is a function of the parent variables
- pa_i refers to the parent nodes of X_i in the graph
- u_i represents the unobserved and/or random components of the model

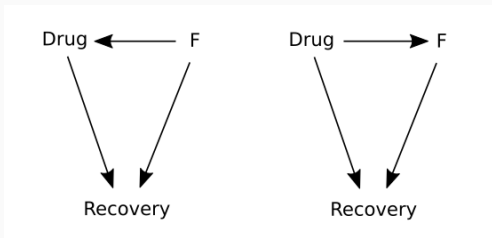
Elementary paths

In a DAG, information flows along paths (both with and against the arrows)

- Fork path: $X \leftarrow Z \rightarrow Y$: blocked by conditioning on Z , because DAG property says X, Y conditionally independent conditional on Z
- Chain path: $X \rightarrow Z \rightarrow Y$: blocked by conditioning on Z , because DAG property says X, Y conditionally independent conditional on Z
- Collider path: $X \rightarrow Z \leftarrow Y$: unblocked by conditioning on Z ; X and Y have no parents in common, so are independent until conditional on Z

Two DAGs

The data from the previous slide could be generated by processes represented by either of the following causal DAGs:



But the inference we should make about the effectiveness of the drug is very different in each case!

The backdoor criterion

A (possibly undirected) path p through a DAG G is said to be *d-separated* or *blocked* by a set of nodes Z if:

1. p contains a chain $X_i \rightarrow M \rightarrow X_j$ or fork $X_i \leftarrow M \rightarrow X_j$ such that $M \in Z$; or,
2. p contains a collider $X_i \rightarrow M \leftarrow X_j$ such that $M \notin Z$ and no descendent of M is in Z .

The *d*-separation (blocking) definition for paths leads to another definition, for sets of variables.

The backdoor criterion

A related definition:

Definiton

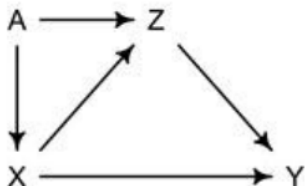
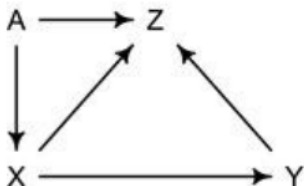
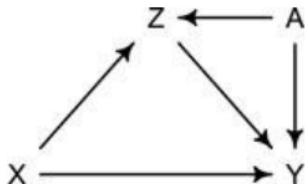
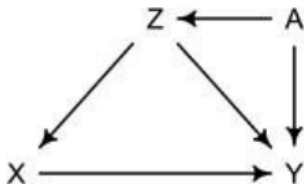
A set of variables Z satisfies the backdoor criterion with respect to an ordered pair of variables (X_i, X_j) in G if:

- 1. no node in Z is a descendent of X_i ; and,*
- 2. Z blocks every path from X_i to X_j that contains an arrow into X .*

To estimate the causal effect of X on Y , condition on a set of variables satisfying the backdoor criterion with respect to (X, Y) .

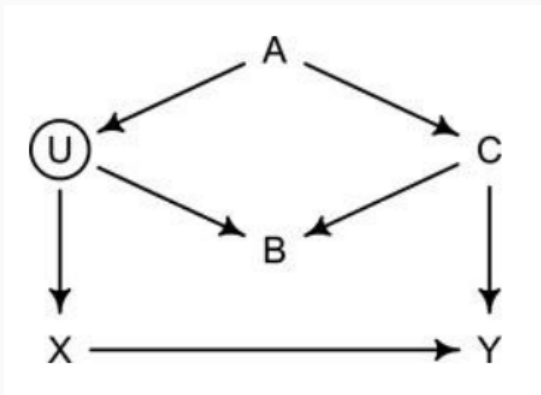
Group exercise

For each DAG, which variable should be conditioned on to estimate total causal influence of X on Y ?



Demonstrative example

Another DAG:



We want to estimate $X \rightarrow Y$. What should we condition on? (U is unobserved; we can't use it.)

Fake data simulation

To demonstrate the effect, let's use a fake data simulation:

$$A \sim \text{Normal}(0, 1)$$

$$U = A + \text{noise}$$

$$C = -2A + \text{noise}$$

$$B = -2C + 3U + \text{noise}$$

$$X = U + \text{noise}$$

$$Y = 1.5X + C + \text{noise}$$

In all cases, $\text{noise} \sim \text{Normal}(0, 0.1)$

What do we seek?

Before we run any regressions, what *should* we see?

- All paths $X \rightarrow Y$ except the direct one are non-causal (backdoor)
- We want to estimate the direct (causal) effect
- We know $Y = 1.5X + \text{other effects}$
- An unconfounded estimate of $\hat{Y} = \beta_x X + \text{others}$ should have $\beta_x \approx 1.5$

Let's go run some...

Let's see some regression results

- Regression including only X :

	mean	sd	hdi_3%	hdi_97%
alpha	0.070	0.031	0.011	0.125
beta_x	-0.465	0.033	-0.525	-0.403
sigma	0.302	0.023	0.261	0.342

Let's see some regression results

- Regression including only X :

	mean	sd	hdi_3%	hdi_97%
alpha	0.070	0.031	0.011	0.125
beta_x	-0.465	0.033	-0.525	-0.403
sigma	0.302	0.023	0.261	0.342

- Condition on A :

	mean	sd	hdi_3%	hdi_97%
alpha	0.027	0.013	0.006	0.054
beta_x	1.461	0.093	1.290	1.636
beta_a	-1.948	0.093	-2.124	-1.782
sigma	0.126	0.009	0.109	0.145

Let's see some regression results

- Regression including only X :

	mean	sd	hdi_3%	hdi_97%
alpha	0.070	0.031	0.011	0.125
beta_x	-0.465	0.033	-0.525	-0.403
sigma	0.302	0.023	0.261	0.342

- Condition on A :

	mean	sd	hdi_3%	hdi_97%
alpha	0.027	0.013	0.006	0.054
beta_x	1.461	0.093	1.290	1.636
beta_a	-1.948	0.093	-2.124	-1.782
sigma	0.126	0.009	0.109	0.145

- Condition on A and B :

	mean	sd	hdi_3%	hdi_97%
alpha	0.020	0.011	-0.002	0.040
beta_x	1.893	0.110	1.687	2.107
beta_a	-1.044	0.190	-1.428	-0.710
beta_b	-0.192	0.036	-0.263	-0.129
sigma	0.112	0.008	0.097	0.127

Let's see some regression results

- Using X alone: badly confounded
- Using X, A : good estimate of $X \rightarrow Y$
- Using X, A, B : confounded again

What other options do we have?

More variables

- Condition on C :

	mean	sd	hdi_3%	hdi_97%
alpha	0.018	0.009	0.002	0.034
beta_x	1.539	0.065	1.410	1.650
beta_c	1.017	0.033	0.957	1.076
sigma	0.087	0.006	0.075	0.098

More variables

- Condition on C :

	mean	sd	hdi_3%	hdi_97%
alpha	0.018	0.009	0.002	0.034
beta_x	1.539	0.065	1.410	1.650
beta_c	1.017	0.033	0.957	1.076
sigma	0.087	0.006	0.075	0.098

- Use everything:

	mean	sd	hdi_3%	hdi_97%
alpha	0.019	0.009	0.003	0.036
beta_x	1.559	0.100	1.372	1.756
beta_a	-0.233	0.186	-0.593	0.102
beta_b	-0.002	0.038	-0.076	0.067
beta_c	0.905	0.121	0.686	1.138
sigma	0.087	0.007	0.075	0.099

Using everything works because the collider path is blocked at C ;
but note precision

Unobserved variables in a DAG

The previous example had an unobserved variable, U :

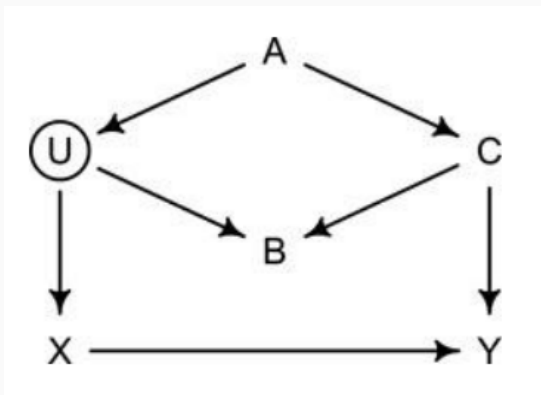
- If a variable is unobserved, then we can't stratify/adjust for it in the regression
- ...but that doesn't mean we are off the hook for thinking about it!
- Unobserved variables can confound estimates
- Unobserved variables can form colliders

Sometimes this means there is no way to make the estimates that you want!

Unobserved variables

Unobserved variables in DAGs

Last time we saw a DAG with an unobserved variable:



We can't control for unobserved variables

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

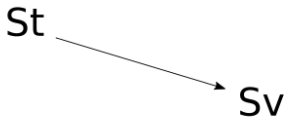
Example: stress and severity of COVID-19 disease

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

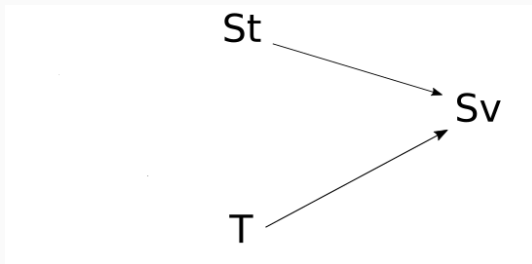
Why this result?

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



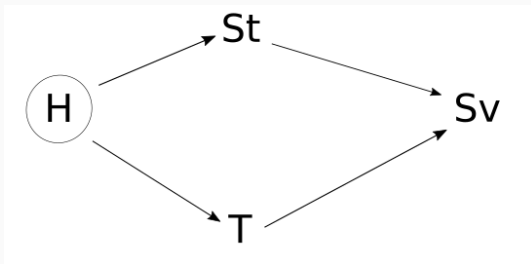
Why this result?

- 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?

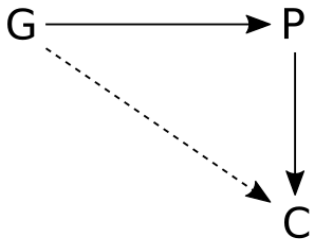
- Possibly, high stress decreases severity – but this would go against all we know about stress and immune response
- Possibly, we have confounding:



Collider bias from unobserved variables

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

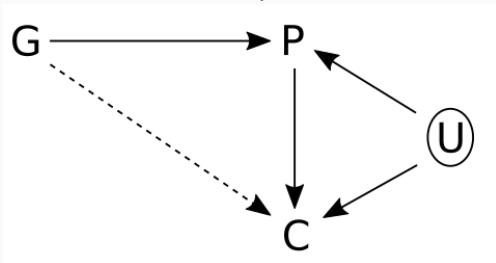
- Goal: infer direct effects of parents (P) and grandparents (G) on children (C)
- Reasonable to assume that parents influence children, grandparents influence parents; do grandparents also directly influence children?



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?



Fake data simulation

```
N = 200  # number of of grandparent-parent-child triads
b_GP = 1  # direct effect of G on P
b_GC = 0  # direct effect of G on C
b_PC = 1  # direct effect of P on C
b_U = 2   # direct effect of U on P and C

U = 2 * np.random.binomial(1, 0.5, N) - 1
G = np.random.normal(size=N)
P = np.random.normal(b_GP * G + b_U * U)
C = np.random.normal(b_PC * P + b_GC * G + b_U * U)
d = pd.DataFrame.from_dict({"C": C, "P": P, "G": G, "U": U})
```


Fitting a model

```
with pm.Model() as gp_model:  
    # Priors  
    a = pm.Normal("a", 0, 1)  
    p_PC = pm.Normal("b_PC", 0, 1)  
    p_GC = pm.Normal("b_GC", 0, 1)  
    sigma = pm.Exponential("sigma", 1)
```

```
    # Model equation
```

```
    mu = a + p_PC * d.P + p_GC * d.G
```

```
    # Likelihood
```

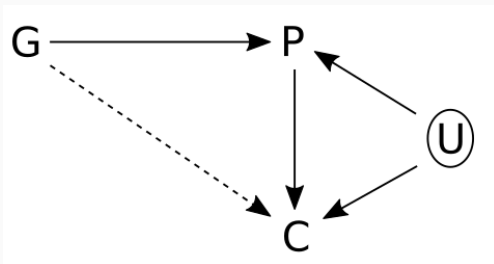
```
    pC = pm.Normal("C", mu, sigma, observed=d.C)
```

```
    # Inference
```

```
    qp = quap()
```

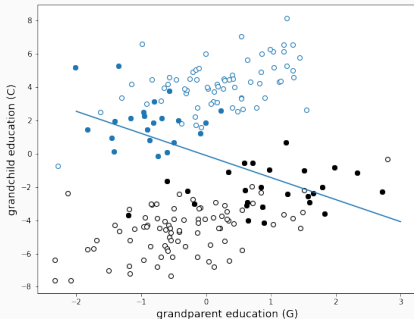
	mean	sd	hdi_3%	hdi_97%
a	-0.122	0.091	-0.292	0.049
b_PC	1.735	0.041	1.658	1.811
b_GC	-0.750	0.100	-0.939	-0.562
sigma	1.278	0.064	1.158	1.398

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!

Collider bias from unobserved variables



- for intuition, focus on 45-60th %ile of parents
- parents in good neighborhoods must have had less educated grandparents
- parents in bad neighborhoods must have had better educated grandparents
- else, they would be in different quantiles

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

Interaction effects:

- effect 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:



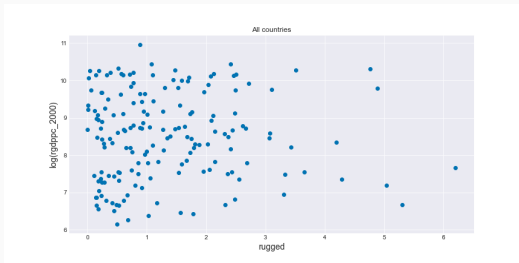
Any time two variables influence a third leaves the possibility for interaction!

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

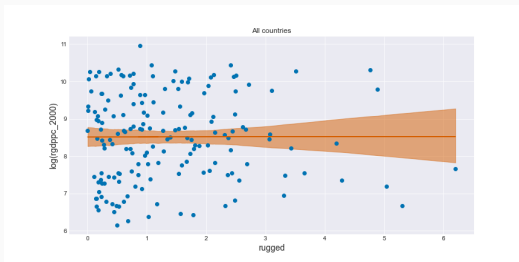


Terrain "ruggedness" and economy

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

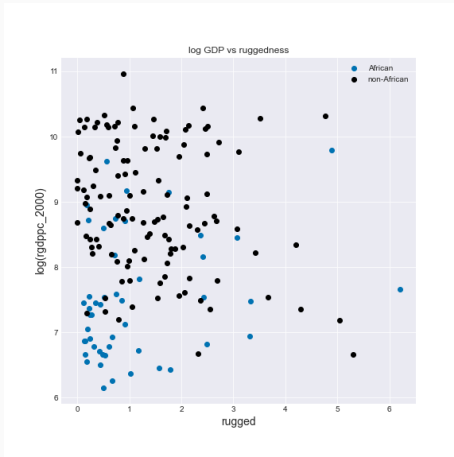
Data: observations on many countries

- 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

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:

$$\log GDP \sim \text{Normal}(\mu_i, \sigma)$$

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

$$\beta_R \sim \text{Normal}(0, 1)$$

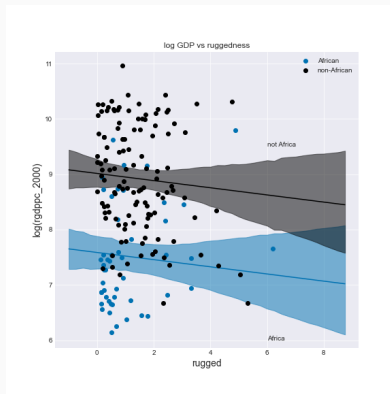
$$\beta_A \sim \text{Normal}(0, 1)$$

$$\alpha_j \sim \text{Normal}(9, 3)$$

$$\sigma \sim \text{HalfCauchy}(5)$$

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 earlier.

Allowing interactions

To add interactions:

$$\log GDP \sim \text{Normal}(\mu_i, \sigma)$$

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

$$\beta_R \sim \text{Normal}(0, 1)$$

$$\beta_A \sim \text{Normal}(0, 1)$$

$$\beta_{AR} \sim \text{Normal}(0, 1)$$

$$\alpha_j \sim \text{Normal}(9, 3)$$

$$\sigma \sim \text{HalfCauchy}(5)$$

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

Why is this the approach?

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

$$\log GDP \sim \text{Normal}(\mu_i, \sigma)$$

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

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

$$\beta_R \sim \text{Normal}(0, 1)$$

$$\beta_A \sim \text{Normal}(0, 1)$$

$$\beta_{AR} \sim \text{Normal}(0, 1)$$

$$\alpha_j \sim \text{Normal}(9, 3)$$

$$\sigma \sim \text{HalfCauchy}(5)$$

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:

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

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

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

Code for the interaction model

```
with pm.Model() as model_product:
    # Priors
    a = pm.Normal('a', mu=9, sd=3)
    bR = pm.Normal('bR', mu=0, sd=1)
    bA = pm.Normal('bA', mu=0, sd=1)
    bAR = pm.Normal('bAR', mu=0, sd=1)
    sigma = pm.HalfCauchy('sigma', 5)

    # Model equations
    gamma = bR + bAR * dd.cont_africa
    mu = a + gamma * dd.rugged + bA * dd.cont_africa

    # Likelihood
    log_gdp = pm.Normal('log_gdp', mu, sigma, observed=np.log(dd

    # Inference
    qp_product = quap()
```

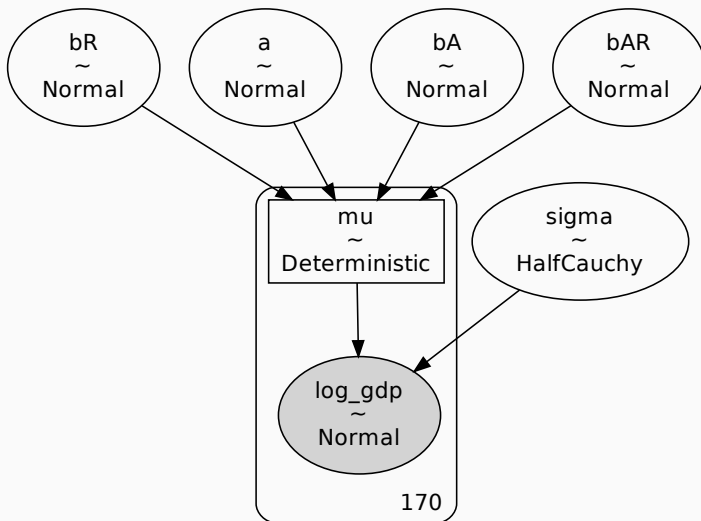
The result

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.

Plate diagram for the product version



The Africa interaction as two independent slopes

Alternative formulation of the model:

$$\log GDP \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha_{\text{CONT}[i]} + \beta_{\text{CONT}[i]} R_i$$

$$\beta_j \sim \text{Normal}(0, 1) \quad j \in \{0, 1\}$$

$$\alpha_j \sim \text{Normal}(9, 3)$$

$$\sigma \sim \text{HalfCauchy}(5)$$

- *index variable*
- allows the model to simply select separate slopes independently

Aside: categorical variables

Categorical variables in regression

Ways to handle a categorical variable c :

- indicator variable
- if c is binary, assign value 0 to one category, 1 to another

$$\mu_j = \alpha + \beta_c c$$

Potential problem: more uncertainty in one category than another

Example: giraffe height

Simple example: modeling giraffe height stratified by sex. Assign $s = 0$ for female, $s = 1$ for male.

$$h \sim \text{Normal}(\mu, \sigma)$$

$$\mu = \alpha + \beta s$$

$$\alpha \sim \text{Normal}(5, 1.5)$$

$$\beta \sim \text{Normal}(0, 0.4)$$



- $\text{Var}\mu$ for males = $\text{Var}\alpha + \text{Var}\beta$
- $\text{Var}\mu$ for females = $\text{Var}\alpha$

Alternatives: one-hot encoding or index variables

One alternative: one-hot encoding

- create an indicator variable for every category

$$\mu = \beta_f f + \beta_m m$$

Why drop α ? Not enough constraints:

- Suppose female giraffes average 4.8 m, males 5.1; then which is correct?

$$\mu = 5 + 0.1m - 0.2f$$

$$\mu = 4 + 1.1m + 0.8f$$

Alternatives: one-hot encoding or index variables

Another: index variables

- create a vector of intercepts and use the value of the categorical variable to index out the right one

$$\mu = \beta_s \quad s \in \{0, 1\}$$

Requires encoding values of the variable as ordinal values $\{0, 1, 2, \dots, n\}$

This is what we did with the Africa ruggedness example:

- Africa indicator variable has values 0, 1
- Use these as an index to select the correct parameter from the vector

Code for the index variable model

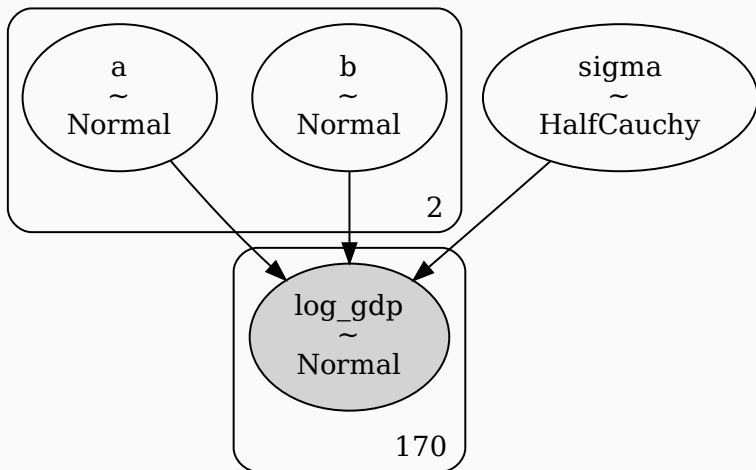
```
with pm.Model() as model_index:
    # Priors
    a = pm.Normal('a', mu=9, sd=3, shape = 2)
    b = pm.Normal('b', mu=0, sd=1, shape = 2)
    sigma = pm.HalfCauchy('sigma', 5)

    # Model equation
    mu = a[data['cont_africa']] + b[data['cont_africa']] * dd.ru

    # Likelihood
    log_gdp = pm.Normal('log_gdp', mu, sigma, observed=np.log(dd

    # Inference
    qp_index = quap()
```

Plate diagram for the index version



Comparing parameters

- Parameters from the product-interaction model:

	mean	sd	hdi_3%	hdi_97%
a[0]	9.221	0.138	8.962	9.480
a[1]	7.283	0.176	6.952	7.615
b[0]	-0.201	0.076	-0.345	-0.058
b[1]	0.186	0.105	-0.010	0.383
sigma	0.933	0.051	0.837	1.028

- Parameters from the index-variable 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

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:

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

Higher-order interactions:

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

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

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:

- Finishing up DAGs
- Criteria for variable inclusion
- Interactions

Next time:

- A bit more on interactions

HW2 up, due next Friday – first steps with PyMC3 and quap;
linear regression model; a DAG