

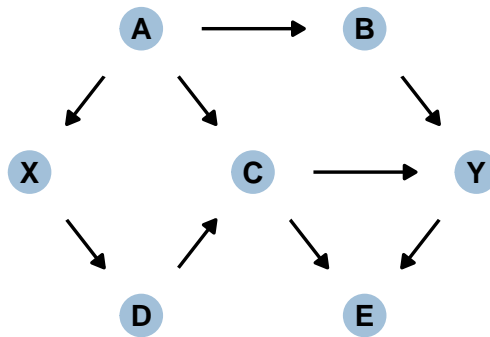
# Test 2 – In Class Problems

Stat 341 — Spring 2021

## Some Computer Output

### Causal DAGs

Researchers are investigating the total causal effect of  $X$  on  $Y$ . They propose the following DAG to represent the direct causal relationships among the 7 variables in the study.



### Kids IQ

You expect that moms with higher IQ tend to have kids with higher IQ, but you want to know whether this relationship is different depending on whether mom is a high school graduate. To check your intuition you gather some data describing some kids and their mothers. Assume your data set is called `IQ` and has three variables:

- `kid_iq`: kid's IQ
- `mom_iq`: mom's IQ
- `mom_hs`: 1 = mom is a highschool grad; 0 = mom is not a high school grad

IQ scores are normalized so that the distribution is approximately normal with a mean of 100 and standard deviation of 15, so most people have an IQ between 70 and 130. (Some consider anything over 140 to be genius level.)

The first few rows of such a data set might look like this:

<code>mom_iq</code>	<code>kid_iq</code>	<code>mom_hs</code>
103	101	0
87	89	1
114	109	0

## Pig feed

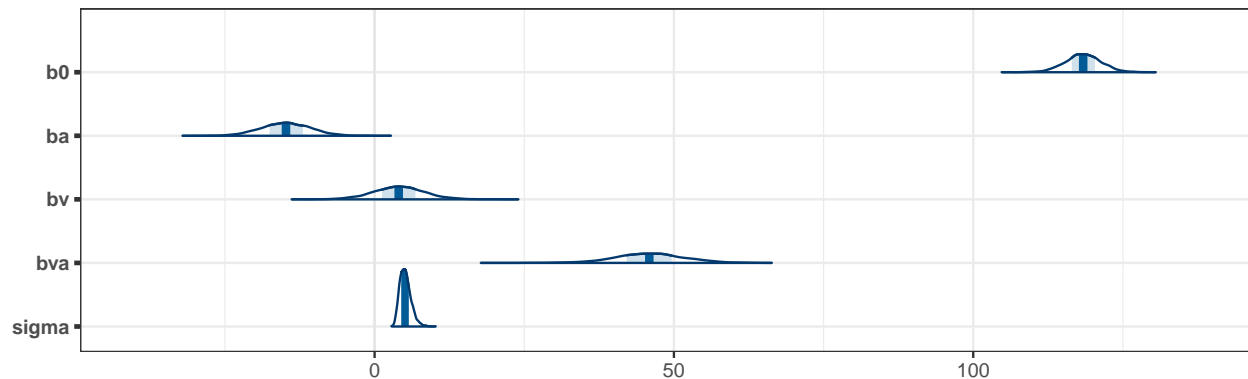
The PigFeed data set has data from an experiment on weight gain in pigs that had the goal of learning whether antibiotics, or vitamin B12, or both, might affect daily weight gain. The scientist running the experiment created four diets using a combination of antibiotics (Yes/No) and vitamin B12 (Yes/No) and randomly assigned them to some young pigs. The response variable is daily weight gain (gain, recorded in hundredths of a pound, averaged over several days). Here are the first few rows of the data:

gain	antibiotics	A	B12	V
152	Yes	1	Yes	1
105	Yes	1	No	0
126	No	0	Yes	1
130	No	0	No	0

```
pig_model <-  
  ulam(  
    data = PigFeed %>% select(gain, A, V),  
    alist(  
      gain ~ dnorm(mu, sigma),  
      mu <- b0 + ba * A + bv * V + bva * A * V,  
      b0 ~ dnorm(0, 100),  
      ba ~ dnorm(0, 30),  
      bv ~ dnorm(0, 30),  
      bva ~ dnorm(0, 30),  
      sigma ~ dexp(1),  
      iter = 2000, chains = 4, cores = 4, file = "pig-model"  
    )  
  )  
precis(pig_model)
```

```
##           mean          sd      5.5%      94.5%    n_eff    Rhat4  
## b0      118.359298  3.0292392 113.528421 123.060463 1207.730 1.001384  
## ba     -14.793100  4.1584097 -21.328042  -8.295485 1349.120 1.000431  
## bv       4.073840  4.3384372  -2.876153  10.868986 1324.790 1.001033  
## bva     45.919455  5.9084701  36.643671  55.358310 1320.981 1.000209  
## sigma   5.145979  0.9480881   3.840890   6.774251 1730.456 1.002159
```

```
mcmc_areas(stanfit(pig_model), pars = vars(-lp_))
```



## Questions

### DAGs

1. For each set of variables below, say whether they may be included (in addition to  $X$  and  $Y$  in a model that measures the total causal effect of  $X$  on  $Y$ , and **explain how you know**.

a.  $A$

b.  $A$  and  $C$

c.  $B$  and  $E$

d.  $A$  and  $E$

## **IQ**

**2.** Will you include interaction in a model that investigates this question? Why or why not? (Whichever way you choose, your answer should make it clear that you understand what it means to include interaction.)

**3.** Write down a model model (the `alist()` that you would use in `ulam()` or mathematical notation) you could use to investigate this question.

**4.** Explain your choice of priors.

## Pig Feed

Answer the following questions as well as you can give the available information.

5. For each of the four diets, determine what the `pig_model` thinks is the average weight gain.
6. Is there evidence of an interaction between antibiotics and vitamin B12 in their effects on weight gain? How strong is the evidence?
7. What would an interaction effect mean in this model?
8. Compute the residual for the first pig in the data set.
9. What does the marginal posterior for `sigma` tell us?