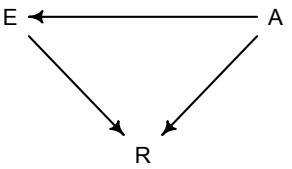
STATISTICAL RETHINKING WINTER 2020/2021
HOMEWORK, WEEK 7 SOLUTIONS

1. This is my DAG:



Age could influence both response R and education E. It could influence R, because people at different ages could have different attitudes. Age could influence education, because the longer you have lived, the more education you could have completed (up to a point). It's like the age causing marriage example from earlier in the course.

To evaluate the causal influence of E on R, we need to block the back-door from E through A to R. So we need a model that conditions on both E and A. Then the estimate for E should be the causal influence of E.

Let's set up the data:

```
library(rethinking)
data(Trolley)
d <- Trolley

# recode these in order
edu_levels <- c( 6 , 1 , 8 , 4 , 7 , 2 , 5 , 3 )
d$edu_new <- edu_levels[ d$edu ]

idx <- 1:nrow(d)
dat <- list(
    y = d$response[idx] ,
    A = d$action[idx],
    I = d$intention[idx],
    C = d$contact[idx],
    E = as.integer( d$edu_new[idx] ),
    edu_norm = normalize( d$edu_new[idx] ),
    age = standardize( d$age[idx] ),
    alpha = rep(2,7) # g prior
)
```

Note that I standardized age above. And now here is the model with both E and A. Really all we need to do is add age to the linear model. The rest you can copy from the model in the chapter.

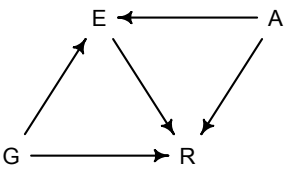
```
m1 <- ulam(
    alist(
        y ~ ordered_logistic( phi , cutpoints ),
        phi <- bE*sum( delta_shell[1:E] ) + bA*A + bC*C + BI*I + bAge*age,
        BI <- bI + bIA*A + bIC*C ,
        c(bA,bI,bC,bIA,bIC,bE,bAge) ~ normal( 0 , 0.5 ),
        cutpoints ~ normal( 0 , 1.5 ),
        vector[8]: delta_shell <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( alpha )
    ), data=dat , chains=4 , cores=4 , cmdstan=TRUE )
precis(m1)
```

```
      mean   sd  5.5% 94.5% n_eff Rhat4
bAge -0.10 0.02 -0.13 -0.07  1084  1.00
bE    0.23 0.10  0.09  0.37   373  1.02
bIC  -1.24 0.10 -1.40 -1.09  1403  1.00
bIA  -0.43 0.08 -0.56 -0.30  1154  1.00
bC   -0.34 0.07 -0.45 -0.23  1336  1.00
bI   -0.29 0.06 -0.38 -0.20  1080  1.00
bA   -0.48 0.05 -0.56 -0.39  1054  1.00
```

You may recall from the chapter that education has a negative effect in the model without age. Now that we include age, education has a positive influence (with some overlap with zero). So age has indeed soaked up some of the previous influence assigned to education. The back-door may be real.

I'd summarize this model, assuming this DAG is true, as saying that age causes people to give slightly lower responses. This could be a cohort effect, and not a causal influence of age. Either way, it is small. Education seems to cause higher responses (more approval). This suggests that education trains people to see some or all of the features A,I,C as more permissible. A model that interacted education with each might shed more light on things. Remember: A DAG doesn't say whether you need an interaction effect or not. That is a separate problem.

2. This is my DAG:



The only new part is the fork coming from gender G. It influences both R and E. If this DAG is correct, then is E confounded? Yes. Let's see the minimum adjustment set necessary to measure the causal influence of E:

```
library(dagitty)
dag2 <- dagitty("dag{
    E -> R <- A
    A -> E
    G -> E
    G -> R
}")
adjustmentSets( dag2 , exposure="E" , outcome="R" , effect="total" )
```

```
 { A, G }
```

This means we'd need to condition on both A and G to get an un-confounded estimate for E. Why? There is a back-door from E to G to R. It is just like the back-door through A.

Here's the model we need, which includes education, age, and gender (female dummy variable):

```
dat$female <- ifelse( d$male==1 , 0L , 1L )
m2 <- ulam(
    alist(
        y ~ ordered_logistic( phi , cutpoints ),
        phi <- bE*sum( delta_shell[1:E] ) + bA*A + bC*C + BI*I +
               bAge*age + bF*female,
        BI <- bI + bIA*A + bIC*C ,
        c(bA,bI,bC,bIA,bIC,bE,bAge,bF) ~ normal( 0 , 0.5 ),
        cutpoints ~ normal( 0 , 1.5 ),
        vector[8]: delta_shell <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( alpha )
    ), data=dat , chains=4 , cores=4 , cmdstan=TRUE )
precis(m2)
```

```
      mean   sd  5.5% 94.5% n_eff Rhat4
bF   -0.56 0.03 -0.62 -0.51  2520     1
bAge -0.07 0.02 -0.11 -0.03  1132     1
bE    0.03 0.16 -0.24  0.26   635     1
bIC  -1.26 0.09 -1.41 -1.11  1201     1
bIA  -0.44 0.08 -0.56 -0.31   909     1
bC   -0.35 0.07 -0.46 -0.24  1291     1
bI   -0.29 0.06 -0.38 -0.20   961     1
bA   -0.48 0.05 -0.56 -0.39  1150     1
```

Age is still negative (and weak), while education is right near zero and straddles both sides. Gender seems to have accounted for all of the previous influenced assigned to education. It looks like female respondents gave lower average responses—indicating less approval.

It would be worth figuring out how gender is associated with education in this sample. It could be true for example that some education levels under-sampled men

or women, and this leads to another kind of confound. Consider for example if older men are less likely to respond, so the sample becomes increasingly female with age. Then education level will also be increasingly female with age. Since the sample is not a representative sample of the population, there are probably some biases of this sort.