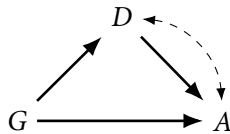**STATISTICAL RETHINKING 2023**
WEEK 6 SOLUTIONS

1. The implied DAG is:

where G is gender, D is discipline, and A is award. The dashed curve between D and A represents possible confounds. The direct causal effect of gender is the path $G \rightarrow A$. The total effect of G includes that path and the indirect path $G \rightarrow D \rightarrow A$. We can estimate the total causal influence (assuming this DAG is correct) with a model that conditions only on gender. I'll use a N(-1,1) prior for the intercepts, because we know from domain knowledge that less than half of applicants get awards.

```
library(rethinking)
data(NWOGrants)
d <- NWOGrants
dat <- list(
    A = as.integer(d$awards),
    N = as.integer(d$applications),
    G = ifelse( d$gender=="f" , 1L , 2L ) ,
    D = as.integer(d$discipline)
)

# for total effect, just G, no D
m1 <- ulam(
    alist(
        A ~ binomial( N , p ),
        logit(p) <- a[G],
        a[G] ~ normal(-1,1)
    ), data=dat , chains=4 , cores=4 )

precis(m1,2)
```

```
      mean   sd  5.5% 94.5% n_eff Rhat4
a[1] -1.74 0.08 -1.87 -1.61  1498     1
a[2] -1.53 0.06 -1.63 -1.43  1453     1
```
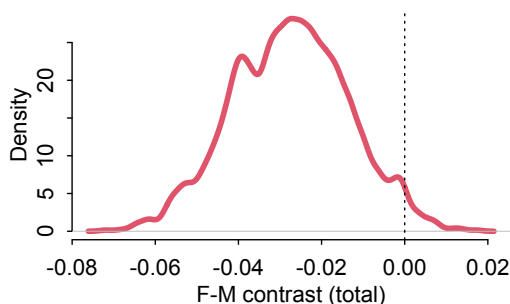
Gender 1 here is female and 2 is male. So males have higher rates of award, on average. How big is the difference? Let's look at the contrast on probability scale:

```
post1 <- extract.samples(m1)
post1$pG1 <- inv_logit( post1$a[,1] )
post1$pG2 <- inv_logit( post1$a[,2] )
post1$G_contrast <- post1$pG1 - post1$pG2

dens( post1$G_contrast , lwd=4 , col=2 , xlab="F-M contrast (total)" )
abline( v=0 , lty=3 )
```


So a 3% difference on average. With such low funding rates (in some disciplines), 3% is a big advantage.

2. Now for the direct influence of gender, we condition on discipline as well:

```
m2 <- ulam(
    alist(
        A ~ binomial( N , p ),
        logit(p) <- a[G,D],
        matrix[G,D]:a ~ normal(-1,1)
    ), data=dat , chains=4 , cores=4 )
precis(m2,3)
```

```
         mean   sd   5.5% 94.5% n_eff Rhat4
a[1,1] -1.07 0.36 -1.65 -0.50  2684    1
a[1,2] -1.76 0.24 -2.15 -1.39  2953    1
a[1,3] -1.43 0.19 -1.75 -1.13  2577    1
a[1,4] -1.28 0.26 -1.70 -0.86  4065    1
a[1,5] -2.06 0.19 -2.37 -1.76  2461    1
a[1,6] -1.20 0.37 -1.77 -0.63  3287    1
a[1,7] -1.24 0.61 -2.24 -0.29  2664    1
a[1,8] -2.02 0.15 -2.26 -1.80  3041    1
a[1,9] -1.32 0.30 -1.79 -0.85  2845    1
a[2,1] -1.03 0.25 -1.43 -0.66  2665    1
a[2,2] -1.14 0.19 -1.44 -0.85  2504    1
a[2,3] -1.77 0.19 -2.07 -1.48  3133    1
a[2,4] -1.99 0.29 -2.49 -1.54  2576    1
a[2,5] -1.46 0.16 -1.72 -1.21  3115    1
```

```
a[2,6] -1.43 0.21 -1.77 -1.10  3079      1
a[2,7] -1.02 0.27 -1.45 -0.62  2521      1
a[2,8] -1.70 0.13 -1.92 -1.49  2798      1
a[2,9] -1.66 0.19 -1.98 -1.35  2544      1
```

It isn't possible to make any sense out of this table. But we can compute the contrast in each discipline to see what is going on:

```
# show contrasts for each discipline
plot( NULL , xlim=c(-0.4,0.4) , ylim=c(0,18) ,
    xlab="F-M contrast for each discipline" , ylab="Density" )
abline( v=0 , lty=3 )
dat$disc <- as.character(d$discipline)
disc <- dat$disc[order(dat$D)]
for ( i in 1:9 ) {
    pG1Di <- link(m2,data=list(D=i,N=1,G=1))
    pG2Di <- link(m2,data=list(D=i,N=1,G=2))
    Gcont <- pG1Di - pG2Di
    dens( Gcont , add=TRUE , lwd=3 , col=i )
    xloc <- ifelse( mean(Gcont) < 0 , -0.35 , 0.35 )
    xpos <- ifelse( mean(Gcont) < 0 , 4 , 2 )
    text( xloc + 0.5*mean(Gcont) , 18-i , disc[2*i] , col=i ,
        pos=xpos , font=2 )
}
```
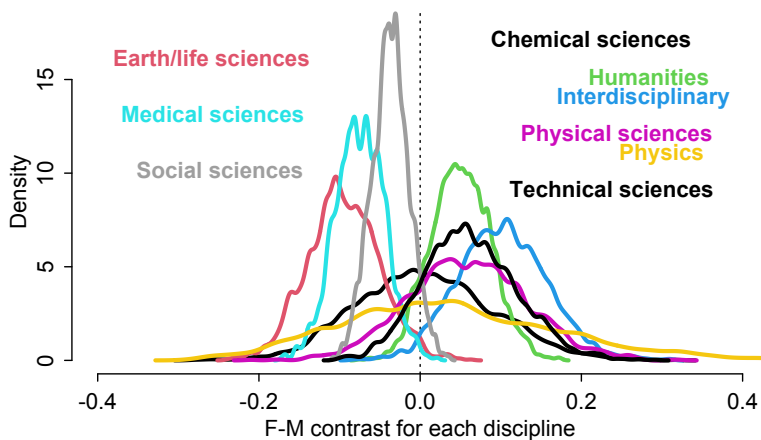

Not the nicest looking figure ever. But it shows the variation across disciplines, with some showing higher rates for women (right) and others for men (left).

Now to compute the average direct effect of gender, imagining an intervention that changes perception of G on each grant application but does not alter D. First let's count up all of the applications and count them also by each discipline.

```
total_apps <- sum(dat$N)
apps_per_disc <- sapply( 1:9 , function(i) sum(dat$N[dat$D==i]) )
```

Next we simulate the same number of applications, to the same disciplines, but changing G to 1 in each case:
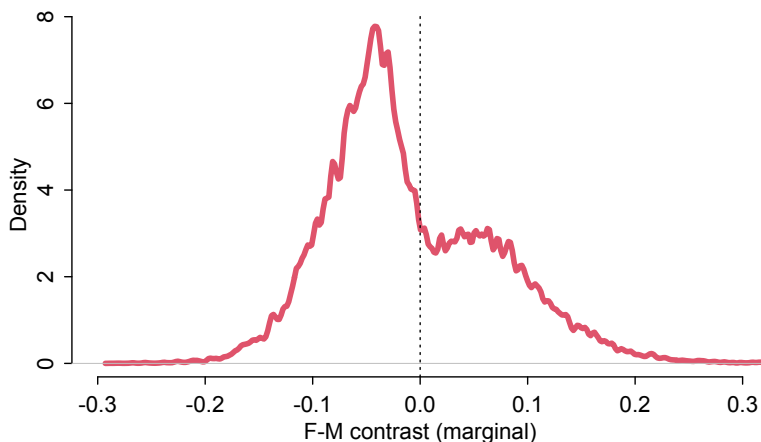
```
pG1 <- link(m2,data=list(
    D=rep(1:9,times=apps_per_disc),
    N=rep(1,total_apps),
    G=rep(1,total_apps)))
```

And the same but with $G = 2$ for all applications:

```
pG2 <- link(m2,data=list(
    D=rep(1:9,times=apps_per_disc),
    N=rep(1,total_apps),
    G=rep(2,total_apps)))
```

We plot the contrast distribution:

```
dens( pG1 - pG2 , lwd=4 , col=2 , xlab="F-M contrast (marginal)" ,
    xlim=c(-0.3,0.3) )
abline( v=0 , lty=3 )
```


The expected direct effect varies by discipline, and so when we post-stratify by discipline, the distribution of expected effects is broad. It is important to note that this density is not the same as the density of the total causal effect of G, but the total effect includes any causal effect on G on D. When we remove that (statistically), the result is that the advantages are rather balanced, with the mean of the density above being almost zero and the 89% interval about -0.12 to 0.12.

3. How can the total effect show a disadvantage to women and the direct a balanced disadvantage to both men and women, depending upon discipline? It's partly the influence of G on D, according to the model at least. Women tend to apply slightly more to departments with lower award rates. It'll help to make a plot:

```
total_f <- sum(d$applications[d$gender=="f"])
pDf <- d$applications[d$gender=="f"] / total_f

total_m <- sum(d$applications[d$gender=="m"])
pDm <- d$applications[d$gender=="m"] / total_m

# overall award rate in each discipline
n_apps <- xtabs( dat$N ~ dat$D )
n_awards <- xtabs( dat$A ~ dat$D )
p_award <- n_awards / n_apps

# f/m award rate in each discipline
post <- extract.samples(m2)
pF <- apply( inv_logit(post$a[,1,]) , 2 , mean )
pM <- apply( inv_logit(post$a[,2,]) , 2 , mean )

plot( pDf , pDm , lwd=3 , col=ifelse(pDf>pDm,2,4) , pch=ifelse(pF>pM,"F","M") )
abline(a=0,b=1, lty=3)
identify( pDf , pDm , round(p_award,2) )
```
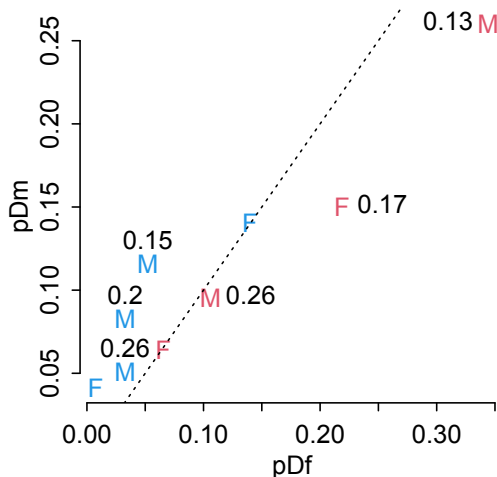

The horizontal axis is the proportion of women applying to each discipline and the vertical the proportion of men applying to each discipline. I've colored the points red when the proportion of women exceeds the proportion of men. Note that this isn't the proportions of men and women in each discipline. More men than women applied to almost all disciplines, in this sample. Rather its within each gender, what

proportions applied to each discipline. So relatively larger proportion of women applied to the disciplines below the diagonal line. The points are named for which gender has a higher award rate *in that discipline*. And the numbers are the overall award rates in each discipline. It is the award rates in the popular disciplines that influence the total effect of gender most. And those disciplines have lower award rates (0.13 and 0.17). And a greater proportion of women apply to those departments. And in the most popular one (upper right) women have a lower award rate even than men. In contrast, the disciplines that more ben apply to (blue) tend to have higher award rates and favor men in those awards.

So both the pattern of who applies where seems to matter, as does the the imbalance in specific disciplines. So these is nuanced story to tell here.

Of course we haven't accounted for confounds. Those could be contributing as well, by masking different award rates within disciplines or by contributing to the pattern of who applies to which discipline. The explanation above isn't really a causal one, but it just reconciles the total estimate with the direct estimates.

4-OPTIONAL CHALLENGE. This problem is two problems in one. The first problem is to figure out a way to use the data to estimate the possible influence of left-handedness on winning. The second is to explain the estimate.

This kind of estimation problem requires a model that is invariant to which fighter is listed as number 1 or number 2. What we want to do is to model the probability that fighter 1 wins any given match (row in the data), but the model should look the same if we switch the fighter labels. The basic idea is that the log-odds that fighter 1 wins is given by:

$$\text{logit}(p_i) = \alpha_{F1[i]} - \alpha_{F2[i]}$$

where the α parameters are unobserved fighter ability scores. We can modify this to account for lefties:

$$\text{logit}(p_i) = (\alpha_{F1[i]} + \beta L_{F1[i]}) - (\alpha_{F2[i]} + \beta L_{F2[i]})$$

where $L_{F1[i]}$ is an indiciator (0/1) of whether fighter 1 is a lefty. So in the event both fighters 1 and 2 are lefties, the advantage cancels. But if only one is a lefty, then that fighter gets a bonus of β on the log-odds scale.

Let's try this approach with the UFC data. I'll start with a model that omits the α parameters. We can talk about those further down.
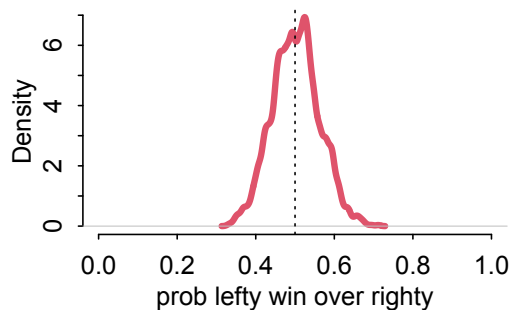
```
library(rethinking)
data(UFClefties)
d <- UFClefties
dat <- list(
    Y = d$fighter1.win,
    L1 = d$fighter1.lefty,
    L2 = d$fighter2.lefty
)
m4a <- ulam(
    alist(
```

```
        Y ~ bernoulli(p),
        logit(p) <- b*(L1 - L2),
        b ~ normal(0,1)
    ) , data=dat , chains=4 , cores=4  )

post <- extract.samples(m4a)
dens( inv_logit(post$b) , xlab="percent lefty advantage" ,
    lwd=4 , col=2 , xlim=c(0,1) )
abline(v=0.5,lty=3)
```


Doesn't seem to be any evidence here of a lefty advantage.

If there isn't any lefty advantage, then why are lefties over-represented among fighters (and tennis players and fencers and etc etc)? The qualitative evidence of lefty advantage is also compelling: professional fighters (and tennis players etc) know what it is like to face off against a lefty. But maybe it's just a persistent myth? If so, then why are lefties more common in such sports?

No one has the answers. But one idea is that by the time we examine professional matches, the sample is biased by...(drum roll)...collider bias! Suppose being a lefty really is an advantage on average. Suppose also that there are other things that matter, and we'll just call those things "ability". So both lefty L and ability A help a fighter win matches. There is no confound there, just competing causes. But the matches in the sample are not beginners. These fighters went through qualifying matches. So suppose a causal structure like this:
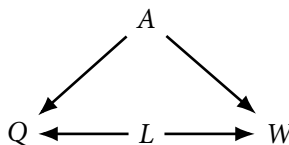
where L is lefty, A is ability otherwise, Q is whether someone qualifies for UFC matches and W is a win in our sample. Since the sampled fights are among qualified fighters, the sample is conditioned on a collider. This induces a non-causal association between A and L, such that lefties are on average lower in ability than righties. Think of it this way: A lefty can qualify with a lower A, because of the lefty

advantage. And a righty must be better than an average lefty to stay in the competition. So after several rounds of qualifying, the lefty advantage can be reduced or erased by the selection.

Here's a little simulation to drive the point home. I'll simulate 5000 fighters in which there is no association between lefties and ability. Then I'll pull out the upper tail of ability to qualify, but I'll let lefties qualify with lower ability.

```
# collider explanation example
N <- 5000
L <- rbern(N,0.1)
A <- rnorm(N)
Q <- rep(1,N)

# qualify if A large enough or Lefty
Q <- ifelse( A > 2 | (A > 1.25 & L==1) , 1 , 0 )

# summarize
table(Q,L)
sum(Q==1 & L==1)/sum(Q==1) # prop lefties
mean( A[L==1 & Q==1] )
mean( A[L==0 & Q==1] )
```

```
> table(Q,L)
   L
Q      0    1
  0 4360  483
  1  110   47
> sum(Q==1 & L==1)/sum(Q==1) # prop lefties
[1] 0.2993631
> mean( A[L==1 & Q==1] )
[1] 1.686975
> mean( A[L==0 & Q==1] )
[1] 2.375568
```

So lefties are only 10% of the simulation 5000, but they are 30% of those qualifying. Now let's simulate matches among those qualifying:

```
# now sim fights among those qualifying
k <- 2 # importance of ability differences
b <- 0.5 # lefty advantage
l <- L[Q==1]
a <- A[Q==1]
M <- sum(Q==1)
W <- rep(NA,M/2) # matches
id1 <- rep(NA,M/2)
id2 <- rep(NA,M/2)
L1 <- id1
L2 <- L1
```

```
for ( i in 1:(M/2) ) {
    a1 <- a[i] + b*l[i]
    a2 <- a[M/2+i] + b*l[M/2+i]
    p <- inv_logit( k*(a1 - a2) )
    f1win <- rbern(1,p)
    W[i] <- f1win
    id1[i] <- i
    id2[i] <- M/2+i
    L1[i] <- l[i]
    L2[i] <- l[M/2+i]
}
```
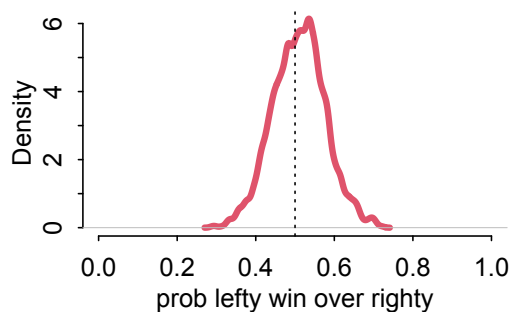
And we can analyze the simulated data to see what the model finds:

```
datx <- list( Y=W , id1=id1, id2=id2 , L1=L1 , L2=L2 , nid=max(id2) )
m4x <- ulam(
    alist(
        Y ~ bernoulli(p),
        logit(p) <- b*(L1 - L2),
        b ~ normal(0,0.5)
    ) , data=datx , chains=4 , cores=4  )
post <- extract.samples(m4x)
dens( inv_logit(post$b) , xlab="prob lefty win over righty" ,
    lwd=4 , col=2 , xlim=c(0,1) )
abline(v=0.5,lty=3)
```


We know there is a lefty advantage here, because we simulated the data with one. But it doesn't show up here, because the sample is conditioned on a collider (qualifying). If you make the lefty advantage bigger, you can still see it. Or if you make qualifying less strict, you can still see the lefty advantage. So it's not clear if this is really a good explanation for the failure to find lefty advantage in the UFC sample. But it's possible.

Another idea is that there is no lefty advantage at the pro level, because pros train for lefties. So being a lefty helps one to become pro, but it doesn't help once someone is pro.

Okay if you go back a few pages and look at the DAG again, you'll see that if we could condition on ability A, we could deal with the sampling bias. How do we get a measure of A? In these data, we can't really, because there aren't enough repeat observations of each fighter. But in a more substantial sample, it might be possible. We'd need a multilevel model though. And that's in later homework.