

Assignment 2 - Methods 4

Study Group 20

2025-04-03

Second assignment

The second assignment uses chapter 3, 5 and 6. The focus of the assignment is getting an understanding of causality.

Chapter 3: Causal Confussion

Reminder: We are trying to estimate the probability of giving birth to a boy I have pasted a working solution to questions 6.1-6.3 so you can continue from here:)

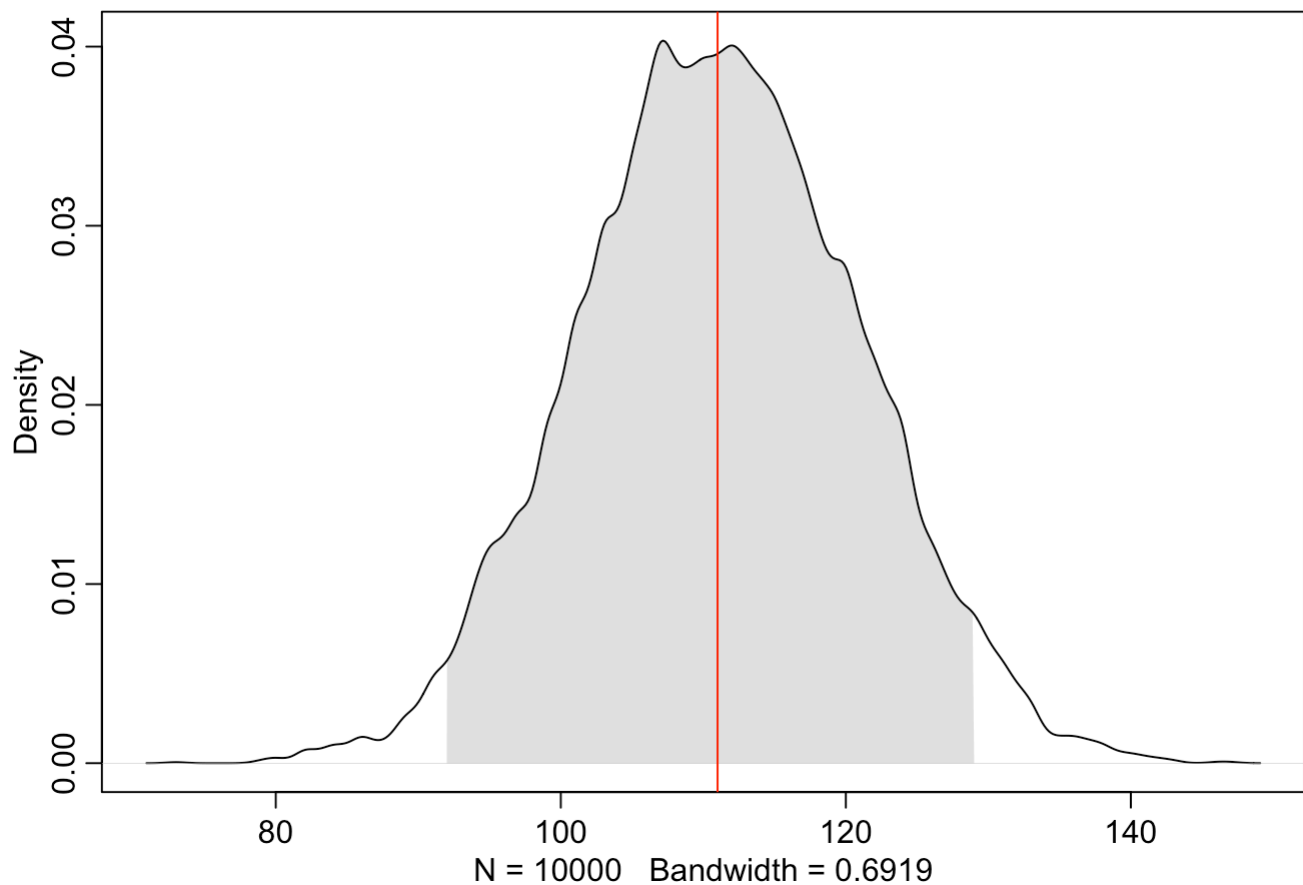
3H3 Use `rbinom` to simulate 10,000 replicates of 200 births. You should end up with 10,000 numbers, each one a count of boys out of 200 births. Compare the distribution of predicted numbers of boys to the actual count in the data (111 boys out of 200 births).

```
# 3H1
# Find the posterior probability of giving birth to a boy:
pacman::p_load(rethinking)
data(homeworkch3)
set.seed(1)
W <- sum(birth1) + sum(birth2)
N <- length(birth1) + length(birth2)
p_grid <- seq(from = 0, to = 1, len = 1000)
prob_p <- rep(1, 1000)
prob_data <- dbinom(W, N, prob = p_grid)
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)

# 3H2
# Sample probabilities from posterior distribution:
samples <- sample(p_grid, prob = posterior, size = 1e4, replace = TRUE)

# 3H3
# Simulate births using sampled probabilities as simulation input, and check if they
# align with real value.
simulated_births <- rbinom(n = 1e4, size = N, prob = samples)
rethinking::dens(simulated_births, show.HPDI = 0.95)
abline(v=W, col="red")
title("Simulated amount of boys in 200 births - red line is real value")
```

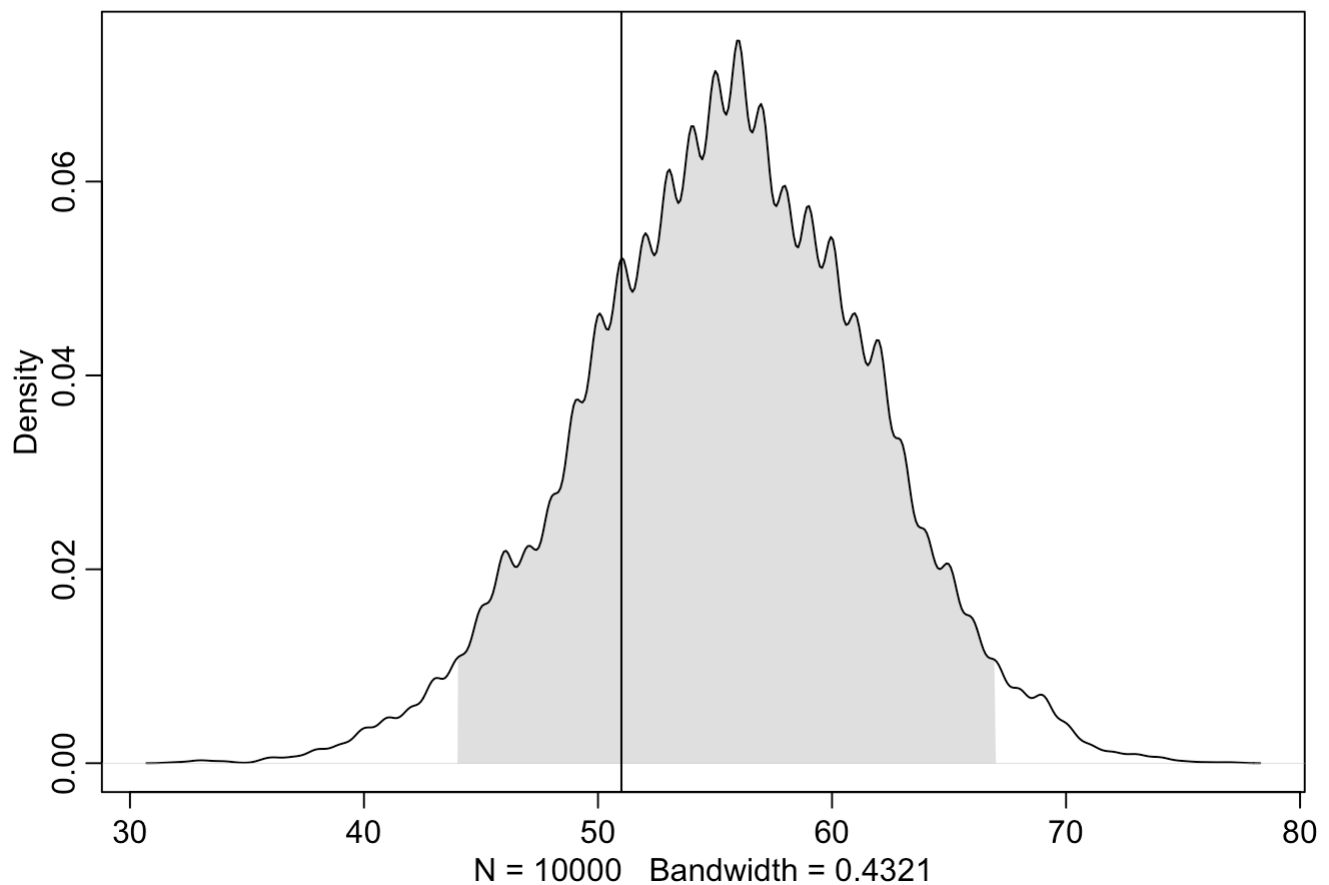
Simulated amount of boys in 200 births - red line is real value



3H4. Now compare 10,000 counts of boys from 100 simulated first borns only to the number of boys in the first births, birth1. How does the model look in this light?

by Marie Storm Rasmussen

```
sim_first <- rbinom(1e4, size = length(birth1), prob = samples)
rethinking::dens(sim_first, show.HPDI = 0.95)
abline(v=sum(birth1))
```



Answer 3H4) It seems that our model overestimates the number of boys being the first born, compared to the observed data.

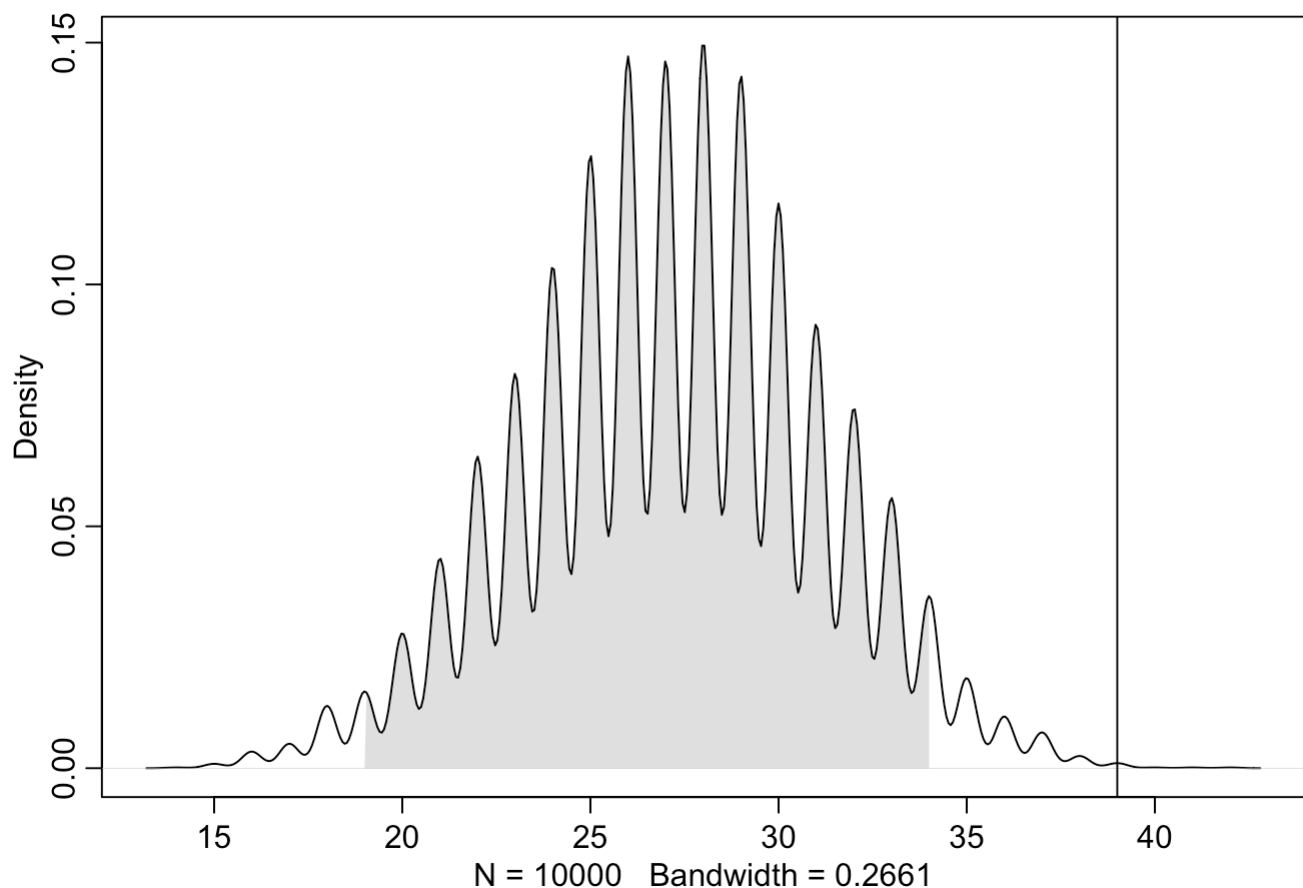
3H5. The model assumes that sex of first and second births are independent. To check this assumption, focus now on second births that followed female first borns. Compare 10,000 simulated counts of boys to only those second births that followed girls. To do this correctly, you need to count the number of first borns who were girls and simulate that many births, 10,000 times. Compare the counts of boys in your simulations to the actual observed count of boys following girls. How does the model look in this light? Any guesses what is going on in these data?

by Marie Storm Rasmussen

```
births <- birth2[birth1 == 0]
firstborn_g <- length(births)

sim_first_g <- rbinom(1e4, size=firstborn_g, prob=samples)
rethinking::dens(sim_first_g, show.HPDI = 0.95)

abline(v = sum(births))
```



Answer 3H5) Our model seems to underfit the data which implies that births are not independent of one another.

Chapter 5: Spurious Correlations

Start of by checking out all the spurious correlations that exists in the world. Some of these can be seen on this wonderfull website: <https://www.tylervigen.com/spurious/random>

(<https://www.tylervigen.com/spurious/random>) All the medium questions are only asking you to explain a solution with words, but feel free to simulate the data and prove the concepts.

5M1. Invent your own example of a spurious correlation. An outcome variable should be correlated with both predictor variables. But when both predictors are entered in the same model, the correlation between the outcome and one of the predictors should mostly vanish (or at least be greatly reduced).

By Matilde Elene Hansen

Answer 5M1) A spurious correlation is a correlation between two variables that appears to exist but is actually caused by one or more unmeasured variables (confounders) or coincidental patterns, rather than a genuine causal relationship. A homebrewed example of this could be predicting happiness from time spent in nature and top 100 IMDB movies watched within a set frame of time. We might see a high correlation between IMDB movies watched and happiness, but perhaps introducing time in nature will diminish this first correlation, implying that time in nature is the primary force of happiness. In other words, the initial correlation between IMDB movies watched and happiness was spurious, because it vanished once the true influencing variable (time spent in nature) was considered.

5M2. Invent your own example of a masked relationship. An outcome variable should be correlated with both predictor variables, but in opposite directions. And the two predictor variables should be correlated with one another.

By Matilde Elene Hansen

Answer 5M2) A masked relationship appears when there are two predictor variables that are correlated with one another, but one of these is positively correlated with the outcome and the other is negatively correlated with it, possibly leading to the variables canceling each other out. An example of this could be the outcome variable "Grades" or some variable measuring academic performance, with the two predictors: time spent studying and stress levels. More time spent studying typically increases academic performance (a positive relationship), but increased studying can also lead to higher stress levels, which in turn negatively affects academic performance (a negative relationship). Since study time and stress levels are positively correlated, their opposing influences on grades can mask each other's true effects in a regression analysis.

5M3. It is sometimes observed that the best predictor of fire risk is the presence of firefighters— States and localities with many firefighters also have more fires. Presumably firefighters do not cause fires. Nevertheless, this is not a spurious correlation. Instead fires cause firefighters. Consider the same reversal of causal inference in the context of the divorce and marriage data. How might a high divorce rate cause a higher marriage rate? Can you think of a way to evaluate this relationship, using multiple regression

By Frederick Richard Rose

Answer 5M3) More people getting divorced leave more people on the market for re-marriage. Adding another predictor might shed light on the causal order of the relationship. But again, more people getting married also increase the number of people getting divorced (assuming the proportion of divorces staying the same). IN order to asses the direction of this relationship, another predictor such as "number of remarriages" could be included.

5M5. One way to reason through multiple causation hypotheses is to imagine detailed mechanisms through which predictor variables may influence outcomes. For example, it is sometimes argued that the price of gasoline (predictor variable) is positively associated with lower obesity rates (outcome variable). However, there are at least two important mechanisms by which the price of gas could reduce obesity. First, it could lead to less driving and therefore more exercise. Second, it could lead to less driving, which leads to less eating out, which leads to less consumption of huge restaurant meals. Can you outline one or more multiple regressions that address these two mechanisms? Assume you can have any predictor data you need.

By Matilde Elene Hansen

Answer 5M5) For the first mechanism we would need a variable corresponding to the time spent exercising, and for the second mechanism we would need a variable corresponding to how often people eat out. This could be expressed in a model as:

$$\mu_i = \alpha + \beta_G G_i + \beta_E E_i + \beta_R R_i$$

where G represents the price of gasoline, E represents an exercise-related variable, and R represents a restaurant related variable (as described above.)

However, one could also go even deeper into the research question by including less subjective ratings and instead measure calories going in (eating out) and amount of calories going out (exercise).

Chapter 5: Foxes and Pack Sizes

All five exercises below use the same data, data(foxes) (part of rethinking).⁸⁴ The urban fox (*Vulpes vulpes*) is a successful exploiter of human habitat. Since urban foxes move in packs and defend territories, data on habitat quality and population density is also included. The data frame has five columns: (1) group: Number of the social group the individual fox belongs to (2) avgfood: The average amount of food available in the territory (3) groupsize: The number of foxes in the social group (4) area: Size of the territory (5) weight: Body weight of the individual fox

5H1. Fit two bivariate Gaussian regressions, using quap: (1) body weight as a linear function of territory size (area), and (2) body weight as a linear function of groupsize. Plot the results of these regressions, displaying the MAP regression line and the 95% interval of the mean. Is either variable important for predicting fox body weight?

By Matilde Elene Hansen

```
data(foxes)

# 1.

m1 <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b * area,
  a ~ dnorm(5, 5),
  b ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m1 <- quap(m1, data = foxes)

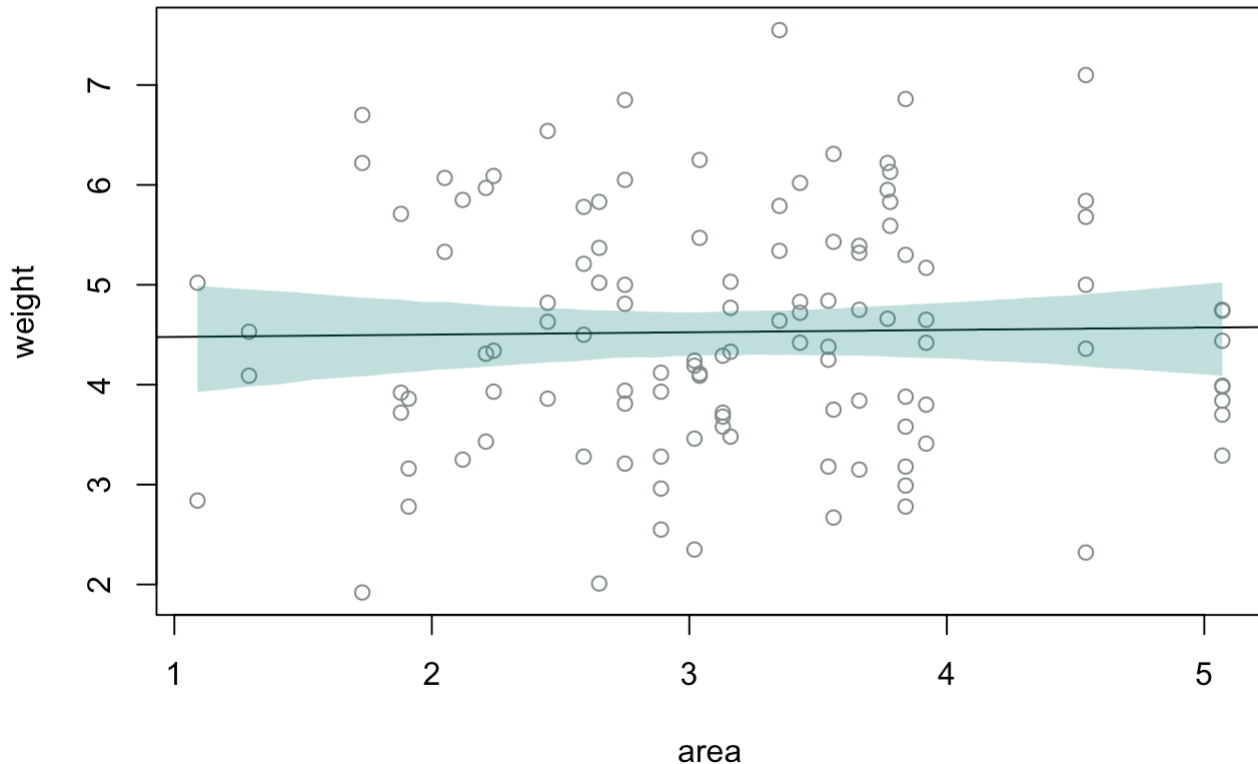
precis(quap_m1)
```

```
##           mean      sd      5.5%    94.5%
## a      4.45430364 0.3895545  3.8317203 5.076887
## b      0.02385729 0.1180300 -0.1647774 0.212492
## sigma 1.17867596 0.0773828  1.0550033 1.302349
```

```
area_sequence <- seq(from = min(foxes$area), to = max(foxes$area), length.out = 1e4)
mu <- link(quap_m1, data = data.frame(area = area_sequence))
mu.PI <- apply(mu, 2, PI, prob = 0.95)
plot(weight ~ area, data = foxes, col = "azure4")
abline(quap_m1)
```

```
## Warning in abline(quap_m1): only using the first two of 3 regression
## coefficients
```

```
shade(mu.PI, area_sequence, col = col.alpha("cyan4", alpha = 0.3))
```



2.

```
m2 <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b * groupsize,
  a ~ dnorm(5, 5),
  b ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m2 <- quap(m2, data = foxes)

precis(quap_m2)
```

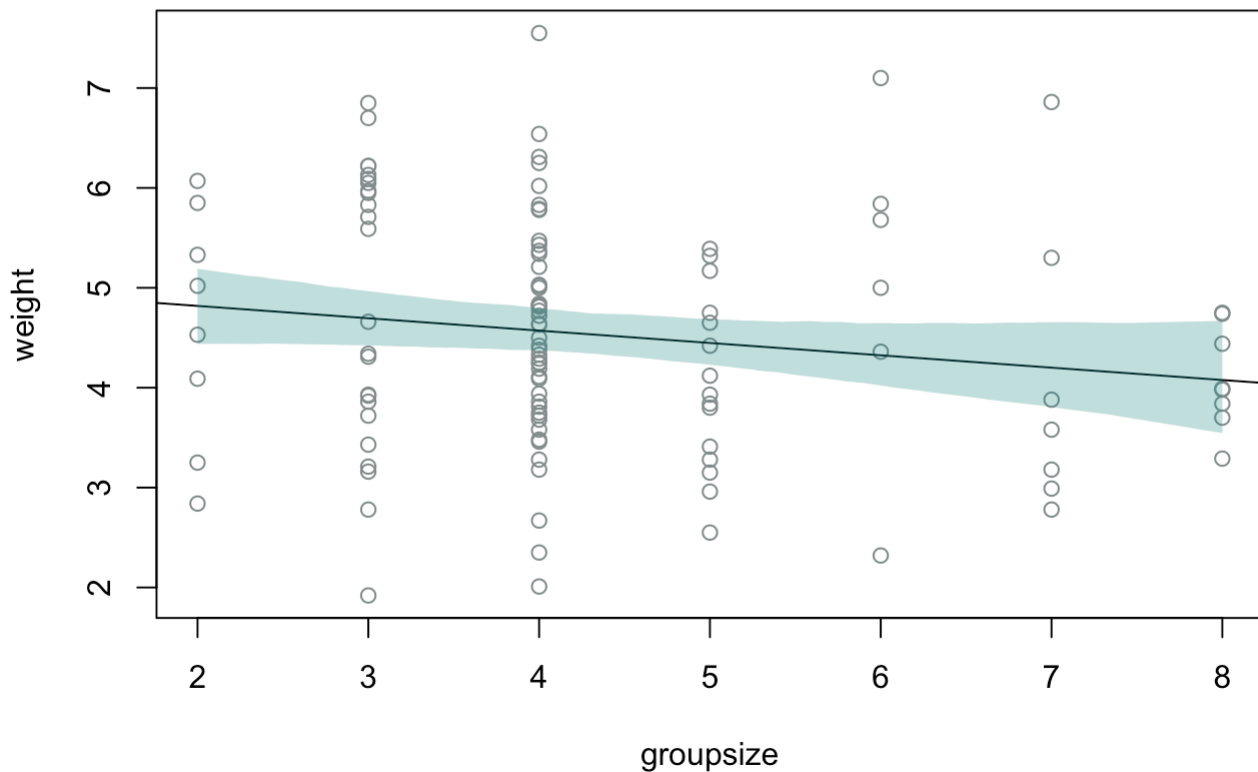
##	mean	sd	5.5%	94.5%
## a	5.0675836	0.32418281	4.5494769	5.58569038
## b	-0.1238161	0.07038361	-0.2363027	-0.01132955
## sigma	1.1635303	0.07638931	1.0414454	1.28561515

```
groupsize_sequence <- seq(from = min(foxes$groupsize), to = max(foxes$groupsize), length.out = 1e4)
mu2 <- link(quap_m2, data = data.frame(groupsize = groupsize_sequence))
mu.PI2 <- apply(mu2, 2, PI, prob = 0.95)
plot(weight ~ groupsize, data = foxes, col = "azure4")
abline(quap_m2)
```



```
## Warning in abline(quap_m2): only using the first two of 3 regression
## coefficients
```

```
shade(mu.PI2, groupsize_sequence, col = col.alpha("cyan4", alpha = 0.3))
```



Answer 5H1) Based on these two bivariate regressions, neither territory area nor group size seems to be very important for the prediction of body weight. Group size might have a slight negative relationship with the weight, however it would be wise to do further analyses to see if other variables influence this as well.

5H2. Now fit a multiple linear regression with weight as the outcome and both area and groupsize as predictor variables. Plot the predictions of the model for each predictor, holding the other predictor constant at its mean. What does this model say about the importance of each variable? Why do you get different results than you got in the exercise just above?

By Frederick Richard Rose

```
m3 <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b1 * area + b2 * groupsize,
  a ~ dnorm(5, 5),
  b1 ~ dnorm(0, 5),
  b2 ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))
```

```
quap_m3 <- quap(m3, data = foxes)
```

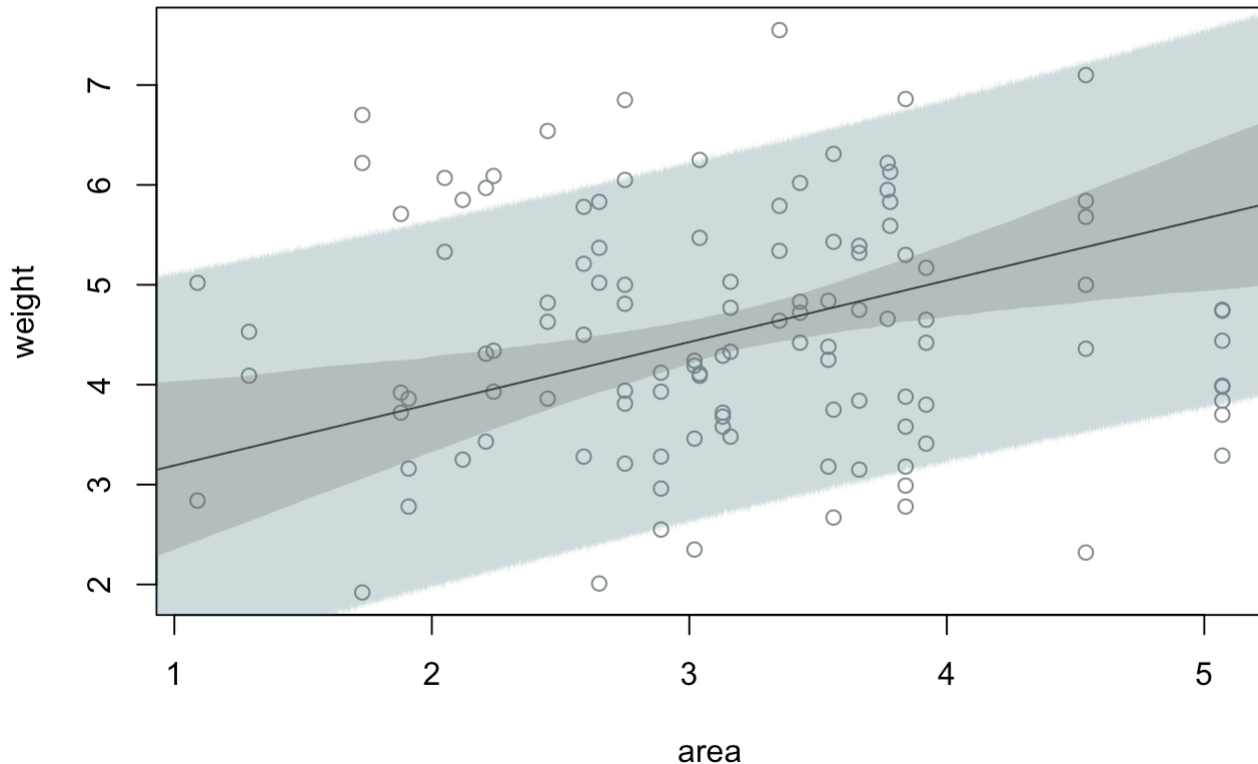
```
precis(quap_m3)
```

##	mean	sd	5.5%	94.5%
## a	4.4541029	0.36977799	3.8631263	5.0450795
## b1	0.6160146	0.19978707	0.2967162	0.9353129
## b2	-0.4318809	0.12066886	-0.6247330	-0.2390288
## sigma	1.1184710	0.07343303	1.0011108	1.2358312

```
groupsize_mean <- mean(foxes$groupsize)
area_sequence2 <- seq(from = 0, to = 6, length.out = 1e4)
predicted_data <- data.frame(groupsize = groupsize_mean, area = area_sequence2)
```

```
mu3 <- link(quap_m3, data = predicted_data)
mu_mean <- apply(mu3, 2, mean)
mu_PI <- apply(mu3, 2, PI, prob = 0.95)
a_sim <- sim(quap_m3, data = predicted_data, n = 1e4)
a_PI <- apply(a_sim, 2, PI)
```

```
plot(weight ~ area, data = foxes, col = "azure4")
lines(area_sequence2, mu_mean)
shade(mu_PI, area_sequence2, col = col.alpha("lightpink4", alpha = 0.3))
shade(a_PI, area_sequence2, col = col.alpha("cadetblue4", alpha = 0.3))
```



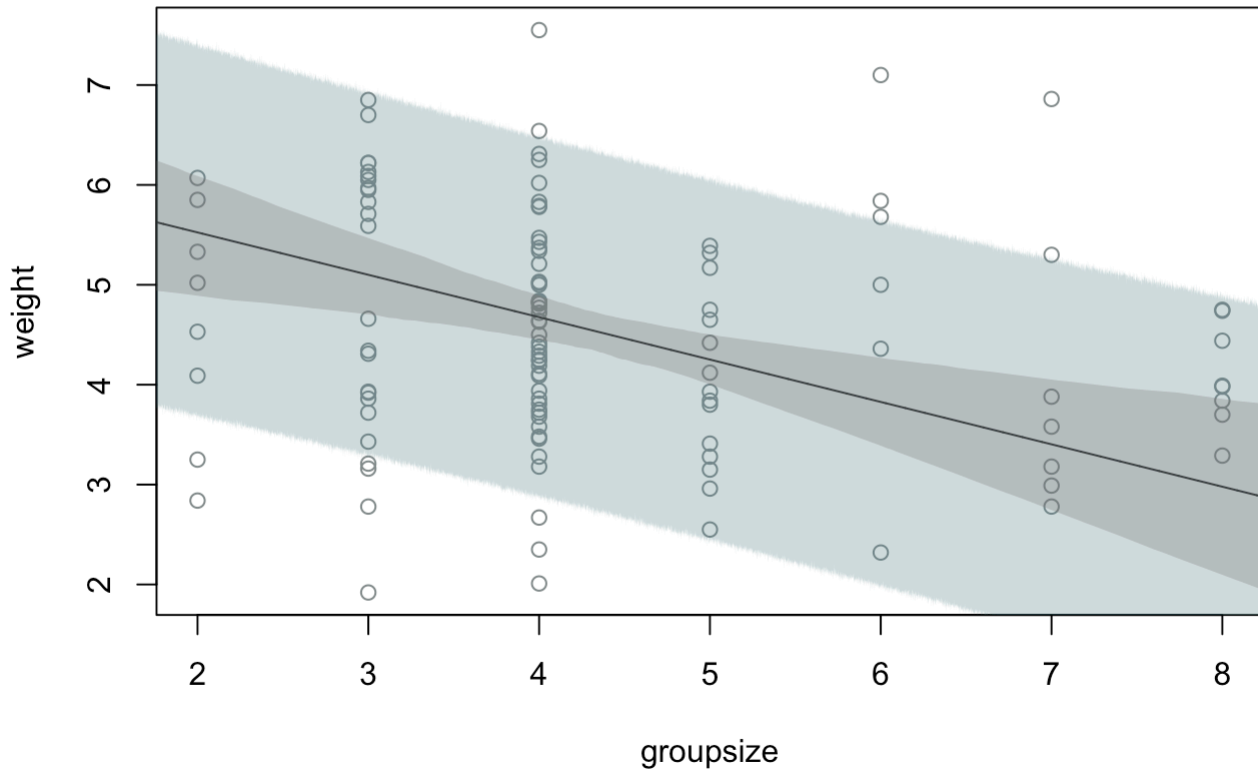
```

area_mean <- mean(foxes$area)
groupsize_sequence2 <- seq(from = 1, to = 10, length = 1e4)
predicted_data2 <- data.frame(groupsize = groupsize_sequence2, area = area_mean)

mu4 <- link(quap_m3, data = predicted_data2)
mu_mean2 <- apply(mu4, 2, mean)
mu_PI2 <- apply(mu4, 2, PI, prob = 0.95)
g_sim <- sim(quap_m3, data = predicted_data2, n = 1e4)
g_PI <- apply(g_sim, 2, PI)

plot(weight ~ groupsize, data = foxes, col = "azure4")
lines(groupsize_sequence2, mu_mean2)
shade(mu_PI2, groupsize_sequence2, col = col.alpha("lightpink4", alpha = 0.3))
shade(g_PI, groupsize_sequence2, col = col.alpha("cadetblue4", alpha = 0.3))

```



Answer 5H2) We separated the two predictors (groupsize and area) in the previous problem, so they showed little to no relationship with weight, but in the multiple linear regression model we can see that as groupsize increases, weight decreases, and as area increases, weight increases. This is due to a masked relationship, where the two predictors cancel each other out.

5H3. Finally, consider the avgfood variable. Fit two more multiple regressions: (1) body weight as an additive function of avgfood and groupsize, and (2) body weight as an additive function of all three variables, avgfood and groupsize and area. Compare the results of these models to the previous models you've fit, in the first two exercises. (a) Is avgfood or area a better predictor of body weight? If you had to choose one or the other to include in a model, which would it be? Support your assessment with any tables or plots you choose.

By Frederick Richard Rose

```
## weight model with two variables - avgfood and groupsize
m4 <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b3 * avgfood + b4 * groupsize,
  a ~ dnorm(5, 5),
  b3 ~ dnorm(0, 5),
  b4 ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m4 <- quap(m4, data = foxes)

precis(quap_m4)
```

##	mean	sd	5.5%	94.5%
## a	4.1824296	0.42504553	3.5031248	4.8617345
## b3	3.5986866	1.17693503	1.7177171	5.4796561
## b4	-0.5426336	0.15272387	-0.7867159	-0.2985514
## sigma	1.1167424	0.07333571	0.9995378	1.2339470

```
## weight model with three variables - avgfood, groupsize, and area
m5 <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b3 * avgfood + b4 * groupsize + b5 * area,
  a ~ dnorm(5, 5),
  b3 ~ dnorm(0, 5),
  b4 ~ dnorm(0, 5),
  b5 ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m5 <- quap(m5, data = foxes)

precis(quap_m5)
```

##	mean	sd	5.5%	94.5%
## a	4.1020677	0.4230737	3.42591422	4.7782211
## b3	2.3014439	1.3935526	0.07427774	4.5286101
## b4	-0.5925804	0.1538494	-0.83846141	-0.3466994
## b5	0.4015713	0.2360869	0.02425877	0.7788838
## sigma	1.1044101	0.0725162	0.98851519	1.2203050

By Marie Storm Rasmussen

```
# creating two more models with standardized avgfood and area
## standardizing avgfood and area
foxes$avgfood_std <- (foxes$avgfood - mean(foxes$avgfood)) / sd(foxes$avgfood)
foxes$area_std <- (foxes$area - mean(foxes$area)) / sd(foxes$area)

## weight model with two variables - avgfood_std and groupsize
m4_avgfood_std <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b3 * avgfood_std + b4 * groupsize,
  a ~ dnorm(5, 5),
  b3 ~ dnorm(0, 5),
  b4 ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m4_avgfood_std <- quap(m4_avgfood_std, data = foxes)

precis(quap_m4_avgfood_std)
```

##	mean	sd	5.5%	94.5%
## a	6.9566876	0.67995673	5.8699854	8.0433898
## b3	0.7450480	0.23844186	0.3639718	1.1261241
## b4	-0.5587963	0.15473520	-0.8060931	-0.3114996
## sigma	1.1166210	0.07331208	0.9994542	1.2337879

```
## weight model with two variables - area_std and groupsize
m4_area_std <- alist(
  weight ~ dnorm(mu, sigma),
  mu <- a + b3 * area_std + b4 * groupsize,
  a ~ dnorm(5, 5),
  b3 ~ dnorm(0, 5),
  b4 ~ dnorm(0, 5),
  sigma ~ dunif(0, 5))

quap_m4_area_std <- quap(m4_area_std, data = foxes)

precis(quap_m4_area_std)
```

##		mean	sd	5.5%	94.5%
## a		6.3903480	0.53148064	5.5409393	7.2397567
## b3		0.5682703	0.18494809	0.2726876	0.8638531
## b4		-0.4283927	0.12001986	-0.6202076	-0.2365777
## sigma		1.1184568	0.07343088	1.0011001	1.2358136

by **Frederick Richard Rose**

Answer 5H3a) I think avgfood is better predictor of weight since it is better correlated with increase in weight, and larger area can't necessarily imply more food and therefore an increase in weight (although we can of course make that assumption). By standardizing both avgfood and area, the relationship is still stronger between weight and avgfood, so I would just model with avgfood.

b. When both avgfood or area are in the same model, their effects are reduced (closer to zero) and their standard errors are larger than when they are included in separate models. Can you explain this result?

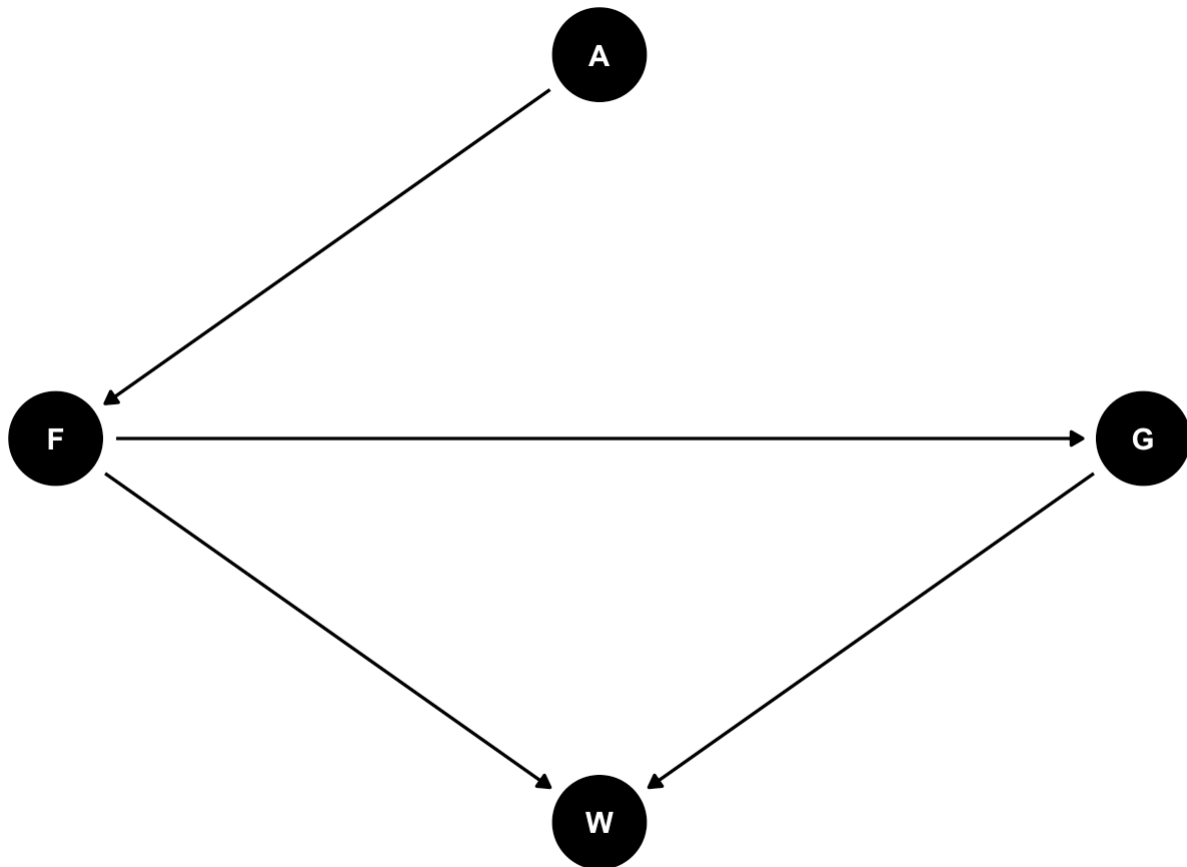
by **Marie Storm Rasmussen**

Answer 5H3b) Both avgfood and area are strongly correlated, so they basically cancel each other out when put into the same model because of multicollinearity.

Defining our theory with explicit DAGs Assume this DAG as an causal explanation of fox weight:

```
pacman::p_load(dagitty,
               ggdag)
dag <- dagitty('dag {
  A[pos="1.000,0.500"]
  F[pos="0.000,0.000"]
  G[pos="2.000,0.000"]
  W[pos="1.000,-0.500"]
  A -> F
  F -> G
  F -> W
  G -> W
}')

# Plot the DAG
ggdag(dag, layout = "circle")+
  theme_dag()
```



where A is area, F is avgfood, G is groupsize, and W is weight.

Using what you know about DAGs from chapter 5 and 6, solve the following three questions:

1. Estimate the total causal influence of A on F. What effect would increasing the area of a territory have on the amount of food inside of it?

By Rune Egeskov Trust

```
foxes$W <- standardize(foxes$weight)
foxes$A <- standardize(foxes$area)
foxes$F <- standardize(foxes$avgfood)
foxes$G <- standardize(foxes$groupsize)
```

```
m6 <- alist(
  F ~ dnorm(mu, sigma),
  mu <- a + bA * A,
  a ~ dnorm(0, 0.2),
  bA ~ dnorm(0, 0.5),
  sigma ~ dexp(1))
```

```
quap_m6 <- quap(m6, data = foxes)
```

```
precis(quap_m6)
```

```
##              mean          sd        5.5%        94.5%
## a      9.144007e-08 0.04231156 -0.06762195 0.06762213
## bA      8.764759e-01 0.04332435  0.80723525 0.94571662
## sigma 4.662631e-01 0.03052536  0.41747764 0.51504848
```

Answer 1) Area has a high influence on avgfood (intuitively we can assume this), and for each one standard deviation increase in area causes a 0.88 increase in avgfood, additionally, since there are no backdoor paths, we don't need to control for additional predictors.

2. Infer the **total** causal effect of adding food F to a territory on the weight W of foxes. Can you calculate the causal effect by simulating an intervention on food?

By Rune Egeskov Trust

```
## no backdoor paths, so only using the one predictor
```

```
m7 <- alist(
  W ~ dnorm(mu, sigma),
  mu <- a + bF * F,
  a ~ dnorm(0, 0.2),
  bF ~ dnorm(0, 0.5),
  sigma ~ dexp(1))
```

```
quap_m7 <- quap(m7, data = foxes)
```

```
precis(quap_m7)
```

```
##              mean          sd        5.5%        94.5%
## a      1.524424e-07 0.08360011 -0.1336090 0.1336093
## bF     -2.421139e-02 0.09088494 -0.1694631 0.1210403
## sigma  9.911431e-01 0.06465844  0.8878064 1.0944797
```


Answer 2) This shows that the increase in avgfood has no effect on the weight (somewhat not intuitive, as one could assume that more food leads to an increase in weight).

3. Infer the **direct** causal effect of adding food F to a territory on the weight W of foxes. In light of your estimates from this problem and the previous one, what do you think is going on with these foxes?

By Rune Egeskov Trust

```
## account for the backdoor path from groupsize G
m8 <- alist(
  W ~ dnorm(mu, sigma),
  mu <- a + bF * F + bG * G,
  a ~ dnorm(0, 0.2),
  c(bF, bG) ~ dnorm(0, 0.5),
  sigma ~ dexp(1))

quap_m8 <- quap(m8, data = foxes)

precis(quap_m8)
```

##		mean	sd	5.5%	94.5%
## a		1.216379e-08	0.08013807	-0.1280761	0.1280761
## bF		4.772531e-01	0.17912322	0.1909796	0.7635266
## bG		-5.735258e-01	0.17914172	-0.8598289	-0.2872228
## sigma		9.420439e-01	0.06175256	0.8433514	1.0407365

```
## checking the effect of avgfood on groupsize
m9 <- alist(
  G ~ dnorm(mu, sigma),
  mu <- a + bF * F,
  a ~ dnorm(0, 0.2),
  bF ~ dnorm(0, 0.5),
  sigma ~ dexp(1))

quap_m9 <- quap(m9, data = foxes)

precis(quap_m9)
```

##		mean	sd	5.5%	94.5%
## a		2.998261e-07	0.03916854	-0.06259858	0.06259918
## bF		8.957170e-01	0.03999383	0.83179914	0.95963488
## sigma		4.301885e-01	0.02816951	0.38516818	0.47520881

Answer 3) We know from the previous problem that more area leads to more avgfood, however more avgfood increases weight, but the backdoor path through groupsize cancels this out. The second model shows us that avgfood has a large effect on groupsize, meaning that more avgfood mean more foxes (groupsize increase), which allows us to infer that the cancelling out is caused by the avgfood being evenly divided amongst the amount of foxes.

Chapter 6: Investigating the Waffles and Divorces

6H1. Use the Waffle House data, `data(WaffleDivorce)`, to find the total causal influence of number of Waffle Houses on divorce rate. Justify your model or models with a causal graph.

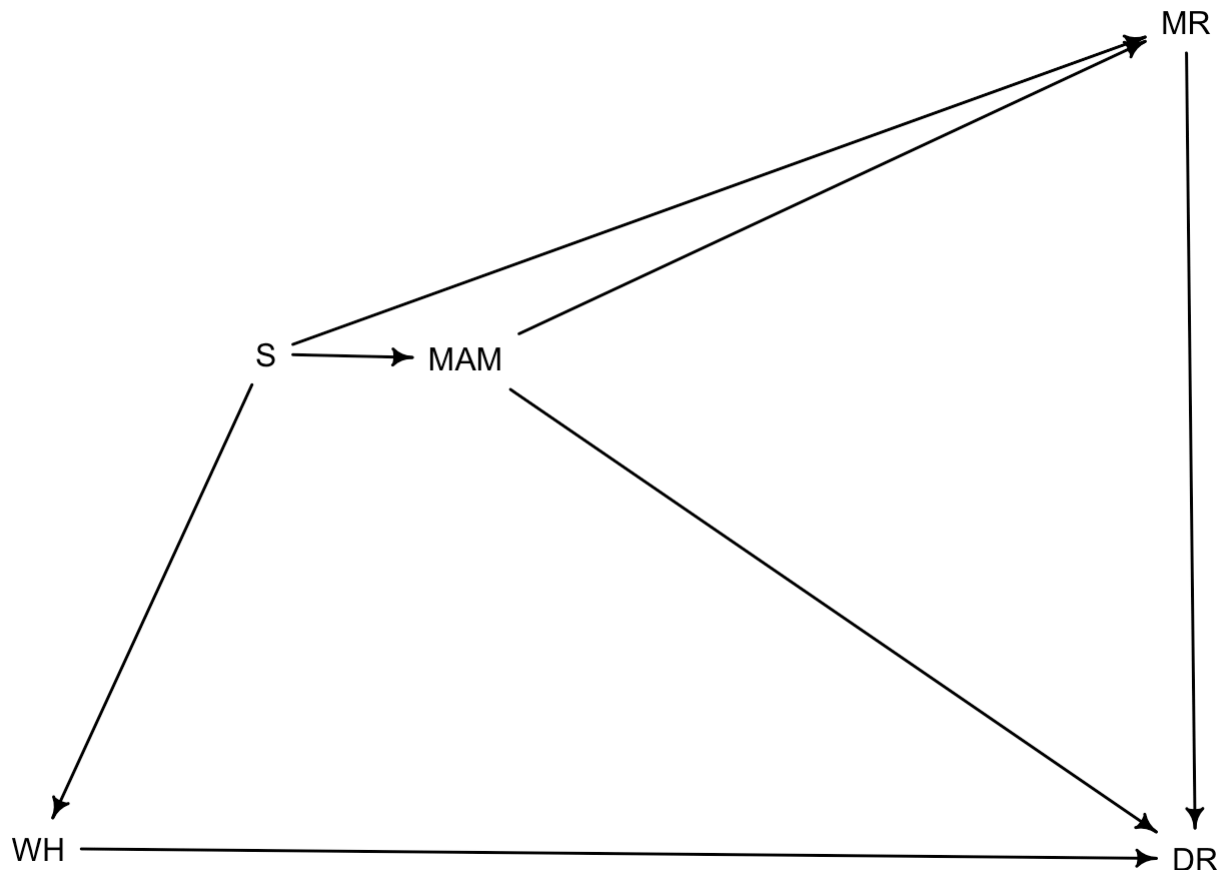
By Markus Lundsryd Jensen

```
data(WaffleDivorce)
```

```
library(dagitty)
```

```
divorce_dag <- dagitty('dag {  
  DR [outcome,pos="1.400,1.621"]  
  MAM [pos="-0.871,0.444"]  
  MR [pos="1.371,-0.348"]  
  S [pos="-1.492,0.433"]  
  WH [exposure,pos="-2.200,1.597"]  
  MAM -> DR  
  MAM -> MR  
  MR -> DR  
  S -> MAM  
  S -> MR  
  S -> WH  
  WH -> DR  
}'  
)
```

```
drawdag(divorce_dag)
```



Answer 6.a) To estimate the total causal influence of number of Waffle Houses on divorce rate given the DAG above, we need to close all backdoors. There are a total of four paths from WH (WaffleHouses) to DR (Divorce). To close the two backdoors (WH->S->MAM->DR and WH->S->SL->MAM->DR and WH->S->MR->DR), we can condition on S, which is a pipe in order to close all the backdoors.

```

WaffleDivorce$WaffleHouses_std <- scale(WaffleDivorce$WaffleHouses)
WaffleDivorce$Divorce_std <- scale(WaffleDivorce$Divorce)

```

```

# Defining the model
m <- quap(
  alist(
    Divorce_std ~ dnorm(mu, sigma),
    mu <- a + bW * WaffleHouses_std + bS * South,
    a ~ dnorm(0, 1),
    bW ~ dnorm(0, 1),
    bS ~ dnorm(0, 1),
    sigma ~ dexp(1)
  ), data = WaffleDivorce
)

precis(m)

```

##		mean	sd	5.5%	94.5%
## a		-0.16779269	0.16390919	-0.42975124	0.09416586
## bW		0.06371281	0.17199033	-0.21116096	0.33858658
## bS		0.60942233	0.36647125	0.02373048	1.19511417
## sigma		0.92094635	0.09094479	0.77559901	1.06629368

Answer 6.b) The total causal effect of WaffleHouses on divorce rate has a mean around zero and 89% credible interval of -0.21 - 0.34. Thus this causal effect is very little if even there.

6H2. Build a series of models to test the implied conditional independencies of the causal graph you used in the previous problem. If any of the tests fail, how do you think the graph needs to be amended? Does the graph need more or fewer arrows? Feel free to nominate variables that aren't in the data.

By Markus Lundsryd Jensen

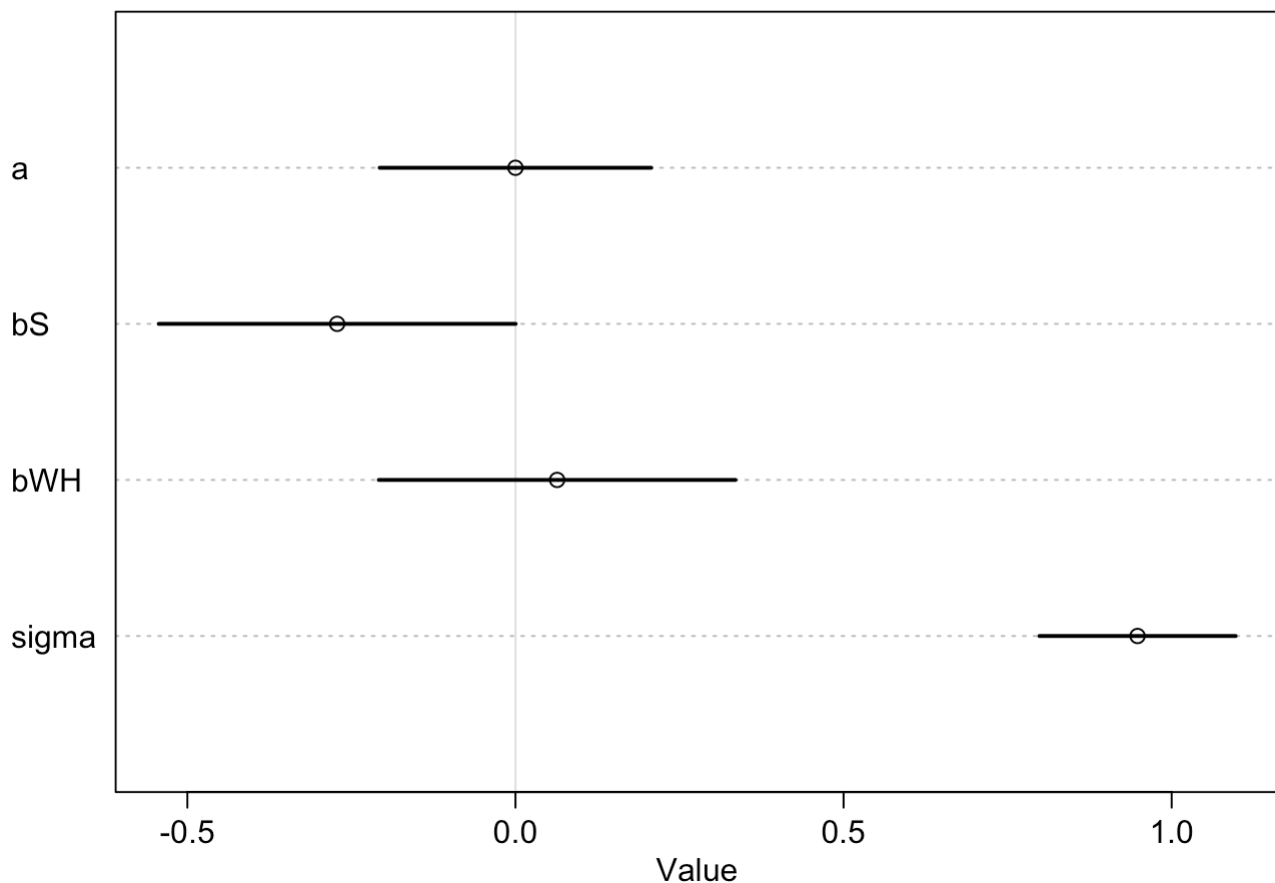
```
impliedConditionalIndependencies( divorce_dag )
```

```
## DR _||_ S | MAM, MR, WH
## MAM _||_ WH | S
## MR _||_ WH | S
```

```
#test the conditional independencies
#scale data

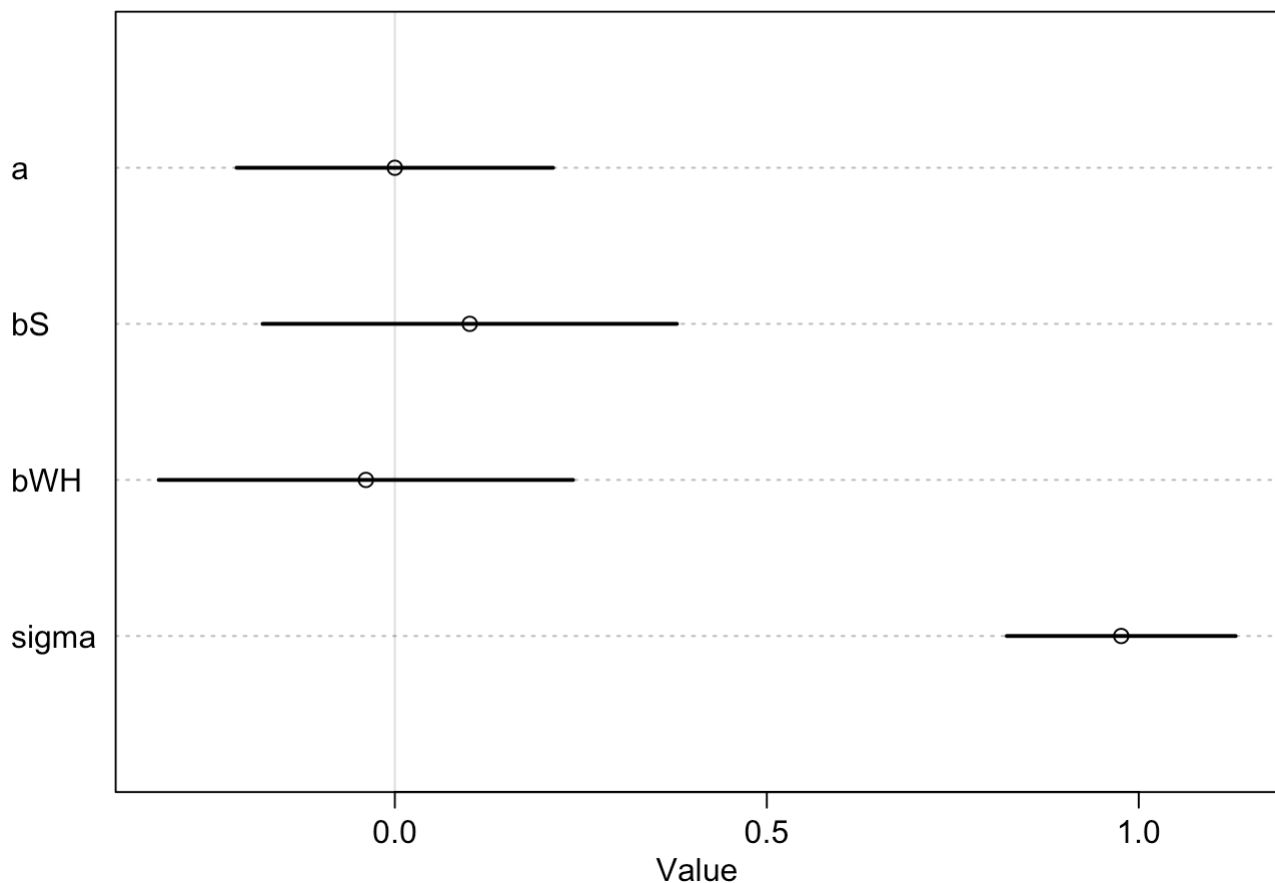
WD <- WaffleDivorce
## Scaling Relevant Variables
WD$MAM <- scale(WD$MedianAgeMarriage)
WD$DR <- scale(WD$Divorce)
WD$MR <- scale(WD$Marriage)
WD$WH <- scale(WD$WaffleHouses)
WD$S <- scale(WD$South)
```

```
# MAM _||_ WH | S
AWS <- quap(
  alist(
    MAM ~ dnorm(mu, sigma),
    mu <- a + bS * S + bWH * WH,
    a ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    bWH ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = WD
)
plot(precis(AWS))
```



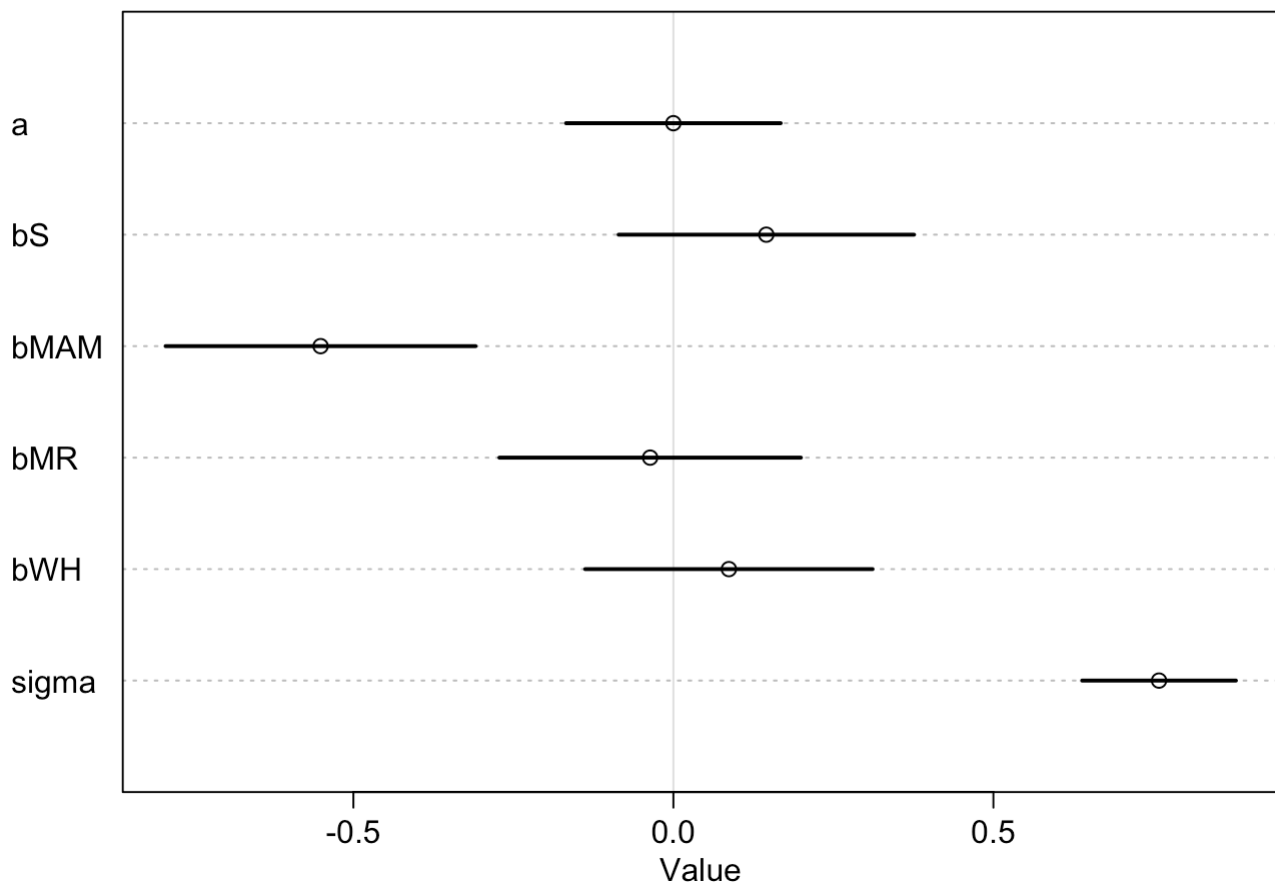
Answer AWS) mean age of marriage is independent of number of Waffle houses when conditioned on South. Once we account for whether a state is in the South (S), the number of Waffle Houses (WH) should no longer have an effect on MAM

```
# MR _||_ WH | S
MRWHS <- quap(
  alist(
    MR ~ dnorm(mu, sigma),
    mu <- a + bWH * WH + bS * S,
    a ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    bWH ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = WD
)
plot(precis(MRWHS))
```



Answer MRWHS) When conditioning on wafflehouse, the effect of South on Marriage rate no longer has a clear effect on Marriage rate

```
# DR _||_ S | MAM, MR, WH
DRSMAMMRWH <- quap(
  alist(
    DR ~ dnorm(mu, sigma),
    mu <- a + bS * S + bMAM * MAM + bMR * MR + bWH * WH,
    a ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    bMAM ~ dnorm(0, 0.5),
    bMR ~ dnorm(0, 0.5),
    bWH ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = WD
)
plot(precis(DRSMAMMRWH))
```



Answer DRSMAMMRWH) When conditioning on mean age of marriage, Marriage rate and Waffle houses, being in the south has no clear influence on divorce rate