Homework-2B

Sam Bashevkin

2022-06-21

```
library(rethinking)
```

```
## Loading required package: rstan

## Loading required package: StanHeaders

##
## rstan version 2.26.11 (Stan version 2.26.1)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
## For within-chain threading using `reduce_sum()` or `map_rect()` Stan functions,
## change `threads_per_chain` option:
## rstan_options(threads_per_chain = 1)

## Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file

## Loading required package: cmdstanr

## This is cmdstanr version 0.5.2

## - CmdStanR documentation and vignettes: mc-stan.org/cmdstanr

## - CmdStan path: C:/Users/sbashevkin/Documents/.cmdstan/cmdstan-2.29.2

## - CmdStan version: 2.29.2

##
## A newer version of CmdStan is available. See ?install_cmdstan() to install it.
## To disable this check set option or environment variable CMDSTANR_NO_VER_CHECK=TRUE.

## Loading required package: parallel

## rethinking (Version 2.21)

##
## Attaching package: 'rethinking'

## The following object is masked from 'package:rstan':
##
##     stan

## The following object is masked from 'package:stats':
##
##     rstudent
```

```
library(knitr)
library(dagitty)
```

1. Construct a linear regression of weight as predicted by height, using the adults (age 18 or greater) from the Howell1 dataset. The heights listed below were recorded in the !Kung census, but weights were not recorded for these individuals. Provide predicted weights and 89% compatibility intervals for each of these individuals. That is, fill in the table below, using model-based predictions.

Reproduce practice data

```
newdat<-data.frame(Individual=1:3, height=c(140, 160, 175))
kable(newdat)
```

Individual	height
1	140
2	160
3	175

Load and filter data

```
data("Howell1")
Howell1_adult<-Howell1[Howell1$age>=18,]
xbar<-mean(Howell1_adult$height)
```

Fit model

```
m1<-quap(alist(
        weight ~ dnorm(mu, sigma),
        mu <- a + b*(height-xbar),
        a ~ dnorm(40,10),
        b ~ dlnorm(0,1),
        sigma ~ dexp(1)
    ),
    data=Howell1_adult)
```

Predict weights

```
pred_weights<-link(m1, newdat)

# 89% Interval
pred_weights_89<-apply(apply(pred_weights, 2, PI), 2, function(x) paste(round(x, 2), collapse=":"))

# median
pred_weights_median<-apply(pred_weights, 2, median)

# Add to table and produce output
newdat$`expected weight`<-pred_weights_median
newdat$`89% interval`<-pred_weights_89

kable(newdat)
```

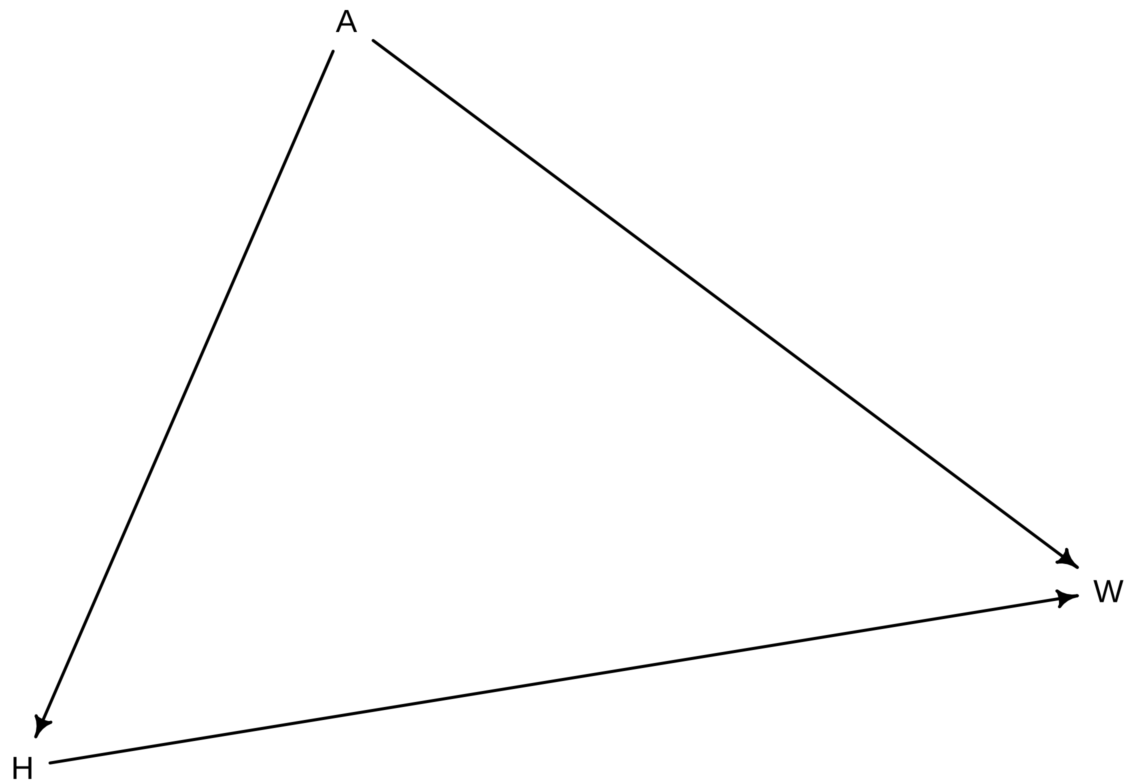
Individual	height	expected weight	89% interval
1	140	35.80466	35.04:36.65

Individual	height	expected weight	89% interval
2	160	48.37287	47.94:48.82
3	175	57.81475	56.79:58.81

2. From the Howell1 dataset, consider only the people younger than 13 years old. Estimate the causal association between age and weight. Assume that age influences weight through two paths. First, age influences height, and height influences weight. Second, age directly influences weight through agerelated changes in muscle growth and body proportions. All of this implies this causal model (DAG) below. Use a linear regression to estimate the total (not just direct) causal effect of each year of growth on weight. Be sure to carefully consider the priors. Try using prior predictive simulation to assess what they imply.

dag

```
dag <- dagitty("dag {
    A -> H -> W
    A -> W
}")
drawdag(dag)
```



data

```
Howell1_under13<-Howell1[Howell1$age<13,]
agebar<-mean(Howell1_under13$age)
heightbar<-mean(Howell1_under13$height)
```

model

```
m2<-quap(alist(
    weight ~ dnorm(mu, sigma),
    mu <- a + Ba*(age-agebar) + Bh*(height-heightbar),
    a ~ dnorm(15, 10),
    Ba ~ dlnorm(0,1),
    Bh ~ dlnorm(0,1),
    sigma ~ dexp(1),

    height ~ dnorm(nu, tau),
    nu <- g + Ba2*(age-agebar),
    g ~ dnorm(100, 20),
    Ba2 ~ dlnorm(0,1),
    tau ~ dexp(1)
),
data=Howell1_under13)
```

Total causal effect

```
post<-extract.samples(m2)
n<-1e4
```

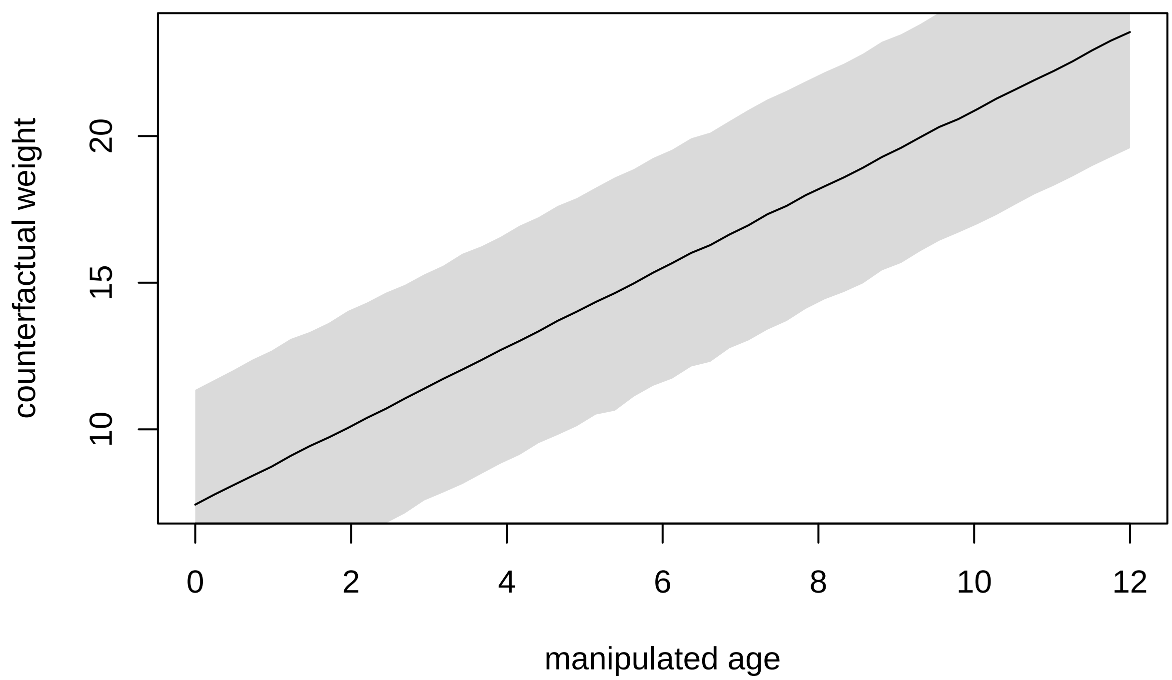
```r
age_seq<-seq(min(Howell1_under13$age), max(Howell1_under13$age), length.out=50)

H_sim <- with(post , sapply( 1:50 ,
function(i) rnorm(n, g + Ba2*(age_seq[i]-agebar), tau )))

W_sim <- with(post , sapply( 1:50 ,
function(i) rnorm(n, a + Ba*(age_seq[i]-agebar) + Bh*(H_sim[,i]-heightbar) , sigma ) ) )
```

Plot the total effect of age on weight

```r
plot(age_seq, colMeans(W_sim), type="l", xlab="manipulated age" , ylab="counterfactual weight")
shade(apply(W_sim,2,PI), age_seq)
```



3. Now suppose the causal association between age and weight might be different for boys and girls. Use a single linear regression, with a categorical variable for sex, to estimate the total causal effect of age on weight separately for boys and girls. How do girls and boys differ? Provide one or more posterior contrasts as a summary.

Prepare index variable

```r
Howell1_under13$sex<-ifelse(Howell1_under13$male==1, 1, 2)
```

model

```r
m3<-quap(alist(
    weight ~ dnorm(mu, sigma),
    mu <- a[sex] + Ba[sex]*(age-agebar),
    a[sex] ~ dnorm(15, 10),
    Ba[sex] ~ dlnorm(0,1),
    sigma ~ dexp(1)
),
data=Howell1_under13)
```

contrasts

```r
muF <- link(m3, data=list(sex=rep(1, 50), age=age_seq, agebar=agebar))
muM <- link(m3, data=list(sex=rep(2, 50), age=age_seq, agebar=agebar))

mu_contrast<-muF-muM
plot(NULL, xlim=range(age_seq), ylim=c(-1, 5),
     xlab="age", ylab="weight contrast (F-M)")
for(p in c(0.5, 0.6, 0.7, 0.8, 0.9, 0.99))
    shade(apply(mu_contrast, 2, PI, prob=p), age_seq)
abline(h=0, lty=2)
```