Statistical Rethinking Winter 2020 – Homework Week 7

Alessandro Gentilini `.@gmail` (just auditing)

January 31, 2021

There are only two problems this week, because the models take a little while to run. Be sure to run the models from the relevant part of Chapter 12 first. That will give you almost all of the model structure you need.

1. In the Trolley data—data(Trolley)—we saw how education level (modeled as an ordered category) is associated with responses. Is this association causal? One plausible confound is that education is also associated with age, through a causal process: People are older when they finish school than when they begin it. Reconsider the Trolley data in this light. Draw a DAG that represents hypothetical causal relationships among response, education, and age. Which statical model or models do you need to evaluate the causal influence of education on responses? Fit these models to the trolley data. What do you conclude about the causal relationships among these three variables?

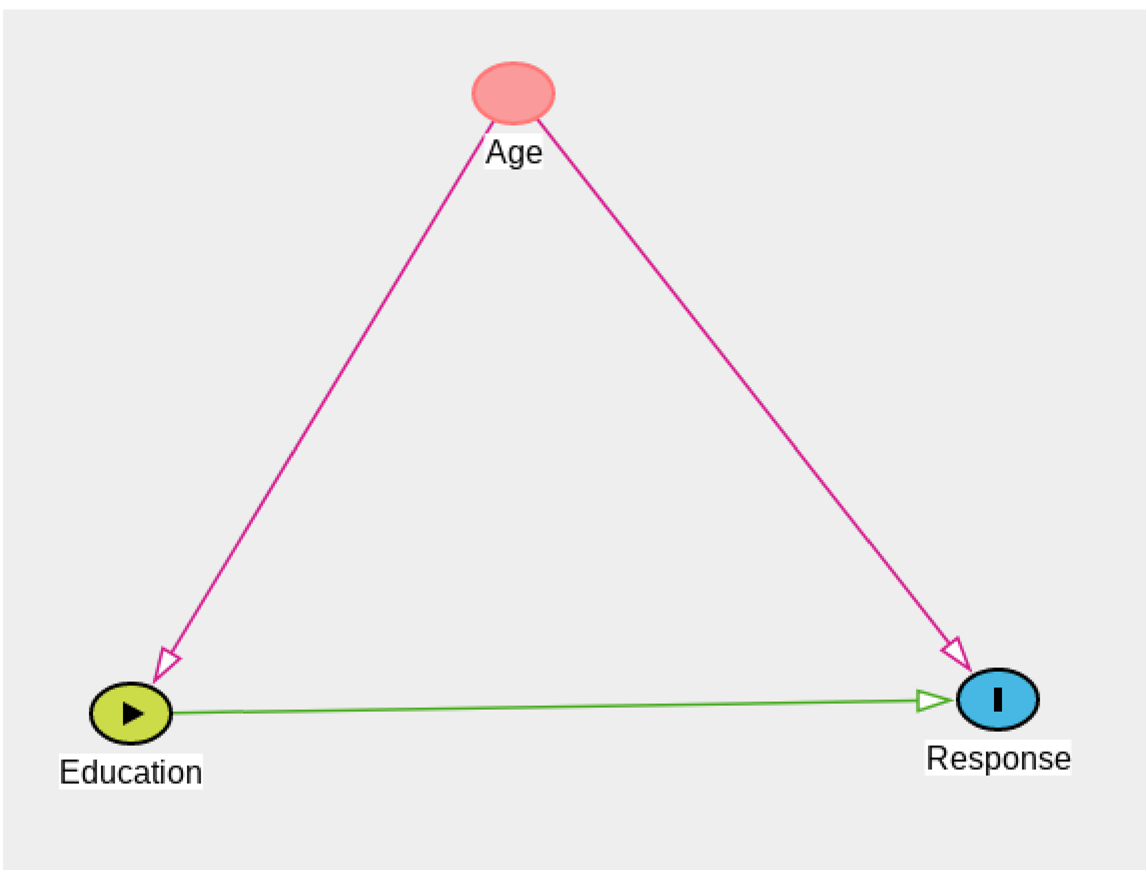


Figure 1: DAG for question 1.

```r
## R code 12.30
library(rethinking)
```

```
## Loading required package: rstan

## Loading required package: StanHeaders

## Loading required package: ggplot2

## rstan (Version 2.21.2, GitRev: 2e1f913d3ca3)

## 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)

## Loading required package: parallel

## rethinking (Version 2.13)

##
## Attaching package: 'rethinking'

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

```r
data(Trolley)
d <- Trolley
levels(d$edu)
```

```
## [1] "Bachelor's Degree"    "Elementary School"    "Graduate Degree"
## [4] "High School Graduate" "Master's Degree"      "Middle School"
## [7] "Some College"         "Some High School"
```

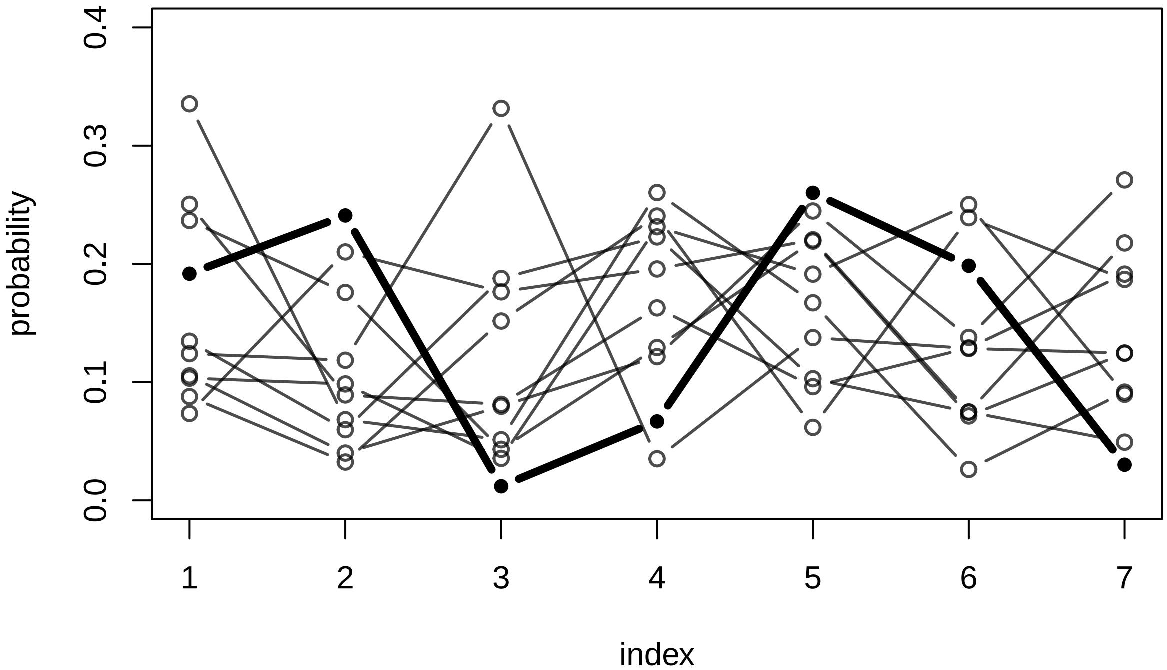
```r
## R code 12.31
edu_levels <- c( 6 , 1 , 8 , 4 , 7 , 2 , 5 , 3 )
d$edu_new <- edu_levels[ d$edu ]

## R code 12.32
library(gtools)
```

```
##
## Attaching package: 'gtools'

## The following object is masked from 'package:rethinking':
##
##     logit
```

```r
set.seed(1805)
delta <- rdirichlet( 10 , alpha=rep(2,7) )
str(delta)
```

```
##  num [1:10, 1:7] 0.1053 0.2504 0.1917 0.1241 0.0877 ...
```

```r
## R code 12.33
h <- 3
plot( NULL , xlim=c(1,7) , ylim=c(0,0.4) , xlab="index" , ylab="probability" )
for ( i in 1:nrow(delta) ) lines( 1:7 , delta[i,] , type="b" ,
    pch=ifelse(i==h,16,1) , lwd=ifelse(i==h,4,1.5) ,
    col=ifelse(i==h,"black",col.alpha("black",0.7)) )
```

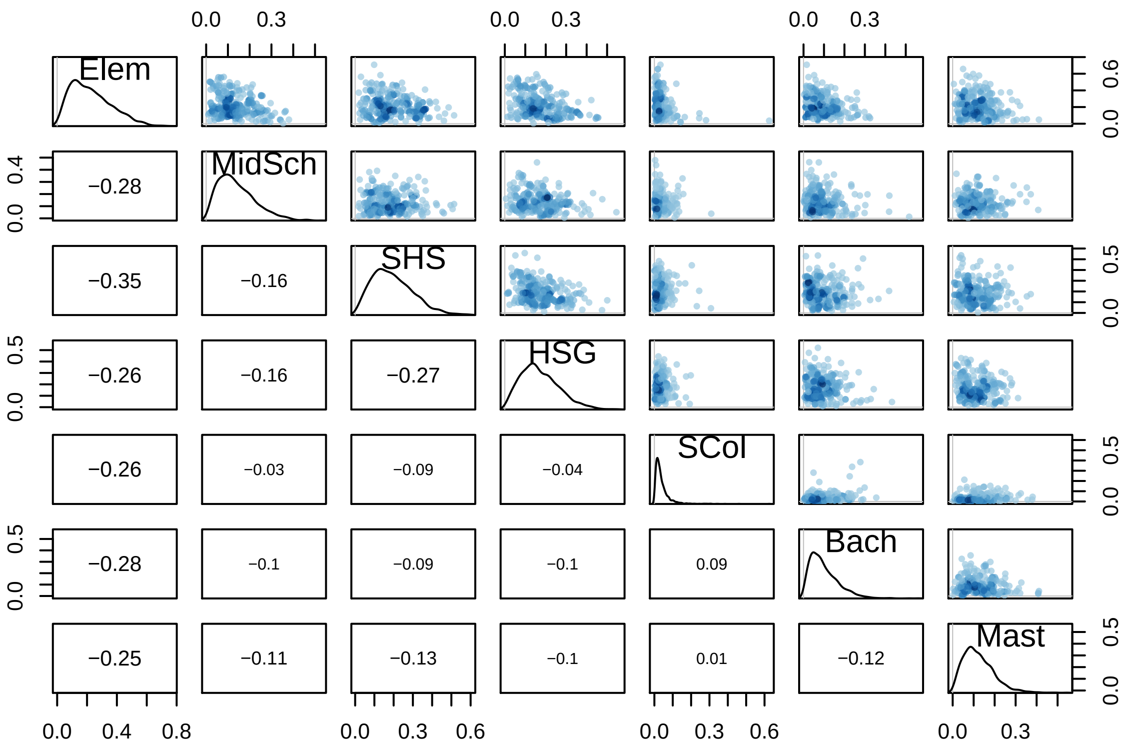



```
# ## R code 12.34
 dat <- list(
     R = d$response ,
     action = d$action,
     intention = d$intention,
     contact = d$contact,
     E = as.integer( d$edu_new ),   # edu_new as an index
     alpha = rep( 2 , 7 ) )       # delta prior

 m12.6 <- ulam(
     alist(
         R ~ ordered_logistic( phi , kappa ),
         phi <- bE*sum( delta_j[1:E] ) + bA*action + bI*intention + bC*contact,
         kappa ~ normal( 0 , 1.5 ),
         c(bA,bI,bC,bE) ~ normal( 0 , 1 ),
         vector[8]: delta_j <<- append_row( 0 , delta ),
         simplex[7]: delta ~ dirichlet( alpha )
     ), data=dat , chains=4 , cores=4 )

# ## R code 12.35
precis( m12.6 , depth=2 , omit="kappa" )
```

```
##                 mean         sd         5.5%       94.5%     n_eff     Rhat4
## bE       -0.31727323 0.16012960 -0.573320112 -0.07822779  800.5012 1.0053344
## bC       -0.95542444 0.05049957 -1.037931536 -0.87665760 2379.9029 0.9993236
## bI       -0.71662173 0.03820488 -0.778303696 -0.65623659 2726.0276 1.0010588
## bA       -0.70300223 0.04160926 -0.769633339 -0.63675830 2262.0406 0.9993690
## delta[1]  0.22726591 0.13587507  0.051012364  0.47654215 1277.8768 1.0021468
## delta[2]  0.14082934 0.08598181  0.030711545  0.30197552 3050.7783 1.0004336
## delta[3]  0.19444887 0.10590654  0.048384532  0.37863629 2314.7446 0.9990815
## delta[4]  0.17215053 0.09383417  0.041109063  0.33791874 2120.0754 0.9985608
## delta[5]  0.04018656 0.04386371  0.005840643  0.10832689  961.5685 1.0025040
## delta[6]  0.10046595 0.07102419  0.016962032  0.23321077 1790.1021 0.9998704
```

```
## delta[7]  0.12465284 0.07550851  0.028851242  0.25562809 2152.9781 1.0002371
```

```
# ## R code 12.36
delta_labels <- c("Elem","MidSch","SHS","HSG","SCol","Bach","Mast","Grad")
pairs( m12.6 , pars="delta" , labels=delta_labels )
```



I add age to the model as an ordinary continuous variable:

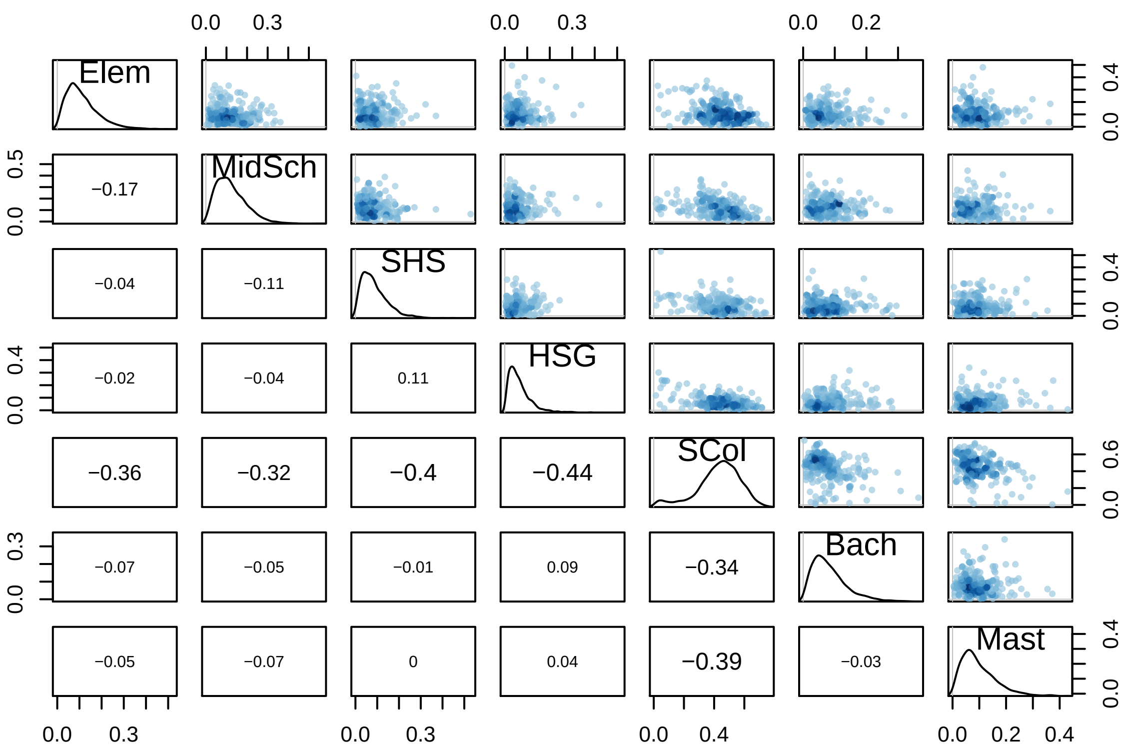
```
dat2 <- list(
    R = d$response ,
    action = d$action,
    intention = d$intention,
    contact = d$contact,
    E = as.integer( d$edu_new ),   # edu_new as an index
    age=d$age,
    alpha = rep( 2 , 7 ) )      # delta prior

m12.6_2 <- ulam(
    alist(
        R ~ ordered_logistic( phi , kappa ),
        phi <- bE*sum( delta_j[1:E] )+ bAge*age + bA*action + bI*intention + bC*contact,
        kappa ~ normal( 0 , 1.5 ),
        c(bA,bI,bC,bE,bAge) ~ normal( 0 , 1 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( alpha )
    ), data=dat2 , chains=4 , cores=4 )
```

```
## Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quant
## Running the chains for more iterations may help. See
## http://mc-stan.org/misc/warnings.html#tail-ess
```

```r
# ## R code 12.35
precis( m12.6_2 , depth=2 , omit="kappa" )
```

```
##                  mean          sd         5.5%        94.5%     n_eff     Rhat4
## bAge     -0.006744346 0.001506974 -0.009010825 -0.004217267  998.0594 1.0008678
## bE        0.228824215 0.111275111  0.038284224  0.370442935  249.7191 1.0102769
## bC       -0.957886487 0.051256933 -1.039287869 -0.875874552 2093.3433 0.9994386
## bI       -0.716874172 0.035711308 -0.773375741 -0.658730065 2411.9922 0.9991339
## bA       -0.705297069 0.042205926 -0.774186114 -0.638379831 2250.1717 0.9988639
## delta[1]  0.112696823 0.074865855  0.023306100  0.250885774 1871.9224 0.9995366
## delta[2]  0.123093027 0.074976721  0.028077598  0.257936789 2495.0870 1.0001308
## delta[3]  0.087849037 0.062986681  0.015696794  0.195996489 1004.9204 0.9999142
## delta[4]  0.066277062 0.053373419  0.012295905  0.162124629  599.3231 1.0013369
## delta[5]  0.429433617 0.143509297  0.140947898  0.627213659  343.3507 1.0046630
## delta[6]  0.083431137 0.055543581  0.017682313  0.190752627 1943.9609 1.0000785
## delta[7]  0.097219298 0.064669635  0.019551815  0.216685136 1501.7350 1.0025907
```

```r
# ## R code 12.36
delta_labels <- c("Elem","MidSch","SHS","HSG","SCol","Bach","Mast","Grad")
pairs( m12.6_2 , pars="delta" , labels=delta_labels )
```



Age seems to be not influential, but after adding it to the model, the overall association of education `bE` is now positive; I cannot explain this result.

---

2. Consider one more variable in the Trolley data: Gender. Suppose that gender might influence education as well as response directly. Draw the DAG now that includes response, education, age, and gender. Using only the DAG, is it possible that the inferences from Problem 1 are confounded by gender? If so, define any additional models you need to infer the causal influence of education on response. What do you conclude?

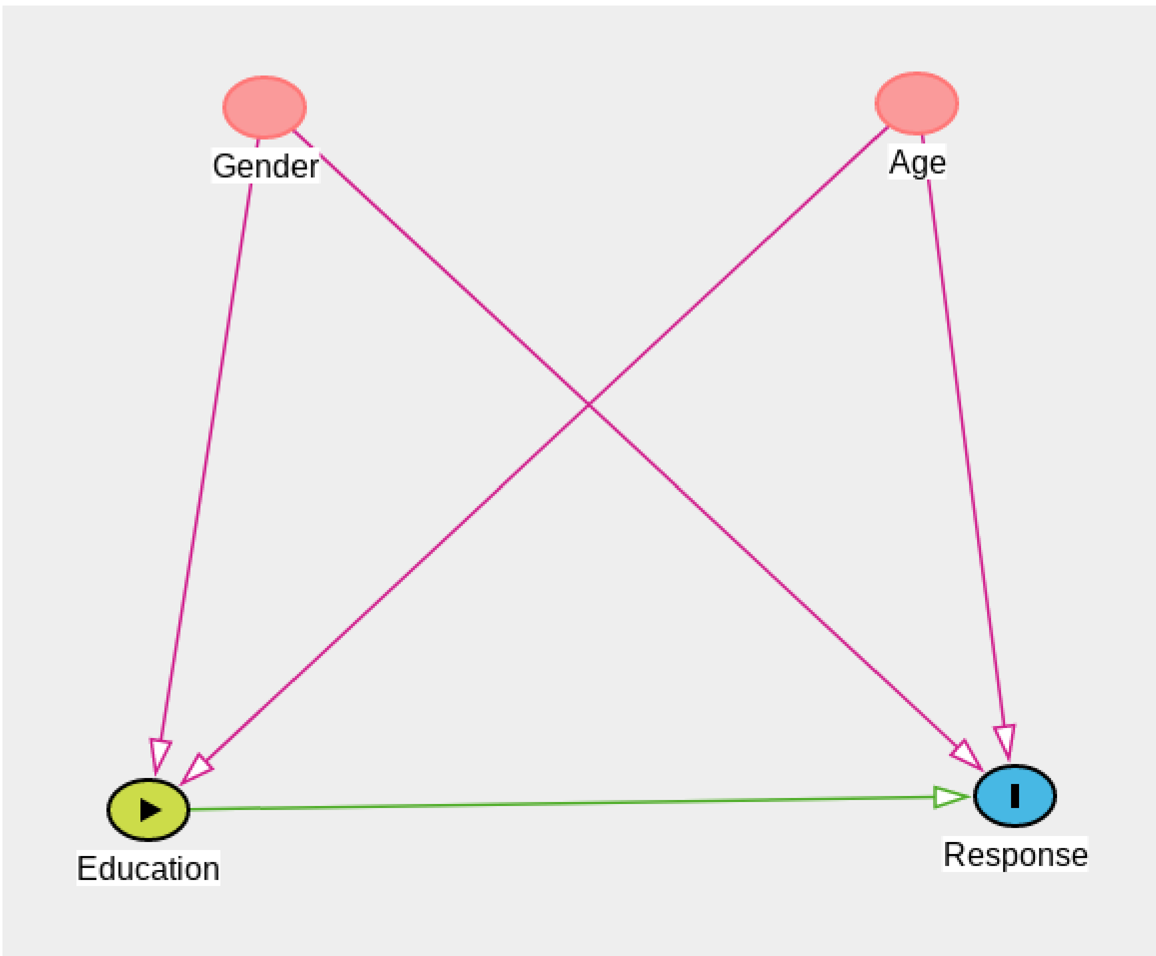


Figure 2: DAG for question 2.

I would say that it is possible that the inferences from Problem 1 are confounded by gender, but I have not a logical exaplanation to that since I did not still fully grasped the causal inference and the DAGs.

I will add an index variable `sex` and then education `bE` and age `bAge` coefficients will get indexed by `sex`.

```
d$sex <- ifelse(d$male==1,2,1)
dat3 <- list(
    R = d$response ,
    sex = d$sex,
    action = d$action,
    intention = d$intention,
    contact = d$contact,
    E = as.integer( d$edu_new ),   # edu_new as an index
    age=d$age,
    alpha = rep( 2 , 7 ) )      # delta prior

m12.6_3 <- ulam(
    alist(
        R ~ ordered_logistic( phi , kappa ),
        phi <- bE[sex]*sum( delta_j[1:E] )+ bAge[sex]*age + bA*action + bI*intention + bC*contact,
        bE[sex] ~ normal( 0 , 1 ),
        bAge[sex] ~ normal( 0 , 1 ),
        kappa ~ normal( 0 , 1.5 ),
        c(bA,bI,bC) ~ normal( 0 , 1 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( alpha )
    ), data=dat3 , chains=4 , cores=4 )

## R code 12.35
precis( m12.6_3 , depth=2 , omit="kappa" )
```

```
##                  mean          sd         5.5%        94.5%    n_eff     Rhat4
## bE[1]    -0.724026097 0.142110052 -0.942174572 -0.494648379 1393.139 1.0009157
## bE[2]     0.322306027 0.148003315  0.089055925  0.562531720 1322.002 1.0046717
## bAge[1]  -0.001073866 0.001997220 -0.004157666  0.002205990 2740.817 0.9992243
## bAge[2]  -0.007330786 0.001952283 -0.010490392 -0.004247653 2351.315 1.0020842
## bC       -0.968589075 0.049314171 -1.047631245 -0.890384020 2079.237 0.9997013
## bI       -0.724000338 0.037689772 -0.784660593 -0.664255388 2689.168 1.0003982
## bA       -0.710839914 0.039666954 -0.773550761 -0.646139452 1934.220 1.0002155
## delta[1]  0.176473184 0.091561912  0.047729628  0.338094154 1595.680 0.9989730
## delta[2]  0.150904535 0.086538274  0.035310428  0.311104237 2058.370 0.9986070
## delta[3]  0.286388887 0.110006167  0.112522779  0.460111166 1428.895 0.9999640
## delta[4]  0.083812239 0.049187998  0.017991117  0.169458569 2351.030 1.0001993
## delta[5]  0.042909937 0.032103811  0.007612822  0.101805807 1547.908 1.0008782
## delta[6]  0.223327020 0.067225560  0.113734351  0.332887130 2032.988 1.0001273
## delta[7]  0.036184199 0.025271662  0.006471726  0.086149936 2501.478 1.0003297
```

Age is still not influential and now there is a positive association of education for men and a negative association of education for women.