Homework 9

Question 1

Revisit the Bangladesh fertility data, data(bangladesh). Fit a model with both varying intercepts by district_id and varying slopes of urban (as a 0/1 indicator variable) by district_id. You are still predicting use.contraception. Inspect the correlation between the intercepts and slopes. Can you interpret this correlation, in terms of what it tells you about the pattern of contraceptive use in the sample? It might help to plot the varying effect estimates for both the intercepts and slopes, by district. Then you can visualize the correlation and maybe more easily think through what it means to have a particular correlation. Plotting predicted proportion of women using contraception, in each district, with urban women on one axis and rural on the other, might also help.

```
data(bangladesh)

d_q1 <- bangladesh

d_q1$district_id <- as.integer(as.factor(d_q1$district))

dat_q1 <- data.frame(
  use.contraception=d_q1$use.contraception,
  district_id = as.factor(d_q1$district),
  urban=d_q1$urban
)

dat_list_q1 <- list(
  C=d_q1$use.contraception,
  D=d_q1$district_id,
  U=d_q1$urban,
  Uid=d_q1$urban + 1L,
  N=1934,
  K=max(d_q1$district_id)
)
```

Lets first fit the fixed effect model using MLE GLM model

```
q1_a <- glm(use.contraception ~ district_id : urban, family=binomial, data=dat_q1)

summary.glm(q1_a, correlation = FALSE)
```

```
##
## Call:
## glm(formula = use.contraception ~ district_id:urban, family = binomial,
##     data = dat_q1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1899  -0.9147  -0.9147   1.4652   2.0393
##
## Coefficients: (15 not defined because of singularities)
```

```
##                       Estimate Std. Error z value Pr(>|z|)    
## (Intercept)           -0.65512    0.05692 -11.510  < 2e-16 ***
## district_id1:urban     0.10173    0.26780   0.380  0.70403    
## district_id2:urban          NA         NA      NA       NA    
## district_id3:urban    15.22119  624.19383   0.024  0.98055    
## district_id4:urban     2.95770    1.05035   2.816  0.00486 ** 
## district_id5:urban     0.65512    1.41536   0.463  0.64346    
## district_id6:urban     1.57141    0.83859   1.874  0.06095 .  
## district_id7:urban          NA         NA      NA       NA    
## district_id8:urban    15.22119  624.19383   0.024  0.98055    
## district_id9:urban     1.34827    1.22607   1.100  0.27148    
## district_id10:urban         NA         NA      NA       NA    
## district_id11:urban         NA         NA      NA       NA    
## district_id12:urban   -0.03803    0.86789  -0.044  0.96505    
## district_id13:urban    0.14429    0.73251   0.197  0.84384    
## district_id14:urban    1.37812    0.21966   6.274 3.52e-10 ***
## district_id15:urban    0.14429    0.73251   0.197  0.84384    
## district_id16:urban   15.22119  624.19383   0.024  0.98055    
## district_id17:urban         NA         NA      NA       NA    
## district_id18:urban    0.65512    0.53754   1.219  0.22295    
## district_id19:urban    1.75373    1.15610   1.517  0.12928    
## district_id20:urban         NA         NA      NA       NA    
## district_id21:urban   -1.29079    1.07056  -1.206  0.22793    
## district_id22:urban         NA         NA      NA       NA    
## district_id23:urban         NA         NA      NA       NA    
## district_id24:urban         NA         NA      NA       NA    
## district_id25:urban    0.43198    0.47774   0.904  0.36589    
## district_id26:urban         NA         NA      NA       NA    
## district_id27:urban    0.24965    0.91464   0.273  0.78489    
## district_id28:urban   -0.44349    1.15610  -0.384  0.70127    
## district_id29:urban    0.94280    0.76588   1.231  0.21832    
## district_id30:urban    1.75373    0.58015   3.023  0.00250 ** 
## district_id31:urban    0.65512    0.81848   0.800  0.42347    
## district_id32:urban         NA         NA      NA       NA    
## district_id33:urban    1.57141    0.83859   1.874  0.06095 .  
## district_id34:urban   -0.03803    0.70939  -0.054  0.95725    
## district_id35:urban    0.85579    0.45306   1.889  0.05890 .  
## district_id36:urban   -0.03803    1.22607  -0.031  0.97526    
## district_id37:urban         NA         NA      NA       NA    
## district_id38:urban    0.94280    0.76588   1.231  0.21832    
## district_id39:urban    0.65512    1.41536   0.463  0.64346    
## district_id40:urban    0.58613    0.37595   1.559  0.11898    
## district_id41:urban  -13.91095  509.65213  -0.027  0.97822    
## district_id42:urban   -0.73117    1.11948  -0.653  0.51367    
## district_id43:urban    1.01179    0.49608   2.040  0.04139 *  
## district_id44:urban         NA         NA      NA       NA    
## district_id45:urban    2.04141    1.11948   1.824  0.06822 .  
## district_id46:urban    1.34827    0.61501   2.192  0.02836 *  
## district_id47:urban    0.65512    0.81848   0.800  0.42347    
## district_id48:urban    0.90643    0.50716   1.787  0.07389 .  
## district_id49:urban         NA         NA      NA       NA    
## district_id50:urban   15.22119  441.37169   0.034  0.97249    
## district_id51:urban    1.01179    0.49608   2.040  0.04139 *  
## district_id52:urban   -0.37450    0.52409  -0.715  0.47487    
```

```
## district_id53:urban   0.33667    0.46813   0.719  0.47204
## district_id55:urban  -0.95432    1.09692  -0.870  0.38430
## district_id56:urban   1.14063    0.45295   2.518  0.01179 *
## district_id57:urban  -0.44349    1.15610  -0.384  0.70127
## district_id58:urban  -0.15581    0.60361  -0.258  0.79631
## district_id59:urban        NA         NA      NA       NA
## district_id60:urban  -0.19218    0.69241  -0.278  0.78136
## district_id61:urban  -0.32571    0.67939  -0.479  0.63164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2590.9  on 1933  degrees of freedom
## Residual deviance: 2456.9  on 1888  degrees of freedom
## AIC: 2548.9
##
## Number of Fisher Scoring iterations: 13
```

Lets fit the fixed effect model using bayesian inference

```
model_q1_1_fit <- stan(file='week09/09_q1_1.stan', data=dat_list_q1, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q1_1.stan'
```

```
loo(model_q1_1_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1318.5 17.7
## p_loo        85.6  2.4
## looic      2636.9 35.5
## ------
## Monte Carlo SE of elpd_loo is 0.2.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
model_q1_2_fit <- stan(file='week09/09_q1_2.stan', data=dat_list_q1, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q1_2.stan'

## Warning: There were 1 divergent transitions after warmup. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
## to find out why this is a problem and how to eliminate them.

## Warning: There were 1 chains where the estimated Bayesian Fraction of Missing Information was low. Se
## http://mc-stan.org/misc/warnings.html#bfmi-low

## Warning: Examine the pairs() plot to diagnose sampling problems

## Warning: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and medians may be
## Running the chains for more iterations may help. See
```

```
## http://mc-stan.org/misc/warnings.html#bulk-ess
```

```
## 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
```

```
loo(model_q1_2_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1240.5 13.8
## p_loo        49.9  1.0
## looic      2480.9 27.7
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
model_q1_3_fit <- stan(file='week09/09_q1_3.stan', data=dat_list_q1, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q1_3.stan'
```

```
loo(model_q1_3_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1234.3 14.0
## p_loo        52.0  1.0
## looic      2468.5 28.1
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
model_q1_4_fit <- stan(file='week09/09_q1_4.stan', data=dat_list_q1, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q1_4.stan'
```

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

```
## 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
```

```
loo(model_q1_4_fit)
```

```
##
```

```
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1233.6 14.1
## p_loo        51.3  1.0
## looic      2467.2 28.2
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

Question 2

Now consider the predictor variables age.centered and living.children, also contained in data(bangladesh). Suppose that age influences contraceptive use (changing attitudes) and number of children (older people have had more time to have kids). Number of children may also directly influence contraceptive use. Draw a DAG that reflects these hypothetical relationships. Then build models needed to evaluate the DAG. You will need at least two models. Retain district and urban, as in Problem 1. What do you conclude about the causal influence of age and children?

Lets first decide on the DAG.

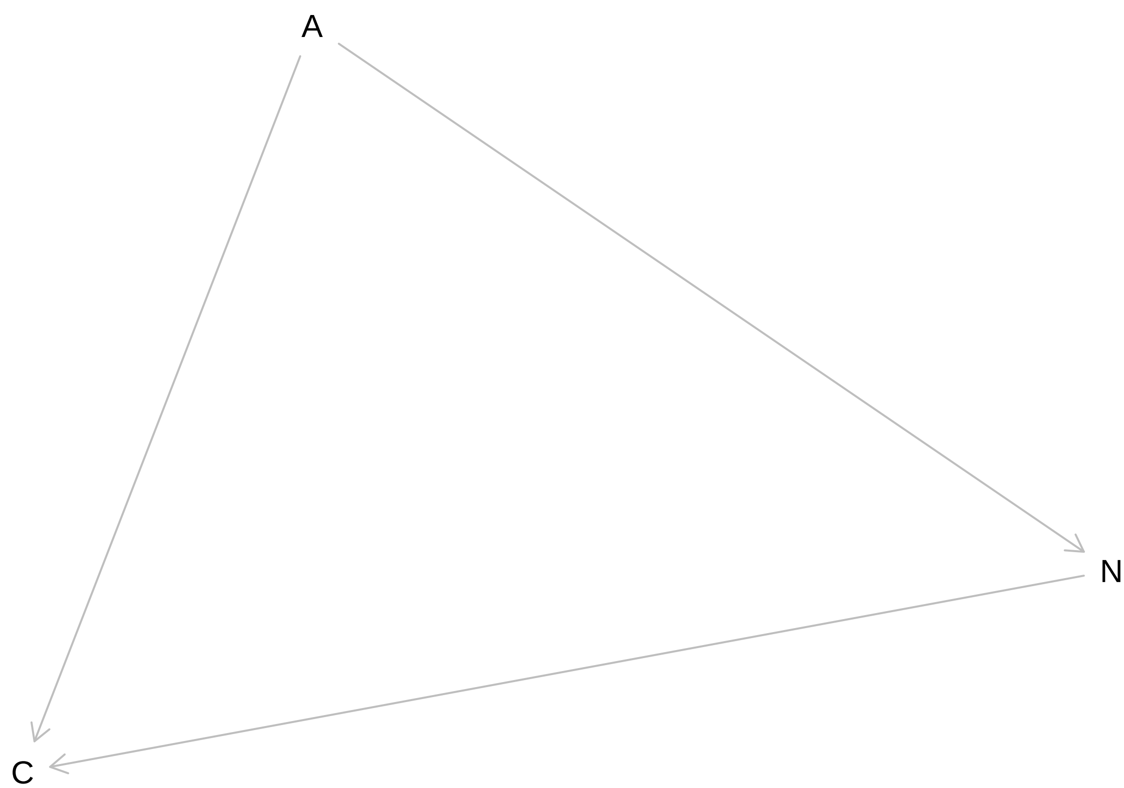
```
dag <- dagitty("dag{
A -> N
A -> C
N -> C
}")

adjustmentSets( dag , exposure="A" , outcome="C" , effect="direct" )
```

```
## { N }
```

```
plot(dag)
```

```
## Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to set you
```




```
d_q2 <- bangladesh

d_q2$district_id <- as.integer(as.factor(d_q2$district))

dat_q2 <- data.frame(
  use.contraception=d_q2$use.contraception,
  district_id = as.numeric(as.factor(d_q2$district)),
  age.centered = standardize(d_q2$age.centered),
  living.children = standardize(d_q2$living.children),
  urban=d_q2$urban
)

dat_list_q2 <- list(
  C=dat_q2$use.contraception,
  D=dat_q2$district_id,
  U=dat_q2$urban,
  Ch=dat_q2$living.children,
  A=dat_q2$age.centered,
  N=1934,
  K=max(dat_q2$district_id)
)
```

```
model_q2_1_fit <- stan(file='week09/09_q2_1.stan', data=dat_list_q2, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q2_1.stan'
```

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

```
loo(model_q2_1_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1233.9 14.1
## p_loo        53.4  1.0
## looic      2467.8 28.3
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
model_q2_2_fit <- stan(file='week09/09_q2_2.stan', data=dat_list_q2, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q2_2.stan'
```

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

```
## 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
```

```
loo(model_q2_2_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1206.7 15.4
## p_loo        55.2  1.1
## looic      2413.5 30.8
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

Question 3

Modify any models from Problem 2 that contained that children variable and model the variable now as a monotonic ordered category, like education from the week we did ordered categories. Education in that example had 8 categories. Children here will have fewer (no one in the sample had 8 children). So modify the code appropriately. What do you conclude about the causal influence of each additional child on use of contraception?

```
d_q3 <- bangladesh
```

```r
d_q3$district_id <- as.integer(as.factor(d_q3$district))

dat_q3 <- data.frame(
  use.contraception=d_q3$use.contraception,
  district_id = as.numeric(as.factor(d_q3$district)),
  age.centered = standardize(d_q3$age.centered),
  living.children = d_q3$living.children,
  urban=d_q3$urban
)

dat_list_q3 <- list(
  C=dat_q3$use.contraception,
  D=dat_q3$district_id,
  U=dat_q3$urban,
  Ch=dat_q3$living.children,
  A=dat_q3$age.centered,
  N=1934,
  K=max(dat_q3$district_id)
)
```

```r
model_q3_1_fit <- stan(file='week09/09_q3_1.stan', data=dat_list_q3, cores=4)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final
## line found on 'C:\Users\Orcun Gumus\OneDrive - McKinsey &
## Company\Desktop\statrethinking_winter2019\week09\09_q3_1.stan'

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

```r
loo(model_q3_1_fit)
```

```
##
## Computed from 4000 by 1934 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo  -1195.7 14.7
## p_loo        52.4  1.0
## looic      2391.4 29.5
## ------
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```