hw4

GISH

2019撟4<9c><88>23

### Q1

# (a)

d <- rugged  
dnoS <- rugged[!(d$country=="Seychelles"),] #remove Seychelles  
dnoS$loggdp <- log(dnoS$rgdppc\_2000)  
  
dF <- dnoS[complete.cases(dnoS$loggdp),]  
  
modelAIS <- map(  
 alist(  
 loggdp ~ dnorm (mu,sigma),  
 mu <- a + bA \* cont\_africa +bR \* rugged +bAR \* cont\_africa \* rugged,  
 a ~ dnorm (0,10),  
 bA ~ dnorm(0,10),  
 bR ~ dnorm(0,10),  
 bAR ~ dnorm(0,10),  
 sigma ~ dunif(0,30)  
   
 ),  
 data = dF  
)  
  
print(modelAIS)

##   
## Maximum a posteriori (MAP) model fit  
##   
## Formula:  
## loggdp ~ dnorm(mu, sigma)  
## mu <- a + bA \* cont\_africa + bR \* rugged + bAR \* cont\_africa \*   
## rugged  
## a ~ dnorm(0, 10)  
## bA ~ dnorm(0, 10)  
## bR ~ dnorm(0, 10)  
## bAR ~ dnorm(0, 10)  
## sigma ~ dunif(0, 30)  
##   
## MAP values:  
## a bA bR bAR sigma   
## 9.2211045 -1.8797313 -0.2019184 0.2964584 0.9253674   
##   
## Log-likelihood: -226.69

d$loggdp <- log(d$rgdppc\_2000)  
d <- d[complete.cases(d$loggdp),]  
  
modelAI <- map(  
 alist(  
 loggdp ~ dnorm (mu,sigma),  
 mu <- a + bA \* cont\_africa +bR \* rugged +bAR \* cont\_africa \* rugged,  
 a ~ dnorm (0,10),  
 bA ~ dnorm(0,10),  
 bR ~ dnorm(0,10),  
 bAR ~ dnorm(0,10),  
 sigma ~ dunif(0,30)  
   
 ),  
 data = d  
)  
  
print(modelAI)

##   
## Maximum a posteriori (MAP) model fit  
##   
## Formula:  
## loggdp ~ dnorm(mu, sigma)  
## mu <- a + bA \* cont\_africa + bR \* rugged + bAR \* cont\_africa \*   
## rugged  
## a ~ dnorm(0, 10)  
## bA ~ dnorm(0, 10)  
## bR ~ dnorm(0, 10)  
## bAR ~ dnorm(0, 10)  
## sigma ~ dunif(0, 30)  
##   
## MAP values:  
## a bA bR bAR sigma   
## 9.2210506 -1.9452132 -0.2018925 0.3921466 0.9326804   
##   
## Log-likelihood: -229.37

有Seychelles的資料比沒Seychelles的資料多了約0.1的bAR(即交互作用的係數)，可見納入Seychelles，使continents影響(ruggedness對於loggdp的效果)增加了0.1。

# (b)

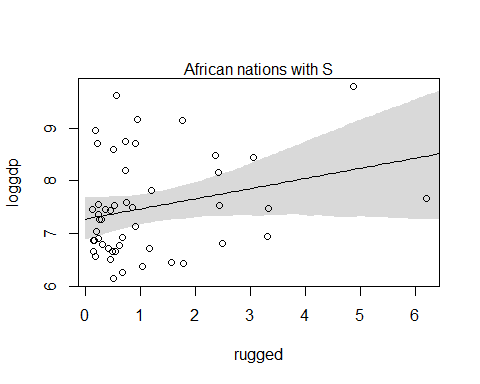
ruggedSeq = seq(from=0, to=7,length.out=100)  
mu.NotAfrica <- link(modelAI,data = data.frame(rugged = ruggedSeq,cont\_africa = 0))

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

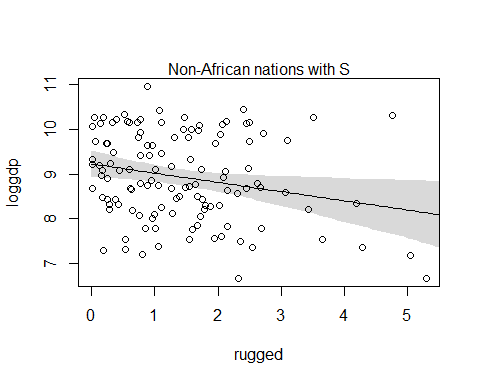
mu.Africa <- link(modelAI,data = data.frame(rugged = ruggedSeq,cont\_africa = 1))

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

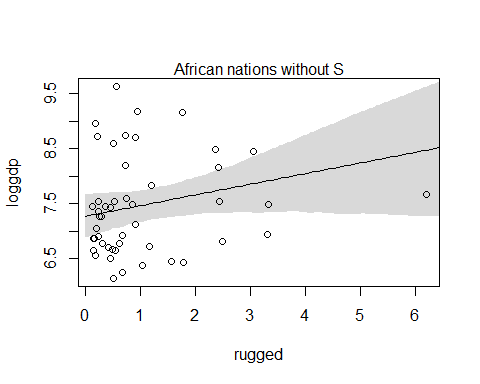
mu.NotAfrica.mean <- apply(X = mu.NotAfrica, MARGIN = 2, FUN = mean)  
mu.NotAfrica.PI <- apply(X = mu.NotAfrica, MARGIN = 2, FUN = PI, prob = 0.97)  
mu.Africa.mean <- apply(X = mu.Africa, MARGIN = 2, FUN = mean)  
mu.Africa.PI <- apply(X = mu.Africa, MARGIN = 2, FUN = PI, prob = 0.97)  
  
  
#with Seychelles,plot Africa regression  
d.A1 <- d[ d$cont\_africa==1 , ]   
d.A0 <- d[ d$cont\_africa==0 , ]  
  
plot(loggdp ~ rugged, data=d.A1)  
lines( ruggedSeq , mu.Africa.mean )  
shade(object = mu.Africa.PI,lim = ruggedSeq)  
mtext( "African nations with S" , 3 )



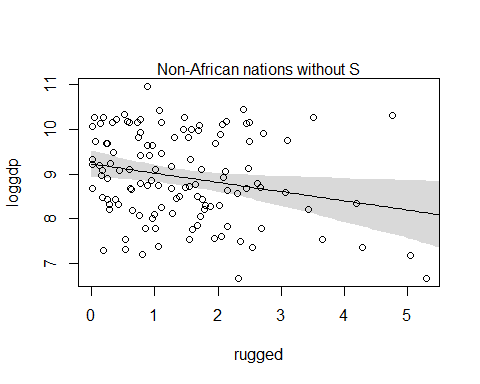
plot(loggdp ~ rugged, data=d.A0)  
lines( ruggedSeq , mu.NotAfrica.mean )  
shade(object = mu.NotAfrica.PI,lim = ruggedSeq)  
mtext( "Non-African nations with S" , 3 )



#without Seychelles,plot africa regression  
dF.A1 <- dF[ dF$cont\_africa==1 , ]   
dF.A0 <- dF[ dF$cont\_africa==0 , ]  
  
plot(loggdp ~ rugged, data=dF.A1)  
lines( ruggedSeq , mu.Africa.mean )  
shade(object = mu.Africa.PI,lim = ruggedSeq)  
mtext( "African nations without S" , 3 )



plot(loggdp ~ rugged, data=dF.A0)  
lines( ruggedSeq , mu.NotAfrica.mean )  
shade(object = mu.NotAfrica.PI,lim = ruggedSeq)  
mtext( "Non-African nations without S" , 3)



# africa的rugged斜率為正，表rugged越大gdp越大。non-africa的rugged斜率為負，表rugged越大gdp越小。

# 而且其實Non-African nations with S = Non african nations without S 因為non african nation本來就沒有Seychelles國家

# 由Africa nations with S 與without S看的出來其實變化很小，斜率只變大一點點，因為Seychelles只佔1/233筆資料

# (c)

modelc1 <- map(  
 alist(  
 loggdp ~ dnorm (mu,sigma),  
 mu <- a + bR \* rugged ,  
 a ~ dnorm (0,10),  
 bR ~ dnorm(0,10),  
 sigma ~ dunif(0,30)  
   
 ),  
 data = dF  
)  
  
modelc2 <- map(  
 alist(  
 loggdp ~ dnorm (mu,sigma),  
 mu <- a + bA \* cont\_africa +bR \* rugged ,  
 a ~ dnorm (0,10),  
 bA ~ dnorm(0,10),  
 bR ~ dnorm(0,10),  
 sigma ~ dunif(0,30)  
   
 ),  
 data = dF  
)  
  
modelc3 <- map(  
 alist(  
 loggdp ~ dnorm (mu,sigma),  
 mu <- a + bA \* cont\_africa +bR \* rugged +bAR \* cont\_africa \* rugged,  
 a ~ dnorm (0,10),  
 bA ~ dnorm(0,10),  
 bR ~ dnorm(0,10),  
 bAR ~ dnorm(0,10),  
 sigma ~ dunif(0,30)  
   
 ),  
 data = dF  
)  
  
compare(modelc1,modelc2,modelc3)

## WAIC pWAIC dWAIC weight SE dSE  
## modelc3 463.5 4.8 0.0 0.81 15.42 NA  
## modelc2 466.4 4.1 2.9 0.19 14.36 4.05  
## modelc1 536.1 2.7 72.6 0.00 13.35 16.00

pdata <- data.frame(rugged = ruggedSeq, cont\_africa = 1)  
mu.ensemble <- ensemble(modelc1,modelc2,modelc3,data = pdata)

## Constructing posterior predictions

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

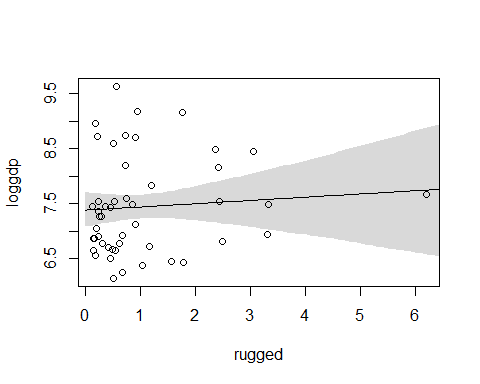
## Constructing posterior predictions

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

## Constructing posterior predictions

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

mu.mean <- apply(X = mu.ensemble$link, MARGIN = 2, FUN = mean)  
mu.PI <- apply(X = mu.ensemble$link, MARGIN = 2, FUN = PI)  
plot(loggdp ~ rugged, data=dF.A1)  
lines( ruggedSeq, mu.mean )  
#lines( ruggedSeq, mu.PI[1,], lty=2 )  
#lines( ruggedSeq, mu.PI[2,], lty=2 )  
shade(object = mu.PI,lim = ruggedSeq)



# 可發現斜率幾乎等於零，rugged與loggdp變成幾乎沒關係

# (b)小題使用的是完整modelc3模型，也就是納入交互作用模型。但此c小題只使用80%的交互作用模型，20%沒有交互作用

# 因此讓交互作用比較沒效果，由a小題可以知道africa其實會讓rugged對gdp有正向影響，但是如果為non-africa則讓rugged對gdp有負向影響

# c小題這邊把continent africa固定為1，即只看africa，然後又消除一些交互作用，此交互作用(只要是africa，就會讓rugged對gdp有正向影響) 的效果被消除了一些，因此斜率變小。

### Q2

data(nettle)  
d <- nettle  
d$lang.per.cap <- (d$num.lang / d$k.pop)  
d$loglpc <- log(d$lang.per.cap)

# (a)

d$logarea <- log(d$area)  
  
model2A <- map(  
 alist(  
 loglpc ~ dnorm(mu,sigma),  
 mu <- a + bA \* logarea + bMGS \* mean.growing.season,  
 a ~ dnorm(0,10),  
 bA ~ dnorm(0,5),  
 bMGS ~ dnorm(5,5), # because growing season與lang per cap為正向關係(依題意)  
 sigma ~ dunif(0,2)  
   
 ),data = d  
 )  
  
precis(model2A)

## Mean StdDev 5.5% 94.5%  
## a -3.73 1.93 -6.81 -0.65  
## bA -0.21 0.13 -0.43 0.00  
## bMGS 0.14 0.06 0.05 0.23  
## sigma 1.39 0.11 1.21 1.57

#看的出來mean growing season之係數(bMGS)為正，且95%信賴區間落在0.05~0.23均大於零，有很明顯的正向關係

# (b)

model2B <- map(  
 alist(  
 loglpc ~ dnorm(mu,sigma),  
 mu <- a + bA \* logarea + bSGS \* sd.growing.season,  
 a ~ dnorm(0,10),  
 bA ~ dnorm(0,5),  
 bSGS ~ dnorm(-5,5), # because standard deviation of growing season與lang per cap為負向關係(依題意)  
 sigma ~ dunif(0,2)  
   
 ),data = d  
)  
precis(model2B)

## Mean StdDev 5.5% 94.5%  
## a -1.96 1.85 -4.91 0.99  
## bA -0.24 0.15 -0.49 0.00  
## bSGS -0.21 0.19 -0.51 0.08  
## sigma 1.44 0.12 1.25 1.63

# 看的出來sd growing season之係數(bSGS)為負，且95%信賴區間落在0.08 ~ -0.51，有還算明顯的負向關係

# (c)

model2C <- map(  
 alist(  
 loglpc ~ dnorm(mu,sigma),  
 mu <- a + bA \* logarea + bMGS \* mean.growing.season + bSGS \* sd.growing.season +bI \*mean.growing.season\*sd.growing.season,  
 a ~ dnorm(0,10),  
 bA ~ dnorm(0,5),  
 bMGS ~ dnorm(5,5),  
 bSGS ~ dnorm(-5,5), # because standard deviation of growing season與lang per cap為負向關係(依題意)  
 bI ~ dnorm(-2,2), #交互作用為負相關(依題意)，但我認為影響沒有mgs sgs大  
 sigma ~ dunif(0,2)  
   
 ),data = d  
)  
precis(model2C)

## Mean StdDev 5.5% 94.5%  
## a -6.63 2.09 -9.98 -3.29  
## bA -0.02 0.16 -0.27 0.23  
## bMGS 0.29 0.08 0.17 0.41  
## bSGS 0.41 0.38 -0.21 1.02  
## bI -0.11 0.05 -0.18 -0.03  
## sigma 1.31 0.11 1.14 1.48

# bI斜率為-0.11且95%信賴區間為-0.03 ~ -0.18皆小於零 ， 可見當growing season變大時，higher variance將使社交需求加劇，導致更需要語言上的統一，因此更小loglpc(log language per capita)