STA 442 Hw2 Q2

Code

2.

Hide

```
dataDir = "../data"
smokeFile = file.path(dataDir, "smoke2014.RData")
if (!file.exists(smokeFile)) {
download.file("http://pbrown.ca/teaching/appliedstats/data/smoke2014.RData", smokeFile)
}
```

```
URL http://pbrown.ca/teaching/appliedstats/data/smoke2014.RData: cannot open destfile
'../data/smoke2014.RData', reason 'No such file or directory'download had nonzero exit s
tatus
```

Hide

```
smoke[1:3,c('Age','ever_cigarettes','Sex','Race',
        'state','school', 'RuralUrban')]
```

	A...	ever_cigarettes	Sex	Race	state	school	RuralUrban
	<dbl>	<lgl>	<fctr>	<fctr>	<fctr>	<chr>	<fctr>
1	18	TRUE	F	white	AL	mdr_00013045	Rural
2	18	TRUE	M	pacific	AL	mdr_00013045	Rural
3	16	TRUE	M	white	AL	mdr_00013045	Rural

3 rows

Hide

```
forInla = smoke[smoke$Age > 10, c("Age", "ever_cigarettes", "Sex", "Race", "state", "sch
ool", "RuralUrban", "Harm_belief_of_chewing_to")]
forInla = na.omit(forInla)
forInla$y = as.numeric(forInla$ever_cigarettes)
forInla$ageFac = factor(as.numeric(as.character(forInla$Age)))
forInla$chewingHarm = factor(forInla$Harm_belief_of_chewing_to,levels = 1:4, labels = c(
"less", "equal", "more", "dunno"))
library("INLA")
toPredict = expand.grid(ageFac = levels(forInla$ageFac),
RuralUrban = levels(forInla$RuralUrban), chewingHarm = levels(forInla$chewingHarm),
Sex = levels(forInla$Sex))
forLincombs = do.call(inla.make.lincombs, as.data.frame(model.matrix(~Sex +ageFac * Rura
lUrban * chewingHarm, data = toPredict)))
fitS2 = inla(y ~ Sex + ageFac * RuralUrban * chewingHarm +f(state, model = "iid", hyper
 = list(prec = list(prior = "pc.prec", param = c(99, 0.05)))), data = forInla, family =
"binomial",control.inla = list(strategy = "gaussian"), lincomb = forLincombs)
```

```
problem writing to connection
```

Hide

```
rbind(fitS2$summary.fixed[, c("mean", "0.025quant", "0.975quant")], Pmisc::priorPostSd(fitS2)$summary[, c("mean", "0.025quant", "0.975quant")])
```

	mean <dbl>	0.025quant <dbl>	0.975quant <dbl>
(Intercept)	-1.974674198	-4.07097020	0.11988733
SexF	-0.125999699	-0.20135859	-0.05070436
ageFac12	0.871242579	-1.32479569	3.06544781
ageFac13	0.601145790	-1.59195517	2.79240463
ageFac14	1.696001107	-0.45489998	3.84510490
ageFac15	1.846332464	-0.31979134	4.01065014
ageFac16	2.157895581	-0.02793248	4.34190286
ageFac17	2.033205169	-0.13692769	4.20152980
ageFac18	2.023183854	-0.19486055	4.23937974
ageFac19	1.616247593	-1.12538218	4.35558812

1-10 of 74 rows Previous **1** 2 3 4 5 6 … 8 Next

Hide

```
theCoef = exp(fitS2$summary.lincomb.derived[, c("0.5quant", "0.025quant", "0.975quant")])
theCoef = theCoef/(1 + theCoef)
toPredict$Age = as.numeric(as.character(toPredict$ageFac))
toPredict$shiftX = as.numeric(toPredict$chewingHarm)/10
toPredict$x = toPredict$Age + toPredict$shiftX
toPlot = toPredict$Sex == "M" & toPredict$RuralUrban == "Rural"
```

Introduction

We want to investigate these hytpotheses are true such that geographic variation (between states) in the rate of smoking among students is substantially greater than variation amongst schools, and the rural-urban difference are greater than differences between states. Moreover, we will figure out the effect of age on smoking for different races.

Method

Hide

```
fitS1 = inla(y ~Sex + ageFac * RuralUrban * chewingHarm+ f(school, model = "iid", hyper
 = list(prec = list(prior = "pc.prec",param = c(99, 0.5)))) +f(state, model = "iid",hype
r = list(prec = list(prior = "pc.prec", param = c(99, 0.5)))), data = forInla, family =
"binomial",control.inla = list(strategy = "gaussian"), lincomb = forLincombs,verbose = T
RUE)
knitr::kable(rbind(fitS2$summary.fixed[, c("mean", "0.025quant", "0.975quant")]), digits
=3,
  caption = "Summary of coefficients in the model",)
```

Hide

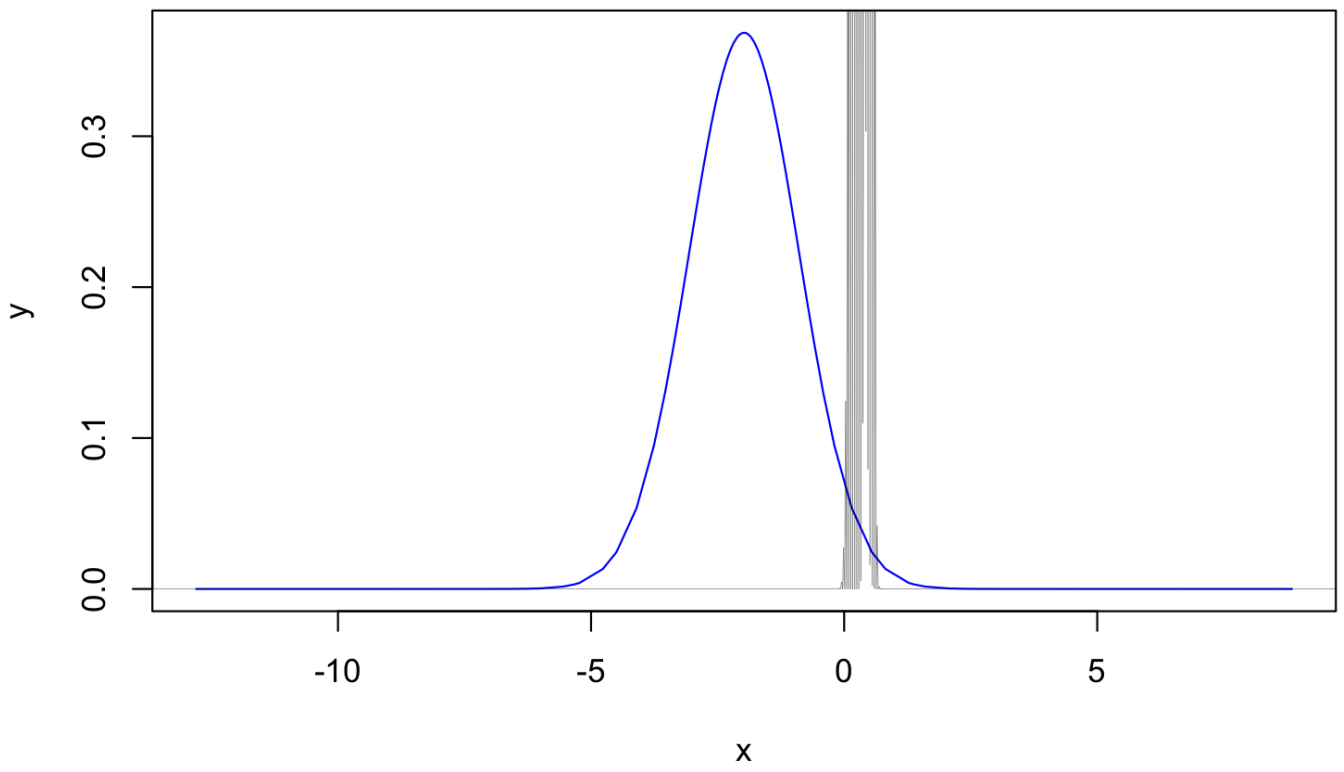
```
fitS1$priorPost<-Pmisc::priorPostSd(fitS1)
fitS1$priorPost$summary
```

	mean <dbl>	sd <dbl>	0.025quant <dbl>	0.5quant <dbl>	0.975quant <dbl>	mode <dbl>
SD for school	0.5031132	0.04488512	0.4230541	0.5001699	0.5991243	0.4918611
SD for state	0.2437120	0.07846945	0.1136908	0.2368795	0.4175086	0.2219188

2 rows

Hide

```
xseq = seq(-100, 100, len=10000)
plot(fitS1$marginals.fixed$'(Intercept)', type='l',col="blue")
lines(xseq, dnorm(xseq, mean=fitS1$priorPost$summary$mean, sd=fitS1$priorPost$summary$s
d),lwd=0.2)
```



Hide

```
forInla = smoke[,c('Age','ever_cigarettes','Sex','Race',
        'state','school', 'RuralUrban')]
forInla = na.omit(forInla)
forInla$y = as.numeric(forInla$ever_cigarettes)
forInla$ageFac = relevel(factor(forInla$Age), '14')

  toPredict = expand.grid(
    ageFac = levels(forInla$ageFac),
    RuralUrban = levels(forInla$RuralUrban),
    Sex = levels(forInla$Sex)
    )
forLincombs = do.call(inla.make.lincombs,
  as.data.frame(model.matrix( ~ ageFac:RuralUrban + Sex,
    data=toPredict)))
fitS2 = inla(y ~ Sex + ageFac:RuralUrban +
    f(state, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(1.1), 0.5)))
    ),
  data=forInla, family='binomial',
  lincomb = forLincombs,
  control.inla = list(strategy='laplace', fast=FALSE))
```

```
problem writing to connection
```

Hide

```
fitS2$summary.hyperpar
```

	mean <dbl>	sd <dbl>	0.025quant <dbl>	0.5quant <dbl>	0.975quant <dbl>	mode <dbl>
Precision for state	6.838206	1.914239	3.763116	6.608638	11.24599	6.181594

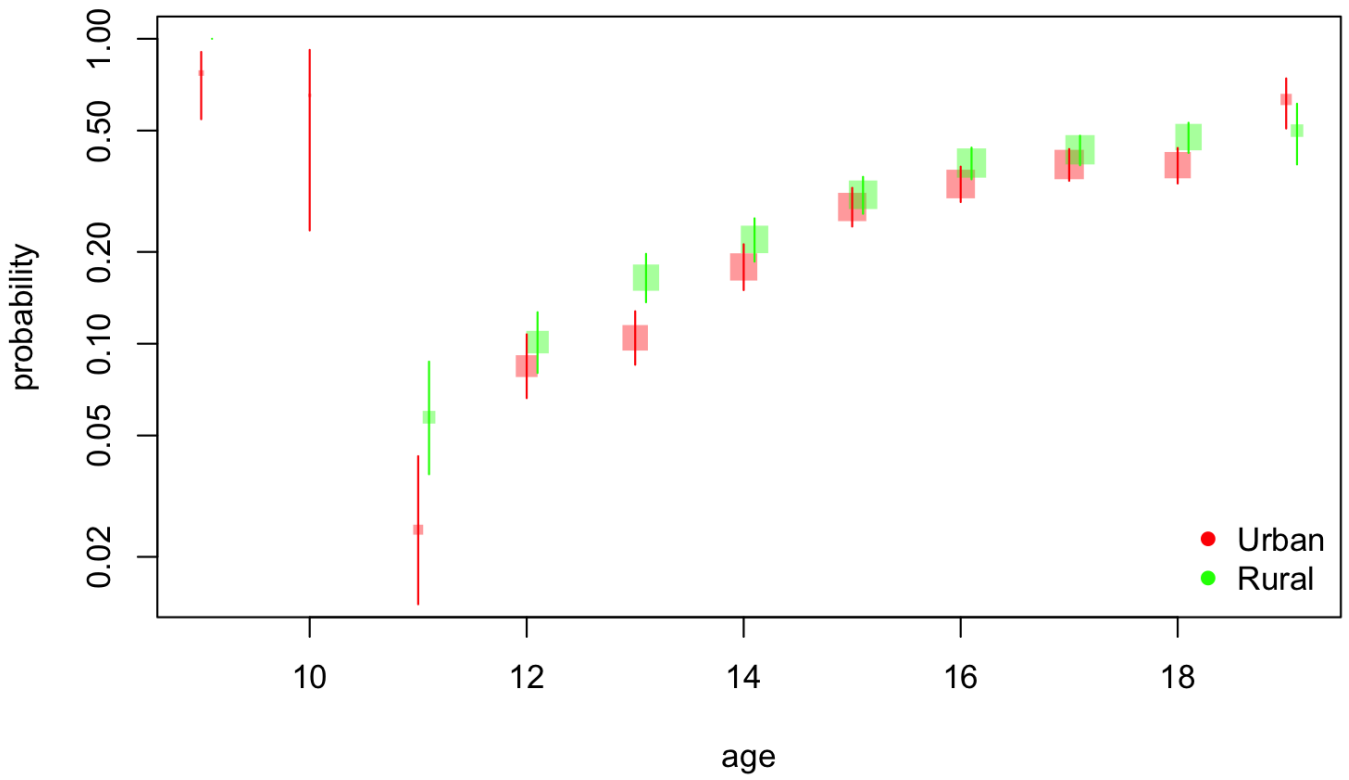
1 row

Hide

```
theSd= Pmisc::priorPost(fitS2)




toPredict$Age = as.numeric(as.character(toPredict$ageFac))

isMale = toPredict$Sex == 'M'
shiftRural = 0.1*(toPredict$RuralUrban == 'Rural')

theSd = fitS4$summary.lincomb.derived[,'sd']
theCex = min(theSd)/theSd

plot(toPredict[isMale,'Age'] + shiftRural[isMale],
  theCoef[isMale,'0.5quant'],
  xlab='age', ylab='probability', ylim = c(0.015, 1),
  pch = 15, log='y',
  cex = 2*theCex,
  col = mapmisc::col2html(
    c(Urban = 'red', Rural = 'green')[as.character(toPredict[isMale,'RuralUrban'])],
    0.4)
  ,cap = "effect of Age")


segments(toPredict[isMale,'Age']+ shiftRural[isMale],
  theCoef[isMale,'0.025quant'],
  y1=theCoef[isMale,'0.975quant'],
  col = c(Urban = 'red', Rural = 'green')[as.character(toPredict[isMale,'RuralUrban'])])
```

Hide

```
legend('bottomright', pch=16, col=c('red','green'), legend = c('Urban','Rural'),
  bty='n')
```

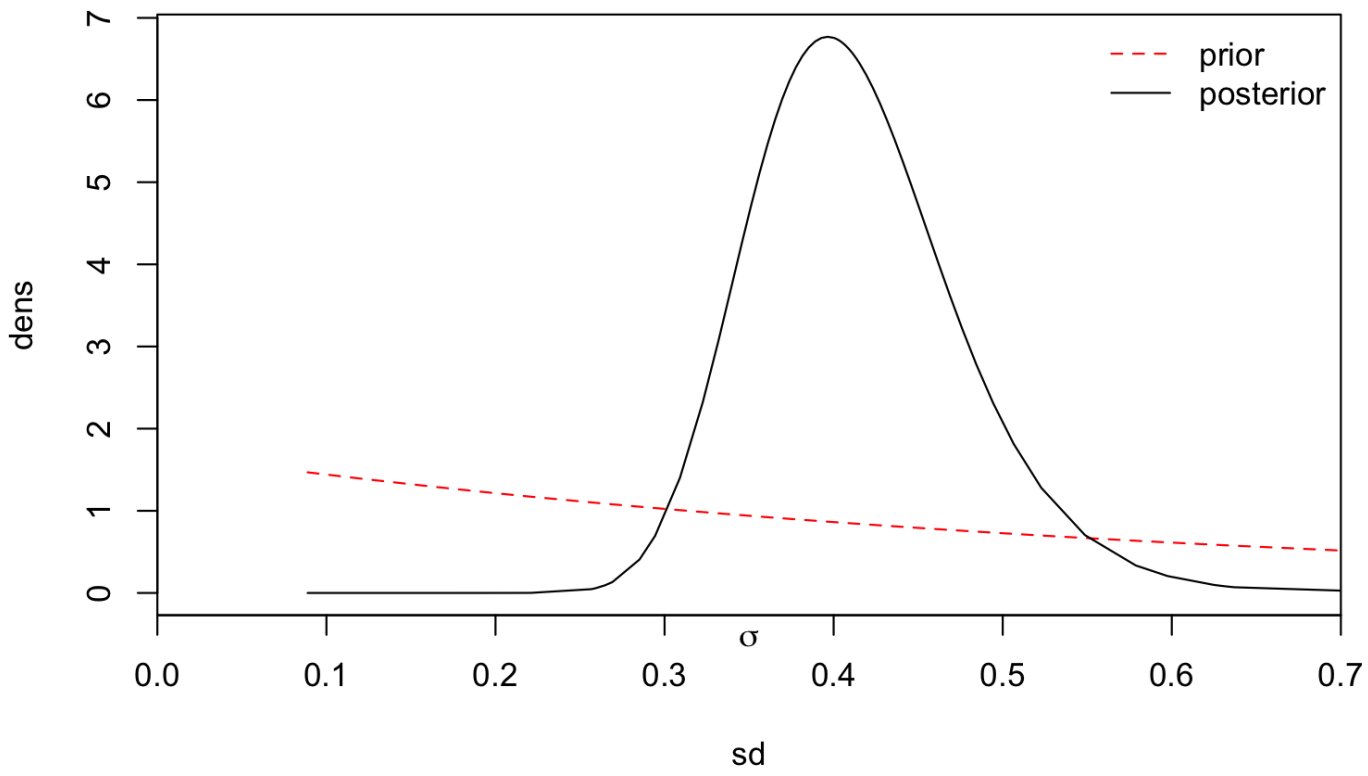


Hide

```
do.call(matplot, theSd4$sd$matplot)
do.call(legend, theSd4$sd$legend)
```

Hide

```
mtext(expression(sigma), side=1)
```



Conclusion

From standard deviation comparison between random effects of state and school the effect of school influences on more about the rate of smoking so that we need to specify the school where students with higher rate of smoking and supply the tobacco control program on that school. The effect of age is stronger than the ruralUrban effect.