Question 1: Math

In this report, I want to investigate the effects of the factors Minority, SES and School on the Mathematical Achievement Scores of 7185 students from 160 different schools. It has been predetermined that the factors Minority and variables SES are fixed effects. I want to see if I should treat School as a random effect or not, by checking if there are substantial differences within schools, or the differences within schools is nearly as big as the differences between students from different schools.

$$MathAch_{ij} = \beta_0 + \beta_1 I_{MinorityYes} + \beta_2 SES_j + b_i + e_{ij}$$

Figure 1. Fitted Mixed Effect Model

I first fit a model (Figure 1), where the random effect bi is assumed to follow a normal distribution with mean 0 and variances σ_b^2 independent of the error term e_{ij} which is iid with mean 0 and variances σ_e^2 .

	Value	Std.Error	t-value	p-value
(Intercept)	13.465962	0.1822278	73.89631	0
MinorityYes	-2.938167	0.2070407	-14.19126	0
SES	2.127597	0.1060196	20.06795	0

Figure 2. Output for Fixed Effect in Mixed Effect Models

Looking at Figure 2, the p-values for the factors Minority and SES are less than 0.05, meaning that they are both statistically significant, thus again confirmed that these two are significant fixed factors in this model. The estimates for these factors and the intercept will remain the same, regardless of the School variable.

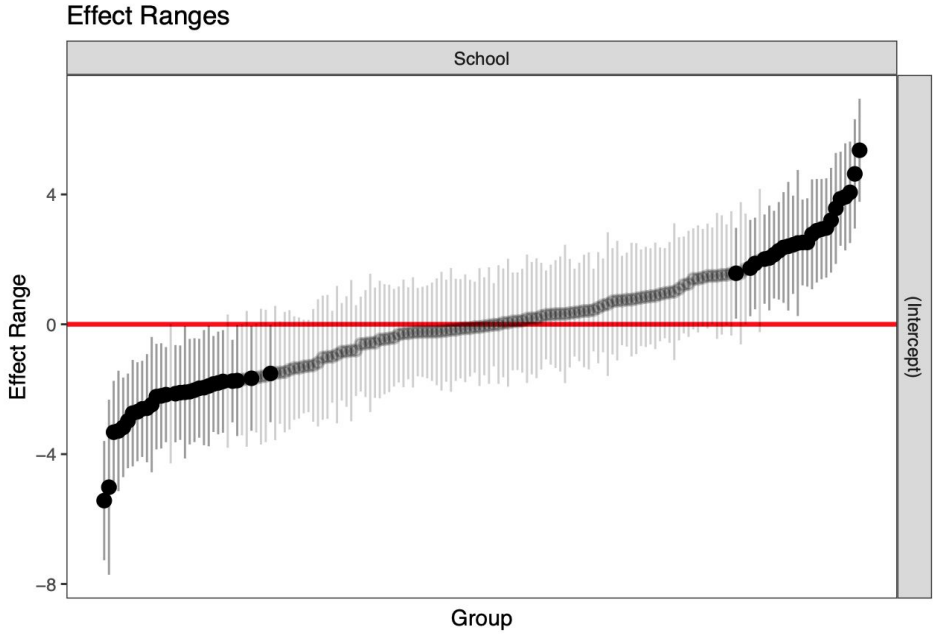


Figure 3. Estimates of Random Effects and their intervals

Before seeing the implication of random effects, I wanted to get estimates of them. In Figure 3, I have plotted the estimated random effects and interval estimates for each school. Knowing that random effects are on average zero, it is interesting to see so many intervals that do not include zero.

	Variance	StdDev
(Intercept)	3.93507	1.983701
Residual	36.14816	6.012334

Figure 4. Output for Random Effects in Mixed Effect Model

Then, I looked into the implication of random effects in Figure 4. The error variance is estimated to be 36.148 while the variance due to the random intercept is estimated to be 3.935. From these two values, I can find the correlation of two measurements within the same school is estimated to be 3.935/(3.935+36.148) = 0.0982 (estimate of within-cluster correlation). When σ_b^2 is relatively small and σ_e^2 is relatively large, we can say that there is a "shrinkage". This means that there is less school-level variance relative to the population variance, and thus the mixed model will produce a group-specific effect that is closer to the overall population effect.

The differences among schools only attributed 9.82% of the variability in Mathematical Achievements Scores. It seems that the differences within the school are nearly as big as the difference between students from different schools. And, it seems unnecessary to treat the School variable as a random effect.

Question 2: Drugs

Introduction

I will be working with annual data on the number and characteristics of persons discharged from public and private substance abuse treatment programs that receive public funding, provided by the TEDS-D. There are two hypotheses I want to investigate in this report. First, I want to see if it is more difficult to treat young persons' addictions to 'hard' drugs (Heroin, Opiates, Methamphetamine, Cocaine) than alcohol or marijuana. Second, I want to see if some American states have significantly more effective treatment programs than others, by comparing the difference in completion rates of the states.

Methods

I will fit a generalized linear mixed model with binomial likelihood using Bayesian inference. The data concerns the likelihood of the subjects in question completing their drug treatments. For the fixed effect variables in the model, I have included the subjects' primary substances of addiction; the subjects' age, gender and race, which are known to be important confounders. For random effects, I have included the subjects' US states and towns in which the treatments are given, to model within-group variations.

In the model, there are unknown parameters $\boldsymbol{\alpha}$,σ_u^2,σ_e^2 for which we need to specify prior distributions. I have chosen to use the penalized complexity prior, which puts an exponential prior to the standard deviation σ_u^2,σ_e^2 and requires me to calibrate the scaling of random effect priors. For both random effect variables (STFIPS and TOWN), I think there is a 99% chance that the between states/town variability is greater than 3 times the residual SD of the model. Hence I set p = 0.001 and U = 3*SD(y).

Results

	0.5quant	0.025quant	0.975quant
(Intercept)			
(Intercept)	0.681	0.546	0.849
SUB1			
ALCOHOL	1.642	1.608	1.677
HEROIN	0.898	0.875	0.921
OTHER OPIATES AND SYNTHET	0.924	0.897	0.952
METHAMPHETAMINE	0.982	0.944	1.022
COCAINE/CRACK	0.876	0.834	0.920
GENDER			
FEMALE	0.895	0.880	0.910
raceEthnicity			
Hispanic	0.829	0.810	0.849
BLACK OR AFRICAN AMERICAN	0.685	0.669	0.702
AMERICAN INDIAN (OTHER TH	0.729	0.680	0.782
OTHER SINGLE RACE	0.864	0.810	0.921
TWO OR MORE RACES	0.851	0.790	0.917
ASIAN	1.133	1.038	1.236
NATIVE HAWAIIAN OR OTHER	0.846	0.750	0.955
ASIAN OR PACIFIC ISLANDER	1.451	1.225	1.720
ALASKA NATIVE (ALEUT, ESK	0.844	0.623	1.144
homeless			
TRUE	1.015	0.983	1.048
SD			
STFIPS	0.693	0.558	0.871
TOWN	0.539	0.485	0.601

Figure 5. Posterior means and quantiles for model parameters

In Figure 5, I can examine the posterior means and credible intervals of the parameters. From the table, the odds of subjects completing drug treatments for ‘hard’ drugs, such as heroin, methamphetamine and cocaine are approximately 0.9 times, the odds of that for marijuana. Whereas the odds of subjects completing drug treatments for alcohol is approximately 1.6 times, the odds of that for marijuana.

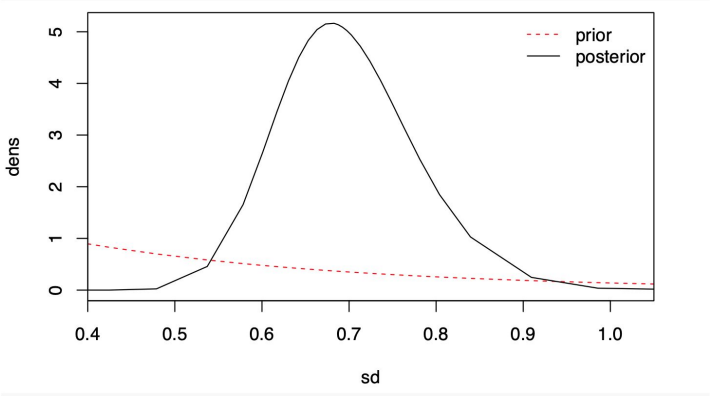


Figure 6. Posterior Distribution of SDs for Random Effects by STFIPS

In Figures 5 and 6, I have found the values and the posterior distribution of the standard deviations for random effects by the variable STFIPS. The value is 0.693 and it has a credible interval of 0.558 to 0.887.

ID	mean	0.025q	0.975q	ID	mean	0.025q	0.975q
ALABAMA	0.2	-0.3	0.8	MONTANA	-0.2	-1.0	0.7
ALASKA	0.0	-0.9	0.8	NEBRASKA	0.8	0.4	1.2
ARIZONA	0.0	-1.3	1.3	NEVADA	-0.1	-0.8	0.6
ARKANSAS	-0.1	-0.7	0.5	NEW HAMPSHIRE	0.2	-0.3	0.7
CALIFORNIA	-0.3	-0.6	0.0	NEW JERSEY	0.5	0.2	0.8
COLORADO	0.5	0.1	1.0	NEW MEXICO	-1.2	-1.9	-0.5
CONNECTICUT	0.1	-0.4	0.7	NEW YORK	-0.3	-0.6	0.0
DELAWARE	1.0	0.7	1.3	NORTH CAROLINA	-0.8	-1.1	-0.5
WASHINGTON DC	-0.3	-0.6	0.1	NORTH DAKOTA	-0.3	-1.0	0.4
FLORIDA	1.0	0.7	1.4	OHIO	-0.2	-0.6	0.1
GEORGIA	-0.2	-0.9	0.4	OKLAHOMA	0.6	0.0	1.1
HAWAII	0.2	-0.6	1.1	OREGON	0.1	-0.3	0.5
IDAHO	-0.2	-1.0	0.6	PENNSYLVANIA	0.0	-1.3	1.3
ILLINOIS	-0.5	-0.8	-0.2	RHODE ISLAND	-0.2	-0.6	0.3
INDIANA	-0.1	-0.9	0.8	SOUTH CAROLINA	0.4	0.0	0.7
IOWA	0.4	0.1	0.7	SOUTH DAKOTA	0.5	-0.3	1.3
KANSAS	-0.2	-0.6	0.1	TENNESSEE	0.3	-0.2	0.7
KENTUCKY	-0.1	-0.5	0.2	TEXAS	0.6	0.3	0.9
LOUISIANA	-0.6	-1.0	-0.1	UTAH	0.1	-0.5	0.7
MAINE	0.1	-0.7	1.0	VERMONT	-0.2	-1.1	0.6
MARYLAND	0.5	0.2	0.8	VIRGINIA	-2.9	-3.3	-2.5
MASSACHUSETTS	0.8	0.4	1.2	WASHINGTON	-0.1	-0.5	0.3
MICHIGAN	-0.4	-0.7	0.0	WEST VIRGINIA	0.0	-1.3	1.3
MINNESOTA	0.4	0.0	0.9	WISCONSIN	0.0	-1.3	1.3
MISSISSIPPI	0.0	-1.3	1.3	WYOMING	0.0	-1.3	1.3
MISSOURI	-0.4	-0.7	-0.1	PUERTO RICO	0.6	-0.1	1.3

Figure 7. Posterior means and quantiles for random effects by states

Looking at the posterior means for the random effects of the different states, if the treatment is given in Massachusetts or Nebraska, the treatment completion rate will be higher. States like Alabama, Colorado, Texas and South Dakota also have a fairly positive effect on the

treatment completion rate. Whereas states like Alaska, Arizona and Pennsylvania have no effect on the treatment completion rate. On the other hand, if a treatment is given in states like Virginia, Vermont and Washington DC, the treatment completion rate will be decreased.

Conclusions

In response to the first hypothesis, the chance of a young person completing their drug treatment is higher when the substance the individual is addicted to is either alcohol or marijuana**.** 'Hard' drugs such as Heroin, Opiates, Methamphetamine and Cocaine will be more difficult to treat.

In response to the second hypothesis, when we are comparing the posterior means of the random effect of the states, it does seem like some states have a positive effect on the odds of its treatment completion rate, whereas some states have none to negative effects. Therefore, we can conclude that some US states, such as Massachusetts and Nebraska, will have particularly effective treatment programs. Whereas states like Virginia and Vermont may have programs that are highly problematic and ineffective.

STA442 HW2 Appedix

Question 1: Math

```
data("MathAchieve",package = "MEMSS")
```

```
library("lme4")
```

```
## Loading required package: Matrix
```

```
library("nlme")
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     MathAchieve
```

```
## The following object is masked from 'package:lme4':
##
##     lmList
```

```
library("tidyverse")
```

```
## -- Attaching packages ------------------------------------------------------------------------ tidyv
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts --------------------------------------------------------------------------- tidyverse_c
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x tidyr::pack()     masks Matrix::pack()
## x tidyr::unpack()   masks Matrix::unpack()
```

```
library("merTools")
```

```
## Loading required package: arm
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

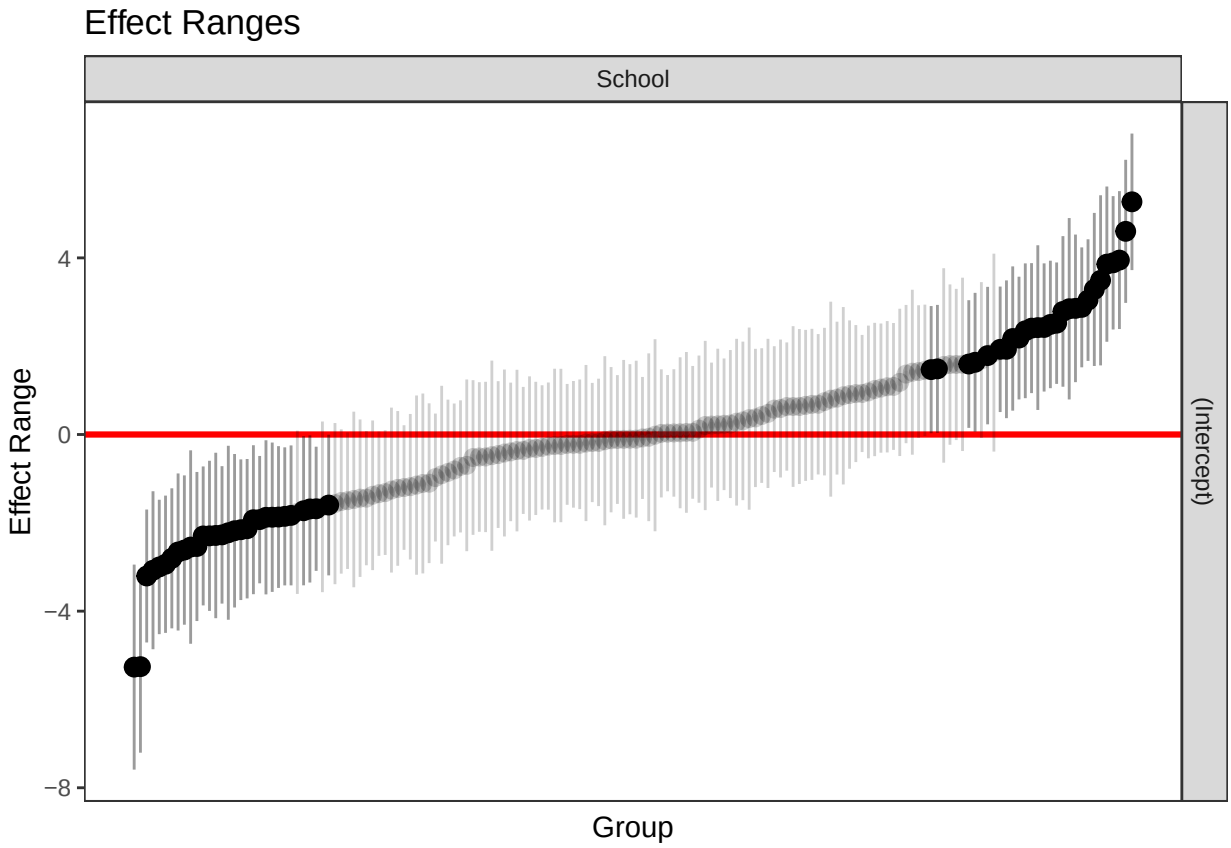
```
##
## arm (Version 1.10-1, built: 2018-4-12)
```

```
## Working directory is /Users/JennaFu/Desktop/2019-2020/STA442/Homework 2
```

```
MathLme = nlme::lme(MathAch ~ Minority + SES, random = ~ 1|School ,data = MathAchieve)
# Figure 2: Output for Fixed Effect in Mixed Effect Models
knitr::kable(summary(MathLme)$tTable[,-3],digital=3)
```

	Value	Std.Error	t-value	p-value
(Intercept)	13.465962	0.1822278	73.89631	0
MinorityYes	-2.938167	0.2070407	-14.19126	0
SES	2.127597	0.1060196	20.06795	0

```
# Figure 3: Estimates of Random Effects and their intervals
MathLme2 = lme4::lmer(MathAch ~ Minority + SES +(1|School), data = MathAchieve)
plotREsim(REsim(MathLme2))
```




```
# Figure 4: Output for Random Effects in Mixed Effect Model
knitr::kable(VarCorr(MathLme))
```

	Variance	StdDev
(Intercept)	3.93507	1.983701
Residual	36.14816	6.012334

Question 2: Drugs

```
library("INLA")
```

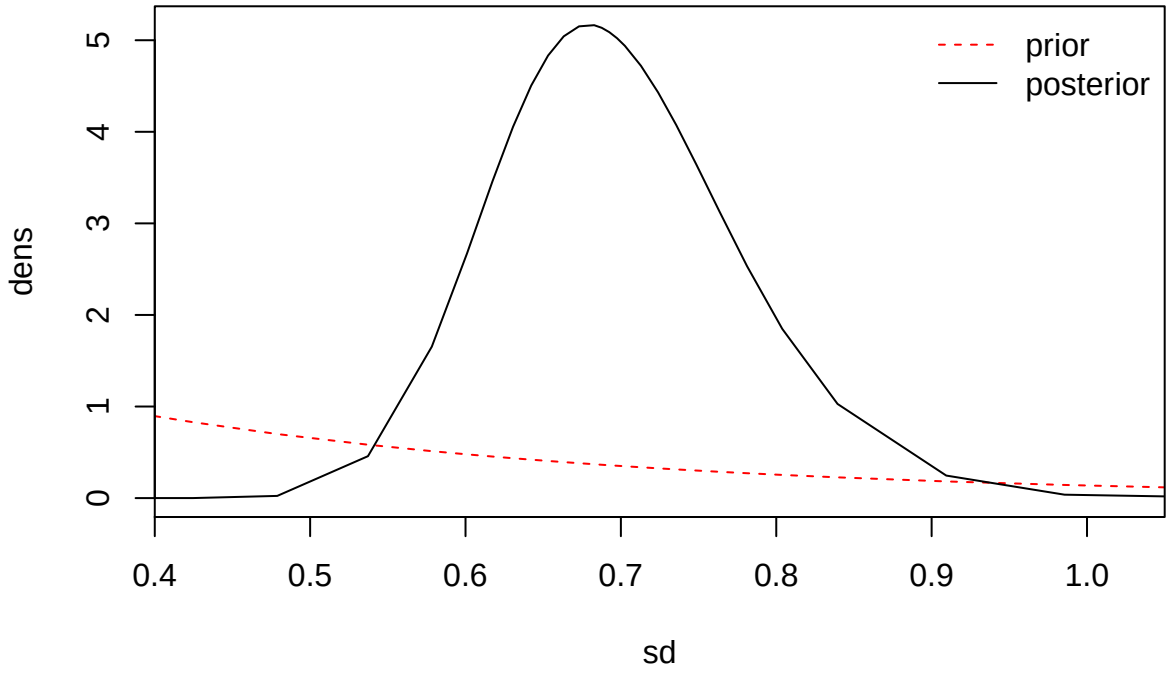
```
## Loading required package: sp
```

```
## Loading required package: parallel
```

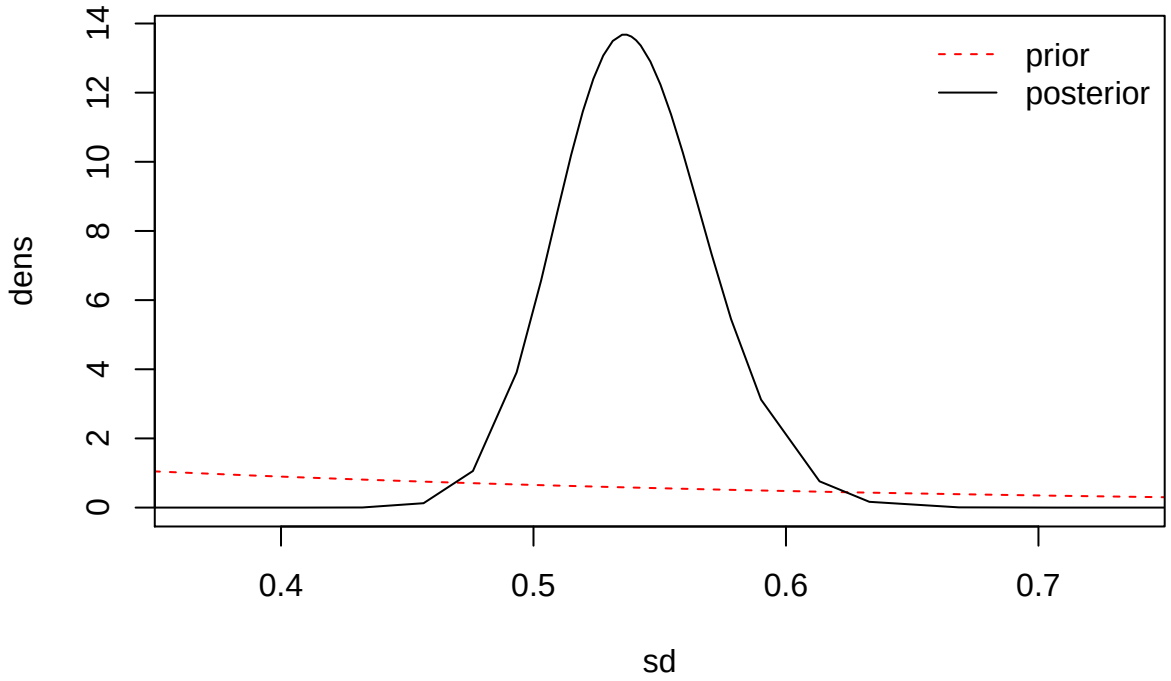
```
## This is INLA_19.09.03 built 2019-09-03 09:07:31 UTC.
## See www.r-inla.org/contact-us for how to get help.
## To enable PARDISO sparse library; see inla.pardiso()
```

```
library("Pmisc")

download.file("http://pbrown.ca/teaching/appliedstats/data/drugs.rds", "drugs.rds")
xSub = readRDS("drugs.rds")
forInla = na.omit(xSub)
forInla$y = as.numeric(forInla$completed)
sdires <- sd(forInla$y)

ires = inla(y ~ SUB1 + GENDER + raceEthnicity + homeless
            + f(STFIPS, hyper=list(prec=list( prior='pc.prec', param=c(3*sdires, 0.01))))
            + f(TOWN, hyper=list(prec=list( prior='pc.prec', param=c(3*sdires, 0.01)))),
          data=forInla, family="binomial",
          control.inla = list(strategy='gaussian', int.strategy='eb'))
```

```
sdState = Pmisc::priorPostSd(ires)
do.call(matplot, sdState$STFIPS$matplot)
do.call(legend, sdState$legend)
```




```
do.call(matplot, sdState$TOWN$matplot)
do.call(legend, sdState$legend)
```




```
# install.packages("data.table", type = "binary")

toPrint = as.data.frame(rbind(exp(ires$summary.fixed[, c(4, 3, 5)]),
                              sdState$summary[, c(4, 3, 5)]))
sss = "^(raceEthnicity|SUB1|GENDER|homeless|SD)(.[[:digit:]]+.[[:space:]]+| for )?"
toPrint = cbind(variable = gsub(paste0(sss, ".*"),
"\\1", rownames(toPrint)), category = substr(gsub(sss,
"", rownames(toPrint)), 1, 25), toPrint)

Pmisc::mdTable(toPrint, digits = 3, mdToTex = TRUE,
               guessGroup = TRUE,
               caption = "Posterior means and quantiles for model parameters.")
```

```
ires$summary.random$STFIPS$ID = gsub("[[:punct:]]|[[:digit:]]", "",
                                     ires$summary.random$STFIPS$ID)
ires$summary.random$STFIPS$ID = gsub("DISTRICT OF COLUMBIA", "WASHINGTON DC",
                                     ires$summary.random$STFIPS$ID)
toprint = cbind(ires$summary.random$STFIPS[1:26, c(1, 2, 4, 6)],
                ires$summary.random$STFIPS[-(1:26), c(1, 2, 4, 6)])
colnames(toprint) = gsub("uant", "", colnames(toprint))
knitr::kable(toprint, digits = 1, format = "latex")
```

Table 3: Posterior means and quantiles for model parameters.

	0.5quant	0.025quant	0.975quant
(Intercept)			
(Intercept)	0.681	0.546	0.849
SUB1			
ALCOHOL	1.642	1.608	1.677
HEROIN	0.898	0.875	0.921
OTHER OPIATES AND SYNTHET	0.924	0.897	0.952
METHAMPHETAMINE	0.982	0.944	1.022
COCAINE/CRACK	0.876	0.834	0.920
GENDER			
FEMALE	0.895	0.880	0.910
raceEthnicity			
Hispanic	0.829	0.810	0.849
BLACK OR AFRICAN AMERICAN	0.685	0.669	0.702
AMERICAN INDIAN (OTHER TH	0.729	0.680	0.782
OTHER SINGLE RACE	0.864	0.810	0.921
TWO OR MORE RACES	0.851	0.790	0.917
ASIAN	1.133	1.038	1.236
NATIVE HAWAIIAN OR OTHER	0.846	0.750	0.955
ASIAN OR PACIFIC ISLANDER	1.451	1.225	1.720
ALASKA NATIVE (ALEUT, ESK	0.844	0.623	1.144
homeless			
TRUE	1.015	0.983	1.048
SD			
STFIPS	0.693	0.558	0.871
TOWN	0.539	0.485	0.601

ID	mean	0.025q	0.975q	ID	mean	0.025q	0.975q
ALABAMA	0.2	-0.3	0.8	MONTANA	-0.2	-1.0	0.7
ALASKA	0.0	-0.9	0.8	NEBRASKA	0.8	0.4	1.2
ARIZONA	0.0	-1.3	1.3	NEVADA	-0.1	-0.8	0.6
ARKANSAS	-0.1	-0.7	0.5	NEW HAMPSHIRE	0.2	-0.3	0.7
CALIFORNIA	-0.3	-0.6	0.0	NEW JERSEY	0.5	0.2	0.8
COLORADO	0.5	0.1	1.0	NEW MEXICO	-1.2	-1.9	-0.5
CONNECTICUT	0.1	-0.4	0.7	NEW YORK	-0.3	-0.6	0.0
DELAWARE	1.0	0.7	1.3	NORTH CAROLINA	-0.8	-1.1	-0.5
WASHINGTON DC	-0.3	-0.6	0.1	NORTH DAKOTA	-0.3	-1.0	0.4
FLORIDA	1.0	0.7	1.4	OHIO	-0.2	-0.6	0.1
GEORGIA	-0.2	-0.9	0.4	OKLAHOMA	0.6	0.0	1.1
HAWAII	0.2	-0.6	1.1	OREGON	0.1	-0.3	0.5
IDAHO	-0.2	-1.0	0.6	PENNSYLVANIA	0.0	-1.3	1.3
ILLINOIS	-0.5	-0.8	-0.2	RHODE ISLAND	-0.2	-0.6	0.3
INDIANA	-0.1	-0.9	0.8	SOUTH CAROLINA	0.4	0.0	0.7
IOWA	0.4	0.1	0.7	SOUTH DAKOTA	0.5	-0.3	1.3
KANSAS	-0.2	-0.6	0.1	TENNESSEE	0.3	-0.2	0.7
KENTUCKY	-0.1	-0.5	0.2	TEXAS	0.6	0.3	0.9
LOUISIANA	-0.6	-1.0	-0.1	UTAH	0.1	-0.5	0.7
MAINE	0.1	-0.7	1.0	VERMONT	-0.2	-1.1	0.6
MARYLAND	0.5	0.2	0.8	VIRGINIA	-2.9	-3.3	-2.5
MASSACHUSETTS	0.8	0.4	1.2	WASHINGTON	-0.1	-0.5	0.3
MICHIGAN	-0.4	-0.7	0.0	WEST VIRGINIA	0.0	-1.3	1.3
MINNESOTA	0.4	0.0	0.9	WISCONSIN	0.0	-1.3	1.3
MISSISSIPPI	0.0	-1.3	1.3	WYOMING	0.0	-1.3	1.3
MISSOURI	-0.4	-0.7	-0.1	PUERTO RICO	0.6	-0.1	1.3