Homework 2, Mixed effects models

STA 442 Methods of Applied Statistics

Due 16 Oct 2019

Math (10 marks)

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

```
  School Minority    Sex    SES MathAch MEANSES
1   1224       No Female -1.528   5.876  -0.428
2   1224       No Female -0.588  19.708  -0.428
3   1224       No   Male -0.528  20.349  -0.428
4   1224       No   Male -0.668   8.781  -0.428
5   1224       No   Male -0.158  17.898  -0.428
6   1224       No   Male  0.022   4.583  -0.428
```

From Maindonald and Braun, ch 10 q 5. In the data set `MathAchieve` (`MEMSS` package), the factors `Minority` (levels `yes` and `no`), and the variable `SES` (socio-economic status) are clearly fixed effects. Carry out an analysis that treats `School` as a random effect. Does it appear that there are substantial differences between schools, or are differences within schools nearly as big as differences between students from different schools? Write a short report (a single page of text plus a few graphs). **is the inter-school differences significant**

Q3: Drugs (20 marks)

http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/35074

The Treatment Episode Data Set – Discharges (TEDS-D) is a national census data system of annual discharges from substance abuse treatment facilities. TEDS-D provides annual data on the number and characteristics of persons discharged from public and private substance abuse treatment programs that receive public funding.

```
download.file("http://pbrown.ca/teaching/appliedstats/data/drugs.rds",
  "drugs.rds")
xSub = readRDS("drugs.rds")
```

```
table(xSub$SUB1)
```

```
           (4) MARIJUANA/HASHISH                      (2) ALCOHOL
                          188406                            97013
                     (5) HEROIN (7) OTHER OPIATES AND SYNTHETICS
                           58511                            45609
            (10) METHAMPHETAMINE                (3) COCAINE/CRACK
                           21606                            11333
```

```
table(xSub$STFIPS)[1:5]
```

```
   (1) ALABAMA     (2) ALASKA    (4) ARIZONA   (5) ARKANSAS (6) CALIFORNIA
           616           1360           4479           1508          48065
```

```
table(xSub$TOWN)[1:2]
```

```
ABILENE, TX   AKRON, OH
         42        1078
```

Each row of the dataset corresponds to an individual admitted to a drug or alcohol addiction treatment facility. The variables above are:

- `completed` is `TRUE` if the individual in question completed their treatment and `FALSE` otherwise.
- `SUB1` is the substance which was the individual's primary addiction.
- `GENDER`, `AGE`, `raceEthnicity` are the individuals age, gender and ethnicity, known to be important confounders.
- `STFIPS`, `TOWN`, the US state and town in which the treatment was given.

Write a short report addressing the hypothesis that chance of a young person completing their drug treatment depends on the substance the individual is addicted to, with 'hard' drugs (Heroin, Opiates, Methamphetamine, Cocaine) being more difficult to treat than alcohol or marijuana. A secondary hypothesis is that some American states have particularly effective treatment programs whereas other states have programs which are highly problematic with very low completion rates.

The report should be on the order of four paragraphs: introduction, methods, results, conclusions. Not more than two pages of text, closer to one page is better.

Some code below may or may not be helpful.

```
forInla = na.omit(xSub)
forInla$y = as.numeric(forInla$completed)

library("INLA")
ires = inla(y ~ SUB1 + GENDER + raceEthnicity + homeless +
```

```
    f(STFIPS, hyper=list(prec=list(
          prior='pc.prec', param=c(0.1, 0.05)))) +
    f(TOWN),
  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)
```

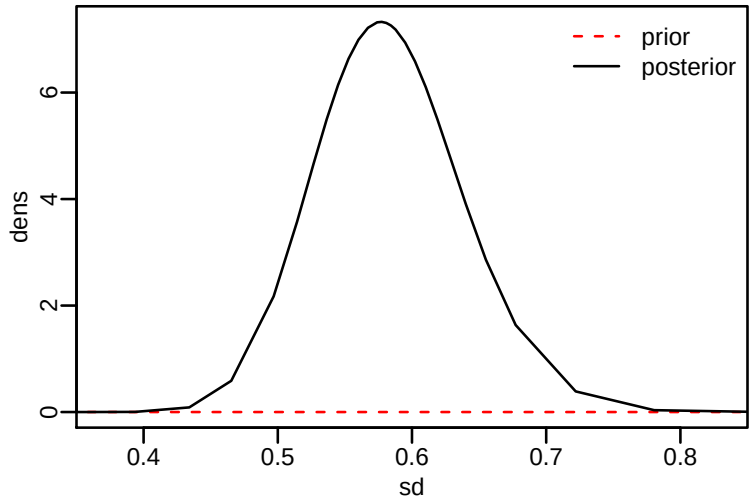

Figure 1: State-level standard deviation

```
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 1: Posterior means and quantiles for model parameters.

	0.5quant	0.025quant	0.975quant
(Intercept)			
(Intercept)	0.682	0.562	0.826
SUB1			
ALCOHOL	1.642	1.608	1.677
HEROIN	0.898	0.875	0.921
OTHER OPIATES AND SYNTHET	0.924	0.898	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	0.730	0.680	0.782
OTHER SINGLE RACE	0.864	0.810	0.920
TWO OR MORE RACES	0.851	0.790	0.917
ASIAN	1.133	1.038	1.236
NATIVE HAWAIIAN OR OTHER	0.847	0.750	0.955
ASIAN OR PACIFIC ISLANDER	1.451	1.225	1.720
ALASKA NATIVE (ALEUT, ESK	0.844	0.623	1.143
homeless			
TRUE	1.015	0.983	1.048
SD			
STFIPS	0.581	0.482	0.698
TOWN	0.537	0.482	0.597

ID	mean	0.025q	0.975q	ID	mean	0.025q	0.975q
ALABAMA	0.2	-0.3	0.7	MONTANA	-0.2	-1.0	0.6
ALASKA	0.0	-0.8	0.8	NEBRASKA	0.8	0.4	1.2
ARIZONA	0.0	-1.1	1.1	NEVADA	-0.1	-0.8	0.5