A2

Yuhan

2019/10/15

Math.

Introduction

Experiment about how different factors affect math achievement of students are performed. Here we are interested in whether differences between schools or within school are substantial.

Methods

Decision about random effect and fixed effect selection is based on aim of the study. Minority, which indicates whether the candidate is a memember of a minority racial group, and socio-economics status as fixed effect. School is selected as random effect since observations obtained from same school are unlikely to be independent, and the research question concerns about the differences in mathematics achievement scores between schools.

For mixed effect model, normality of response variable is assumed. Distribution of MathAch data is shown in Figure1. Though normality assumption is not perfectly met, the distribution is sufficiently good to be regarded as normal distributed.

Result

Fitted model is shown in table 1. P-value of both fixed effects are significant. students from minorty ethic groups are getting math achievement scores 2.94 units lower than others. For each units increase in socio-economics status scores, resultant math achievement scores increase by 2.13 units. To evaluate the research question, evaluate $\sigma^2/(\sigma^2+\tau^2) = 9.8\%$, which is quite small. This small percentage indicated that differences between school are not substantial.

Conclusion

From the fitted mixed effect model, it can be concluded that student math achievement differences between schools is not substantial compared to differences within school.

Q3:Drugs.

Introduction

The Treatment Episode Data Set is a dataset contains information about candidates. By analyzing this dataset, we want to clarify two hypothesis. 1. Chance of a candidate completing drug treatments is affected by type of substances he/she addicted to. More specifically, 'hard' drugs(Heroin, Opiates, Methamphetamine, Cocaine) are more difficult to treat than alcohol or marjuana.

2. Rates of completing treatment is also affected by states. Some states have more effective treatment programs while others may have problematic treatment programs with low completion rates.

Methods

STFIPS and TOWN are chosen as random effect and the other variables besides EDUC and AGE are chosen as fixed effect. A generalized linear model is fitted.

$log(\mu_i/(1-\mu_i)) = X_{it}\beta + U_i + V_i$ where U_i and V_i are individual level random effects, stands for states and town.

Use bayseian inference to find estimate of random effect. In order to do so,a prior distribution is needed. Penalized complexity prior is chosen for this analysis. Since not much information is known about actual distribution. I tried different value set $P(\theta \geq a) \leq 0.05$ for prior distribution. Then checked the area under prior curve of resultant prior-posterior plot. After several attempts, 0.75 is selected as value of a. Then bayes rule $pr(\theta = a|Y = y) = pr(Y = y|\theta = a)pr(\theta = a)/pr(Y = y)$ can be used to compute posterior.

Result

As shown in table 2 only odds of alcohol is greater than 1. Since marijuana is the reference group, estimated value for marijuana is 1. Odds of all other substances groups are smller than 1, which indicated candidates from these 'hard' drugs groups will have a smaller chance to complete treatment.

State is an individual level random effect. Positive value of this random effect means higher average response value. In this case, probability of treatment completion. As the last table shows, there are states with positive estimate and states with negative estimate.

Conclusion

Hypothesis 1 is true, compared to those candidates who are addicted to 'hard' drugs, candidates addicted to alcohol and marijuana is more likely to complete treatment.

Hypothesis 2 is true, effectiveness of treatment programs is vary for different states. Some states has more effective treatment programs while some states have problematic treatment programs that result in low rate of completion.

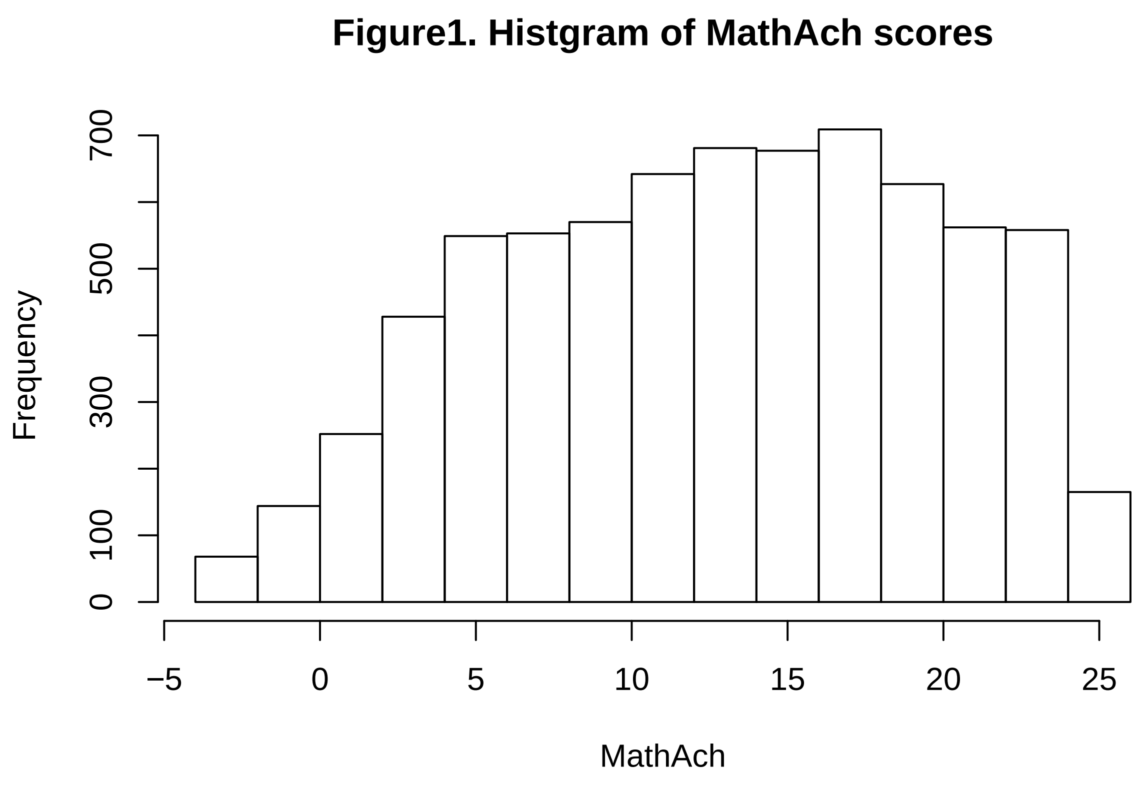
Table 1: Estimation of MathAchieve mixed model

	MLE	Std.Error	DF	t-value	p-value
(Intercept)	13.47	0.18	7023	73.90	0
factor(Minority)Yes	-2.94	0.21	7023	-14.19	0
SES	2.13	0.11	7023	20.07	0
σ	1.98	NA	NA	NA	NA
τ	6.01	NA	NA	NA	NA

Appendix

Math

```r
hist(MathAchieve$MathAch, xlab = "MathAch",main = paste('Figure1. Histgram of MathAch scores'))
```

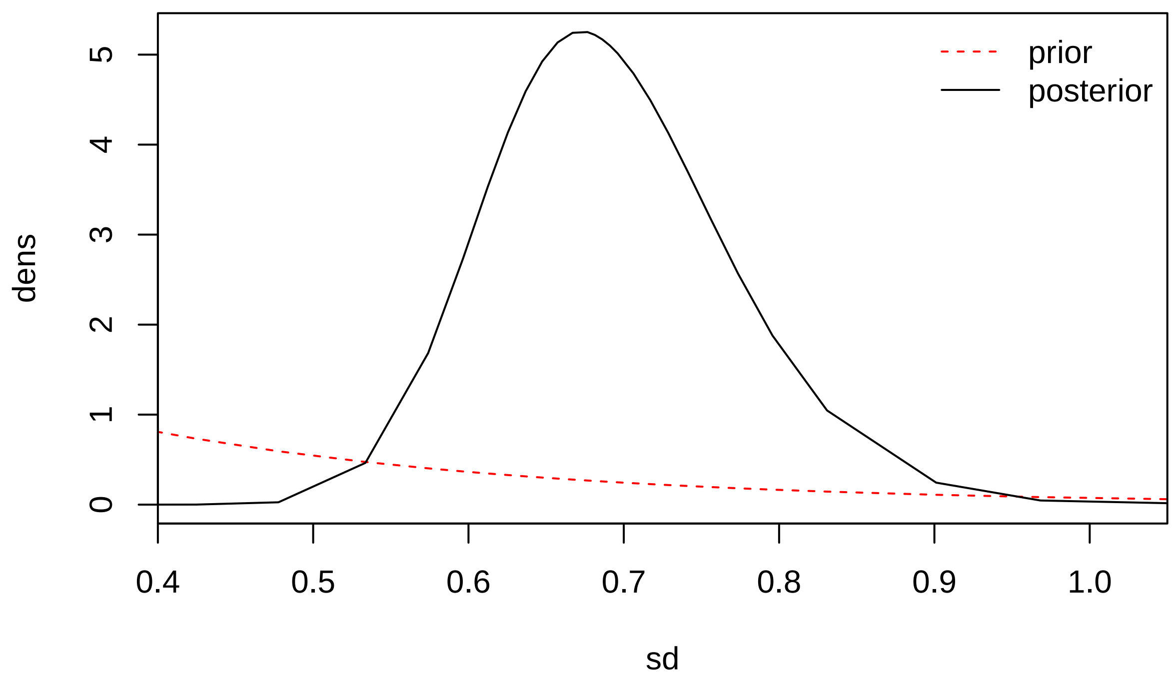



```r
Q1model = lme(MathAch ~ factor(Minority)+SES, random = ~1 |School, data = MathAchieve)

knitr::kable(Pmisc::lmeTable(Q1model), digits = 2, escape = FALSE, format = "latex",caption = "Estimati
```

Q3drugs

```
forInla = na.omit(xSub)
forInla$y = as.numeric(forInla$completed)
library("INLA")
ires = inla(y ~ AGE + SUB1 + GENDER + raceEthnicity + homeless +
              f(STFIPS, hyper=list(prec=list(
                prior='pc.prec', param=c(0.75, 0.05)))) +
              f(TOWN,hyper=list(prec=list(
                prior='pc.prec', param=c(0.75, 0.05)))),
            data=forInla, family='binomial',
            control.inla = list(strategy='gaussian', int.strategy='eb'), verbose = TRUE)
sdState = Pmisc::priorPostSd(ires)
do.call(matplot, sdState$STFIPS$matplot)
do.call(legend, sdState$legend)
```




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

Table 2: Posterior means and quantiles for model parameters.

	0.5quant	0.025quant	0.975quant
(Intercept)			
(Intercept)	0.716	0.575	0.891
AGE18-20			
AGE18-20	0.935	0.916	0.953
AGE15-17			
AGE15-17	0.926	0.905	0.947
AGE12-14			
AGE12-14	0.972	0.934	1.012
SUB1			
ALCOHOL	1.609	1.574	1.645
HEROIN	0.872	0.849	0.896
OTHER OPIATES AND SYNTHET	0.901	0.874	0.929
METHAMPHETAMINE	0.955	0.917	0.994
COCAINE/CRACK	0.855	0.814	0.898
GENDER			
FEMALE	0.893	0.878	0.909
raceEthnicity			
Hispanic	0.832	0.812	0.851
BLACK OR AFRICAN AMERICAN	0.682	0.666	0.699
AMERICAN INDIAN (OTHER TH	0.728	0.679	0.781
OTHER SINGLE RACE	0.865	0.812	0.923
TWO OR MORE RACES	0.855	0.794	0.921
ASIAN	1.132	1.038	1.235
NATIVE HAWAIIAN OR OTHER	0.845	0.748	0.953
ASIAN OR PACIFIC ISLANDER	1.454	1.227	1.723
ALASKA NATIVE (ALEUT, ESK	0.845	0.624	1.145
homeless			
TRUE	1.005	0.973	1.037
SD			
STFIPS	0.686	0.555	0.863
TOWN	0.539	0.486	0.602

```r
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")
```

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	-1.2	-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	-0.8	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	0.7	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	-0.2	-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