University of Toronto

Math score & Drug fatalitiy analysis

YiFei Gu

1 Math

In the data set `MathArchive`, whether there exists substantial difference between schools for the mathematics achievement score, in addition, the difference within schools would be analyzed. Further comparison of variance across schools and within schools are to be investigated. Note that school should be treated as a random effect if one wants to generalize results to other schools. In the mixed effect model used, the Let S_{ij} a numerical parameter denotes the score jth students from the ith school, ranging from 1 to 160. Then:

$$S_{ij}|\ \text{School}_i \sim N(\mu_{ij}, \sigma^2), \text{with}$$

$$\mu_{ij} = \beta_0 + \beta_1 \mathbb{1}_{\texttt{MinorityYes}} + \beta_2 \mathbb{1}_{\texttt{SexMale}} + \beta_3 \texttt{SES}$$

The fixed effect contained in the model are the indicator variable with $\mathbb{1}_{\texttt{MinorityYes}} = 1$ if one belongs to minority and 0 otherwise; $\mathbb{1}_{\texttt{SexMale}} = 1$ if one is male and 0 otherwise and one's social economic status, denoted as `SES`.

Table 1: Estimation of fixed effects in linear mixed model in MathAchieve

	MLE	Std.Error	DF	t-value	p-value
(Intercept)	12.885	0.193	7022	66.593	0
MinorityYes	-2.961	0.206	7022	-14.393	0
SexMale	1.230	0.163	7022	7.558	0
SES	2.089	0.106	7022	19.766	0
σ	1.917	NA	NA	NA	NA
τ	5.992	NA	NA	NA	NA

From the table above, since difference between schools, denoted $\sigma = 1.92$, and difference across schools $\tau = 5.99$, their variance differs significantly, which indicates that the variance of students' math achievement score within school are larger than across the school variation. Furthermore, all of the predictors are significant. The coefficient of `MinorityYES` indicates that being minority have a negative effect on one's score. The coefficient of `SexMale` indicates that male students tend to score higher than female students, and The coefficient of `SES` indicates that with a higher social economic status, student tend to score higher.

2 Drugs

2.1 Introduction:

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. In the report shown below, 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 would be verified. Further more, the hypothesis that, some American states have particularly effective treatment programs whereas other states have programs which are highly problematic with very low completion rates would also be analyzed.

2.2 Methods:

The method used for the analyze is generalized linear model, given that there are fixed effects: each individuals gender, race, homeless status, drugs and age; and random effects: US state and town in which the treatment was given.
Consider the model where μ_{ij} is the probability of jth person in the ith state and town that successful complete the drug treatment. As the result of whether or not completing the drug treatment is binary, hence the response follows Bernoulli distribution as stated below:

$$Y_{ij}|\text{state}_i, \text{town}_i \sim \text{Bern}(\mu_{ij})\ i.i.d$$

$$\text{logit}(\mu_{ij}) = \beta X + \text{state}_i + \text{Town}_i$$

$$\text{where state}_i \sim MVN(0, \sigma_1)\ i.i.d \text{ and town}_i \sim MVN(0, \sigma_2)\ i.i.d$$

The plot with a new prior for states and town:

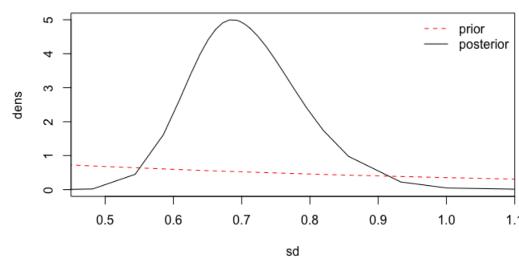

Figure 1: State-level standard deviation, with priors for precision $(u, \alpha) = (0.8, 0.35)$.

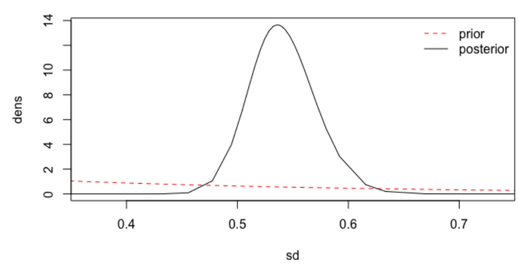

Figure 2: town-level standard deviation, with priors for precision $(u, \alpha) = (0.7, 0.1)$.

2.3 Results:

As the table shown below, the credible interval between 0.025 and 0.975 quantile need to be investigated, if the credible interval does not contain 1, then the variable is significant, otherwise not significant. In the table shown below, the reference group through analysis is marijuana. The odd ratio can also be investigated by dividing the value from 0.5 quantile to 1, which indicates the probability of completing the program in comparison to the reference group, in this case, marijuana, which is set to be 1.

Table 2: Posterior means and quantiles for model parameters.

	0.5quant	0.025quant	0.975quant
(Intercept)	0.716	0.573	0.894
SUB1(2) ALCOHOL	1.609	1.574	1.645
SUB1(5) HEROIN	0.872	0.849	0.896
SUB1(7) OTHER OPIATES AND SYNTHETICS	0.901	0.874	0.929
SUB1(10) METHAMPHETAMINE	0.955	0.916	0.994
SUB1(3) COCAINE/CRACK	0.855	0.814	0.898
GENDER(2) FEMALE	0.893	0.878	0.909
AGE18-20	0.935	0.916	0.953
AGE15-17	0.926	0.905	0.947
AGE12-14	0.972	0.934	1.012
raceEthnicityHispanic	0.832	0.812	0.851
raceEthnicityBLACK OR AFRICAN AMERICAN	0.682	0.666	0.699
raceEthnicityAMERICAN INDIAN (OTHER THAN ALASKA NATIVE)	0.728	0.679	0.781
raceEthnicityOTHER SINGLE RACE	0.866	0.812	0.923
raceEthnicityTWO OR MORE RACES	0.855	0.794	0.921
raceEthnicityASIAN	1.132	1.038	1.235
raceEthnicityNATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	0.845	0.748	0.953
raceEthnicityASIAN OR PACIFIC ISLANDER	1.454	1.227	1.723
raceEthnicityALASKA NATIVE (ALEUT, ESKIMO, INDIAN)	0.845	0.624	1.145
homelessTRUE	1.005	0.973	1.037
SD for STFIPS	0.702	0.566	0.891
SD for TOWN	0.540	0.487	0.603

As highlighted, the insignificant variable are: individuals with age between 12-14, ethnicity of Alaska native, and individuals that are homeless. To compare the difficulty of treatment, the odd ratio indicates that individual addicted with alcohol are around 61% more likely to complete the treatment than marijuana, further more, heroin, opiates and synthetics, methamphetamine, cocaine are around 13%, 10%, 5%, 14.5% less likely to complete the treatment in comparison to marijuana, imlpying that the first hypothesis is true.

Table 3: Posterior means and quantiles for model parameters.

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	0.9	NEBRASKA	0.8	0.4	1.2
ARIZONA	0.0	-1.4	1.4	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	-2.0	-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	-1.4	1.4
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	1.3	WASHINGTON	-0.1	-0.5	0.3
MICHIGAN	-0.4	-0.7	0.0	WEST VIRGINIA	0.0	-1.4	1.4
MINNESOTA	0.4	0.0	0.9	WISCONSIN	0.0	-1.4	1.4
MISSISSIPPI	0.0	-1.4	1.4	WYOMING	0.0	-1.4	1.4
MISSOURI	-0.4	-0.7	-0.1	PUERTO RICO	0.6	-0.1	1.3

States with a credible interval that contain 0 are insignificant, and those with credible interval that is to the right of 0 may imply that they have a better program offered for the treatment, and the completion rates are higher than the other states, as highlighted in green. In contrast, those with a credible interval that is to the left of 0 may imply they have a bad treatment program that leads to a low completion rate.

2.4 Conclusion

In conclusion, addressing the first hypothesis, i.e 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", is verified to be True in the above results. Other than Alcohol which is easier to complete treatment, young people that are addicted with Heroin, Opiates and Synthetics, Methamphetamine or Cocaine/Crack are are harder to be treated, with the probability of them completing the treatment being lower. In regard to the second hypothesis, the above results have shown that some states, like Delaware, Florida and Massachusetts,have better treatment program and the rate of completing the program for people with addiction is higher. Meanwhile most of the states does not differ significantly, New Mexico and Virginia have significant lower rate of completion for the treatment, meaning that their program, in comparison to other states, are are highly problematic. Both of the hypothesis have been verified true.

3 Appendix

```
> data("MathAchieve", package = "MEMSS")
> model <- lme(MathAch ~ Minority + Sex + SES, random = ~1 | School, data=MathAchieve)
> knitr::kable(Pmisc::lmeTable(model), digits = 3, caption = "Estimation of fixed effects
↪  in linear mixed model in MathAchieve",escape = FALSE,format = "latex")
\begin{table}

\caption{\label{tab:}Estimation of fixed effects in linear mixed model in MathAchieve}
\centering
\begin{tabular}[t]{l|r|r|r|r|r}
\hline
  & MLE & Std.Error & DF & t-value & p-value\\
\hline
(Intercept) & 12.885 & 0.193 & 7022 & 66.593 & 0\\
\hline
MinorityYes & -2.961 & 0.206 & 7022 & -14.393 & 0\\
\hline
SexMale & 1.230 & 0.163 & 7022 & 7.558 & 0\\
\hline
SES & 2.089 & 0.106 & 7022 & 19.766 & 0\\
\hline
$\sigma$ & 1.917 & NA & NA & NA & NA\\
\hline
$\tau$ & 5.992 & NA & NA & NA & NA\\
\hline
\end{tabular}
\end{table}
> 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
           616            1360           4479           1508
(6) CALIFORNIA
         48065
> table(xSub$TOWN)[1:2]

ABILENE, TX   AKRON, OH
         42        1078
> forInla = na.omit(xSub)
> forInla$y = as.numeric(forInla$completed)
> ires = inla(y ~ SUB1 + GENDER + AGE + raceEthnicity + homeless + f(STFIPS,
↪  hyper=list(prec=list( prior='pc.prec', param=c(0.8, 0.35)))) + f(TOWN,
↪  hyper=list(prec=list( prior='pc.prec', param=c(0.7, 0.1)))), 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)
> toPrint = as.data.frame(rbind(exp(ires$summary.fixed[, c(4, 3, 5)]), sdState$summary[,
↪  c(4, 3, 5)]))
> knitr::kable(toPrint, digits = 3, format = "latex")

\begin{tabular}{l|r|r|r}
\hline
  & 0.5quant & 0.025quant & 0.975quant\\
\hline
(Intercept) & 0.716 & 0.573 & 0.894\\
\hline
SUB1(2) ALCOHOL & 1.609 & 1.574 & 1.645\\
\hline
SUB1(5) HEROIN & 0.872 & 0.849 & 0.896\\
\hline
SUB1(7) OTHER OPIATES AND SYNTHETICS & 0.901 & 0.874 & 0.929\\
\hline
SUB1(10) METHAMPHETAMINE & 0.955 & 0.916 & 0.994\\
\hline
SUB1(3) COCAINE/CRACK & 0.855 & 0.814 & 0.898\\
\hline
GENDER(2) FEMALE & 0.893 & 0.878 & 0.909\\
\hline
AGE18-20 & 0.935 & 0.916 & 0.953\\
\hline
AGE15-17 & 0.926 & 0.905 & 0.947\\
\hline
AGE12-14 & 0.972 & 0.934 & 1.012\\
\hline
```

```
raceEthnicityHispanic & 0.832 & 0.812 & 0.851\\
\hline
raceEthnicityBLACK OR AFRICAN AMERICAN & 0.682 & 0.666 & 0.699\\
\hline
raceEthnicityAMERICAN INDIAN (OTHER THAN ALASKA NATIVE) & 0.728 & 0.679 & 0.781\\
\hline
raceEthnicityOTHER SINGLE RACE & 0.866 & 0.812 & 0.923\\
\hline
raceEthnicityTWO OR MORE RACES & 0.855 & 0.794 & 0.921\\
\hline
raceEthnicityASIAN & 1.132 & 1.038 & 1.235\\
\hline
raceEthnicityNATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER & 0.845 & 0.748 & 0.953\\
\hline
raceEthnicityASIAN OR PACIFIC ISLANDER & 1.454 & 1.227 & 1.723\\
\hline
raceEthnicityALASKA NATIVE (ALEUT, ESKIMO, INDIAN) & 0.845 & 0.624 & 1.145\\
\hline
homelessTRUE & 1.005 & 0.973 & 1.037\\
\hline
SD for STFIPS & 0.702 & 0.566 & 0.891\\
\hline
SD for TOWN & 0.540 & 0.487 & 0.603\\
\hline
\end{tabular}
> 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")
```