STA442 Assignment 2

Hongbo Du
1003568089
hongbo.du@mail.utoronto.ca

Contents

Question 1

Introduction

This analysis contains all of the primary statistical results obtained from the data set `MathAchieve` (`MEMSS` package), by using R programming language. The insterest of this analysis is that if treats `School` as a random effect, whether it appear that there are substantial differences between schools or there are differences between students within schools nearly as big as differences between students from different schools.

Method

In order to see the difference between individuals versus the difference between different schools, the analysis carries out a linear mixed model that treats `School` as random effect and the factor `Minority` and variable `SES` as fixed effects. The random effects model can be mathematically expressed as

$$Y_{ij}|U_i \overset{ind}{\sim} \mathcal{N}(\mu_{ij}, \tau^2)$$
$$\mu_{ij} = X_{ij}\beta + U_i$$
$$U_i \overset{ind}{\sim} \mathcal{N}(0, \sigma^2).$$

Therefore, looking at the the standard deviations in the random effects model would provide enough information to check which group has a large variance in this analysis. Lastly, checking if the histogram of the random effect coefficient follows a Gaussian distribution is necessary.

Result

By fitting a linear mixed effect model, we notice that $\sigma = 1.984$ and $\tau = 6.012$ (in Table 1.1) where σ represents the standard deviation of different schools (school level) and τ represents the standard deviation of individuals within the school (individual level). The results show that the variance of individuals within school is greater than the variance of amongst different schools. In terms of fixed effects, SES has a MLE of 2.128 and MinorityYes has a MLE of -2.938.

In addition, we can check the following histogram of the coefficient of random effects (Figure 1.1) to see if the coefficients of random effects are normally distributed. We noticed that the histogram follows a Gaussian distribution with a mean at 0 (e.g., the vertical black dashed line). This result follows our theoretical random effects model.

Conclusion

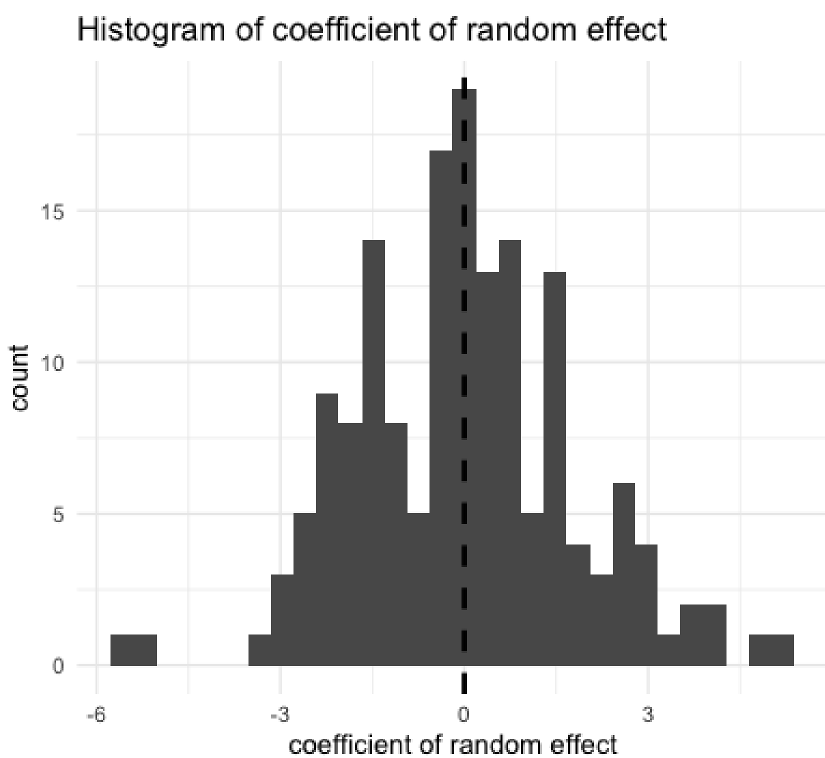
According to our statistical results, we conclude that there are substantial differences between individual students within schools by analyzing the math achievement; however, comparatively, there is no difference between schools as big as between individuals within school. Meanwhile, SES has positive relationship with the math achievements and minority students have negative relationship with the math achievements.

Tables and Figures

Table 1.1

	MLE	Std.Error	DF	t-value	p-value
(Intercept)	13.466	0.182	7023	73.896	0
SES	2.128	0.106	7023	20.068	0
MinorityYes	−2.938	0.207	7023	−14.191	0
σ	1.984	NA	NA	NA	NA
τ	6.012	NA	NA	NA	NA

Figure 1.2



Question 2

Introduction

This analysis contains the statistical results obtained from the data set of The Treatment Episode Data Set – Discharges (TEDS-D) via using R programming language. The data contains the variables of `completed` (whether completed treatment), `SUB1` (primary addition), `GENDER`, `AGE`, `raceEthnicity`, `STFIPS`, and `TOWN`. The goal of this report is to investigate the relating topic of drug treatment programs in the United States.

Methods

In order to complete this analysis, we start by constructing the hypothesis as follows:

- Hypothesis 1: the 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.
- Hypothesis 2: whether some American states have particularly effective treatment programs whereas other states have programs which are highly problematic with very low completion rates.

Therefore, we run INLA to analyze the data, and we obtain the prior density and posterior density as in the Figure 2.1.

Results

For the primary hypothesis, if taking a look at the Table 2.1, we observe that the 0.5 quantile is 1.642 for `ALCOHOL` which is higher than other hard drugs, such as the 0.5 quantile of `HEROIN` at 0.898, the 0.5 quantile of `OTHER OPIATES AND SYNTHET` at 0.924, the 0.5 quantile of `METHAMPHETAMINE` at 0.982, and the 0.5 quantile of `COCAINE/CRACK` at 0.982.

According to the Table 2.2, we notice that the means are at different levels by comparing the mean of different states. For example, `DELAWARE` has a mean of 1.0 while `VIRGINIA` has a mean of -2.9. From the Mean with .95 quantile plot (Figure 2.1), we observe that the mean line is not a straight line and it has very high volatility, as well as the .95 quantile lines.

Conclusions

The statistical results implies that there is a higher probability of completing treatment in alcohol and marijuana than the probability of completing treatment in other hard drugs. Similarly speaking, the 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.

For the secondary hypothesis, we can conclude that there are differences between different American states, that is, some American states have particularly effective treatment programs whereas other states have programs which are highly problematic with very low completion rates.

Tables and Figures

Table 2.1: Posterior means and quantiles for model parameters.

		0.5quant	**0.025quant**	**0.975quant**
(intercept)				
	(intercept)	0.682	0.652	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 (OHTER 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

Table 2.2

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
ARKANSAS	−0.1	−0.7	0.4	NEW HAMPSHIRE	0.2	−0.3	0.7
CALIFORNIA	−0.3	−0.5	0.0	NEW JERSEY	0.5	0.2	0.8
COLORADO	0.5	0.1	0.9	NEW MEXICO	−1.1	−1.8	−0.4
CONNECTICUT	0.1	−0.4	0.6	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	−0.9	0.4
FLORIDA	1.0	0.7	1.3	OHIO	−0.2	−0.5	0.1
GEORGIA	−0.2	−0.8	0.4	OKLAHOMA	0.5	0.0	1.1
HAWAII	0.2	−0.6	1.0	OREGON	0.1	−0.3	0.4
IDAHO	−0.2	−1.0	0.6	PENNSYLVANIA	0.0	−1.1	1.1
ILLINOIS	−0.5	−0.8	−0.2	RHODE ISLAND	−0.2	−0.6	0.2
INDIANA	0.0	−0.8	0.8	SOUTH CAROLINA	0.4	0.1	0.7
IOWA	0.4	0.1	0.7	SOUTH DAKOTA	0.4	−0.4	1.2
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	−0.5	−0.9	−0.1	UTAH	0.1	−0.5	0.6
MAINE	0.1	−0.7	0.9	VERMONT	−0.2	−1.0	0.6
MARYLAND	0.5	0.2	0.8	VIRGINIA	−2.9	−3.2	−2.5
MASSACHUSETTS	0.8	0.4	1.2	WASHINGTON	−0.1	−0.4	0.2
MICHIGAN	−0.4	−0.7	0.0	WEST VIRGINIA	0.0	−1.1	1.1
MINNESOTA	0.4	0.0	0.9	WISCONSIN	0.0	−1.1	1.1
MISSISSIPPI	0.0	−1.1	1.1	WYOMING	0.0	−1.1	1.1
MISSOURI	−0.4	−0.7	−0.1	PUERTO RICO	0.5	−0.1	1.2

Figure 2.1

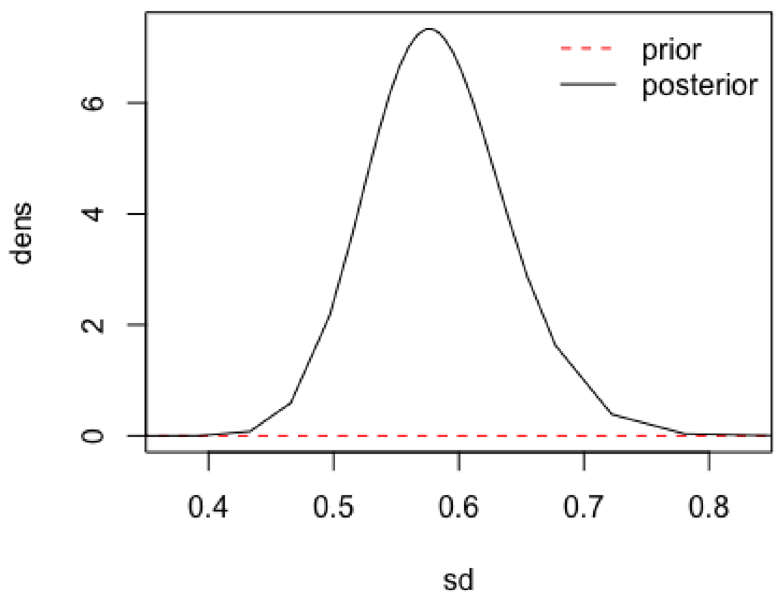


Figure 2.2