Multi-Level Analysis

Ahmed Al-Ali

Research Question and Data

The basic structure is that we have six intelligence measures for 400 children from 60 families. Thus, we have children nested within families. If intelligence has a genetic or heritable component, the between-family differences may be of particular interest. The measures are word list, cards, matrices, figures, animals, and occupations. One possibility is that the intelligence tests all form a single g factor at the between-family level such that some families have higher average scores on intelligence measures than others. Another possibility is that the subscales form two factor: numeric (word list, cards, matrices) versus perceptual (figures, animals, occupations) intelligence.

As in the case of standard MLM, we often wish to examine the amount of variation at each level in an unconditional model.Examining the independence and saturated models is informative. , the independence model gives us a worst case fit where we do not permit item covariation. And the saturated model gives us perfect fit without any parsimony or prediction.In addition we can examine the spectrum between independence and saturated at both the within and between levels.

Data

Cleaning data

Data is given in long format which means that each child in family has a score for each test section and then indicator informs us which section it is , converting it to wide format

Feature engineering

We separate and create 2 data sets one for children level and one for family level instead of compiled data set to make ot easier to illustrate, understand etc… we can minimize the number of variables by combining the sections into numeric and perceptual sections as in general IQ tests

Family size: we would seperate those families into large and small families according to the number of children previously had , mean of 3.5 corresponds to 6 children so any family with value greater than this is considered large, this variable is added because the more number of children have the more the influence over all family it has

Exceptional: binary varaible where if the family IQ is greater than the 75th percentile then 1 is given

Level one-covaraites: Child ID, Scores among each children

Level two-covaraites : Aggregated score among family IQ ,family size,Exceptional,number of children

Statistical Modeling

In order to run a multilevel modeling we should be clear about what are fixed and random parameters in our model,our data exhibit a hierarchies level modeling. We consider splitting our Modeling into level one and level two.

Fixed effects: *Family ID* Child ID within family

Random effects: *numeric* perceptual *family.size* Number of children *exceptional* IQ

For children within families we can model the data as (representation purposes only)

$$
\begin{aligned}
y_{ij} &= \beta_o + u_j + \epsilon_{ij} \\
\epsilon_{ij} &\sim N(0, \sigma_u^2) \\
\epsilon_{ij} &\sim N(0, \sigma_e^2)
\end{aligned}
$$

where y_{ij} is the IQ for ith child in the jth family, β_o is the mean IQ , u_j is the departure for the jth family form the mean and the error is the departure of child from family, note that model changes accordingly.

Level one covaraite modeling The simplest model is to allow for a unconditional random intercept model

$$
\begin{aligned}
y_{ij} &= \beta_{0j} + u_{ij} \\
\beta_{0j} &= \gamma_{00} + v_{0j}
\end{aligned}
$$

ICC of null model

$$ICC = \frac{\sigma_u^2}{\sigma^2 + \sigma_u^2} = \frac{9.245}{9.245 + 8.470} = 0.522$$

mean variation between family of same children is about 52.2%

Random intercept , fixed slope model

$$
\begin{aligned}
y_{ij} = \beta_{0j} + \beta_{1j} numeric_{ij} + \beta_{2j} perceptual_{ij} + u_{ij} \\
\beta_{0j} = \gamma_{00} + v_{0j}
\end{aligned}
$$

Random intercept , random slope model

$$
\begin{aligned}
y_{ij} = \beta_{0j} + \beta_{1j} numeric_{ij} + \beta_{2j} perceptual_{ij} + u_{ij} \\
\beta_{0j} = \gamma_{00} + v_{0j} \\
\beta_{1j} = \gamma_{10} + v_{1j} \\
\beta_{2j} = \gamma_{20} + v_{2j}
\end{aligned}
$$

Before moving on lets compare the models

According to those tests we can infer that random intercept and fixed slopes for lvl one covaraites are better according to the p-value we accept the reduced model in the final test over the random slope included, going forward we will build on random intercept fixed slope model.

Level two modeling:Family level

Random intercept,fixed slope model with the addition of level two covaraite as fixed

$$
\begin{aligned}
y_{ij} = \beta_{0j} + \beta_{1j} numeric_{ij} + \beta_{2j} perceptual_{ij} + u_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} familysize_j + \gamma_{02} exceptional_j + v_{0j}
\end{aligned}
$$

Random intercept,fixed slope model with the addition of level two covaraite as random

$$
\begin{aligned}
y_{ij} = \beta_{0j} + \beta_{1j} numeric_{ij} + \beta_{2j} perceptual_{ij} + u_{ij} \\
\beta_{0j} = \gamma_{00} + \gamma_{01} familysize_j + \gamma_{02} exceptional_j + v_{0j} \\
\gamma_{01} = \alpha_{01} + k_{01} \\
\gamma_{02} = \alpha_{02} + k_{02}
\end{aligned}
$$

compare the two models

Add an interaction term

Compare models

we accept the reduced model Random intercept,fixed slope model with the addition of level two covaraite as fixed

$$y_{ij} = \beta_{0j} + \beta_{1j} numeric_{ij} + \beta_{2j} perceptual_{ij} + u_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} familysize_j + \gamma_{02} exceptional_j + v_{0j}$$

Interpretation

Given that our final model , we should interpret parameters now which are slopes and random affects 8 parameters including ICC . $\sigma^2 = 4.288 \times 10^{-7}$the variance in within-family deviations between child $\hat{\sigma}_u^2 = 1.005 \times 10^{-5}$ the variance in between-family deviations between children means and the overall mean across all children $\beta_o = 5.226 \times 10^P - 4$ the mean IQ level across all families $\beta_1 = 0.5$ for every one increase in numeric section the IQ level increases by half point $\beta_2 = 0.5$for every one increase in perceptual section the IQ level increases by half point $\gamma_1 = 1.474 \times 10^{-4}$ an exceptional intelligent family has a higher mean score by given value $\gamma_2 = -1.013 \times 10^{-4}$ a small size family has a reduction in IQ level by a mean of given value $ICC = \frac{\sigma_u^2}{\sigma^2 + \sigma_u^2} = \frac{1.005 \times 10^{-5}}{1.005 \times 10^{-5} + 4.288 \times 10^{-7}} = 0.959$ mean variation between different families is about 95%

Conclusion

In conclusion we can say that family properties don't influence a child IQ in general but in more specificity on small scale it may differ from one family to another considering other economic factors about parents may play an important role and proportion of female to male for future studies which conclude that intelligence of kid cant be genetically correlated

Appendix

```
## ----include=FALSE------------------------------------------------------------
library(tidyr)
library(ggplot2)
library(haven)
library(lme4)
library(lmerTest)
library(tidyverse)


## ----include=FALSE------------------------------------------------------------
path = file.path("FamIQDataLong.sav")
iq = data.frame(read_sav(path))
head(iq)
str(iq)


## ----include=FALSE------------------------------------------------------------
iq.wide=data.frame("family"=as.numeric(),
              "child"=as.numeric(),
              "wordlist"=as.numeric(),
```

```r
              "cards"=as.numeric(),
              "matrices"=as.numeric(),
              "figures"=as.numeric(),
              "animals"=as.numeric(),
              "occupations"=as.numeric())


for (i in seq(0,2400,6)){
  temp=data.frame("family"=iq[i+1,"family"],
              "child"=iq[i+1,"child"],
              "wordlist"=iq[i+1,"score"],
              "cards"=iq[i+1,"score"],
              "matrices"=iq[i+2,"score"],
              "figures"=iq[i+3,"score"],
              "animals"=iq[i+4,"score"],
              "occupations"=iq[i+5,"score"])
  iq.wide=rbind(iq.wide,temp)
}


## ----include=FALSE-----------------------------------------------------------
iq.wide$numeric=round((iq.wide$wordlist+iq.wide$cards+iq.wide$matrices)/3,2)
iq.wide$perceptual=round((iq.wide$figures+iq.wide$animals+iq.wide$occupations)/3,2)
iq.wide[,3:8]=NULL


## ----include=FALSE-----------------------------------------------------------
iq.wide$IQ=round((iq.wide$numeric+iq.wide$perceptual)/2,2)
iq.wide=iq.wide[1:400,]
iq.child=iq.wide
iq.child$child=seq(1:nrow(iq.child))


## ----include=FALSE-----------------------------------------------------------
family.children=iq.wide %>% group_by(family) %>% slice(which.max(child)) %>% select(family,child)
mean.iq.sections=round(aggregate(iq.wide[, 3:5], list(iq.wide$family), mean),2)
iq.family=cbind(family.children,mean.iq.sections[,2:4])

#colnames(mean.iq)=c("family","children","wordlist","cards","matrices","figures","animals","occupations


## ----include=FALSE-----------------------------------------------------------
iq.family$family.size=ifelse(iq.family$child>6, yes="Large" , no="Small")


## ----include=FALSE-----------------------------------------------------------
iq.family$exceptional=ifelse(iq.family$IQ>quantile(iq.family$IQ,0.75),yes=1,no=0)


## ----include=FALSE-----------------------------------------------------------
knitr::kable(head(iq.child))
```

```
## ----include=FALSE-----------------------------------------------------------
knitr::kable(head(iq.family))


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## theme.1 <- theme(axis.title.x = element_text(size = 14),
##   axis.title.y = element_text(size = 14),
##   plot.title=element_text(hjust=.9,face="italic",size=12))


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## iq.child%>%
##   pivot_longer(c(numeric,perceptual,IQ), names_to = "section", values_to = "score") %>%
##   ggplot(aes(x = family,y=score, colour = family)) +
##   facet_wrap(vars(section), ncol = 3) +
##   geom_point(stat = "identity") +
##   geom_line(stat = "identity") +
##   labs(x = "Family ", y = "Score")
##


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## # scatter plot
## ggplot(data = iq.child) +
##   geom_point(mapping = aes(x = child, y = IQ, color = family))


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## # Lattice plot
## ggplot(iq.wide, aes(x=child,y=IQ)) +
##   geom_line(aes(group=family), color="blue") +
##   facet_wrap(~family, ncol=10) +
##   geom_smooth(se=FALSE,color="black")+theme.1
##


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## #Box plot
## par(cex.axis=0.75) # is for x-axis
## boxplot(iq.wide[,c(3,4)])
##
## boxplot(iq.wide[,c(5)],xlab="IQ score")
##


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## iq.family%>%
##   pivot_longer(c(numeric,perceptual,IQ), names_to = "section", values_to = "score") %>%
##   ggplot(aes(x = family,y=score, colour = family.size)) +
##   facet_wrap(vars(section), ncol = 3) +
##   geom_point(stat = "identity") +
##   geom_line(stat = "identity") +
##   labs(x = "Family ", y = "Score")
```

```r
## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## iq.family %>%
##   pivot_longer(c(child,exceptional), names_to = "section", values_to = "score") %>%
##   ggplot(aes(x=score, colour = family.size)) +
##   facet_wrap(vars(section), ncol = 3) +
##   geom_point(stat = "count") +
##   geom_line(stat = "count") +
##   labs(x = "Score ", y = "Count")


## ----eval=FALSE, message=FALSE, warning=FALSE, include=FALSE, indent=""-------
## #  plot
## ggplot(iq.family) +
##   geom_line(mapping = aes(x = family, y = IQ, color = family.size))


## ----include=FALSE------------------------------------------------------------
con.int<-lmer(IQ ~ 1 + (1|family),data=iq.wide)
summary(con.int)


## ----include=FALSE------------------------------------------------------------
rand.int.fix.slp<-lmer(IQ ~ 1+numeric+ perceptual+ (1|family),data=iq.wide)
summary(rand.int.fix.slp)


## ----include=FALSE------------------------------------------------------------
rand.int.rand.slp<-lmer(IQ ~ 1+numeric+ perceptual+ (1+numeric+ perceptual|family),data=iq.wide)
summary(rand.int.rand.slp)


## ----include=FALSE------------------------------------------------------------
anova(con.int,rand.int.fix.slp)
anova(con.int,rand.int.rand.slp)
anova(rand.int.fix.slp,rand.int.rand.slp)


## ----include=FALSE------------------------------------------------------------
iq.family.an<-merge(iq.wide,iq.family,by="family")
iq.family.an[,c(6,7,8,9)]=NULL
colnames(iq.family.an)=list("family","child","numeric","perceptual","IQ","family.size","exceptional")


## ----include=FALSE------------------------------------------------------------
model1<-lmer(IQ ~ 1+numeric+ perceptual+exceptional+family.size +(1|family),data=iq.family.an)
summary(model1)


## ----include=FALSE------------------------------------------------------------
model2<-lmer(IQ ~ 1+numeric+ perceptual+exceptional+family.size +(1+exceptional+family.size|family),data
summary(model2)
```

```
## ----include=FALSE-------------------------------------------------------------
anova(model1,model2)


## ----include=FALSE-------------------------------------------------------------
model3<-lmer(IQ ~ 1+numeric+ perceptual+(family.size:exceptional)+(1|family),data=iq.family.an)
summary(model3)


## ----include=FALSE-------------------------------------------------------------
knitr::kable(anova(model1,model3))


## ----include=FALSE-------------------------------------------------------------
summary(model1)


## ----code = readLines(knitr::purl("/Users/ahmed/Documents/Penn state university/Semesters/Fall 2021/S
## NA
```