Homework 2

140.656 Multi-level models

BEFORE YOU GET STARTED:

Here are the rules of conduct for Homework 2:

* We encourage you to work in groups on completing the required data analysis for the homework assignment, i.e. Parts A – E.
* Post questions about Stata/R commands! You should make a reasonable attempt at figuring out the commands (Stata help is great and so is a Google search); however, do not spend HOUR(S) on a single command/figure. Please ask for help!
* Parts F and G of the assignment may be completed in teams of up to 4. Note: if you are submitting as a team, all team members will receive the same score for the assignment.

**Background:** In 2003, the Maryland General Assembly enacted the Public Charter School Act. The act created Maryland’s first public charter school program “to establish an alternative means within the existing public school system in order to provide innovative learning opportunities and creative educational approaches to improve the education of students." Charter schools in Maryland are public schools where admission is based on a lottery that is open to all families within the specific county or Baltimore City. Since the passage of the Public Charter School Act in 2003, roughly 35 charter schools have opened and are operating in Maryland; the vast majority of charter schools are in Baltimore City.

There is on-going debate about whether charter schools add value above what students receive at traditional schools; i.e. the child’s neighborhood public school.

**Your assignment:** You will evaluate whether students in charter schools perform better academically than students in traditional schools within Baltimore City. The Maryland School Assessment (MSA) is an annual assessment program that tests students in grades 3 through 8 in reading and mathematics (<http://reportcard.msde.maryland.gov/>).

Your assignment is to analyze data from the MSA mathematics test for 3rd, 4th and 5th graders during the **2017-18 academic year** and write a short abstract detailing whether there is evidence in Baltimore City that children enrolled in charter schools have different achievement in mathematics as measured by the mathematics MSA.

Your abstract should address the following objectives:

1. Compare the proportion of students who “pass” the mathematics MSA, as defined as student performance being *Level-3* to *Level-5*, across charter and traditional schools, separately for 3rd, 4th and 5th graders. Include proper summaries of the statistical evidence for or against differences between the charter and traditional schools (i.e. confidence intervals, p-values).
2. Determine if the school enrollment (i.e. number of student attending the school), attendance rateand the proportion of students who qualify for the federal free/reduced cost lunch program (FARMS) confounds the relationship between “passing” the mathematics MSA and school type (charter vs. traditional) at each of the grade levels.

**Submitting your assignment:**

Your Homework 2 submission should consist of:

* The one-page abstract
  + Your abstract should contain an objective, study design (including where you obtained the data), methods, results and conclusion sections.
  + Your abstract should be no longer than 1 page, single spaced with standard 1” margins.
  + **You should include a descriptive figure summarizing variation across schools in the proportion of students who “pass” the mathematics MSA stratified by grade level and school type (charter vs. traditional school)**
  + You may include up to 2 additional tables/figures, please reference these appropriately in your abstract.
  + Document should be single-spaced, 10 point Times font, one inch margins.
  + Code does not need to accompany results for the main abstract, but code SHOULD accompany short answer questions. Since we grade for correctness, we will need to see your code.
  + A title and author list are at the top of the document (may be included in a header above the one inch margin)
  + Word docs are preferred, but PDFs are ok. (R users can knit to word docs in R studio, too!)
* The descriptive figure summarizing the proportion of students who “pass” stratified by grade level and school type
* The at most 2 additional tables/figures that are referenced in the abstract
* Answers to **Short answer questions,** see below
* If you are submitting as a team, then you should also include the following contribution section on a separate page at the end of your Homework document:
  + A list of names defining the team.
  + A statement about who contributed what to the data analysis (e.g. Initials1, Initials2, Initials3 met as a working group to complete the data analysis in Parts A – E)
  + A statement about who contributed what to the abstract writing (e.g. Initials1 drafted the Objective, Study Design and Conclusion section. All three authors contributed equally to the writing of the methods and results section. Initials2 edited the final abstract.)
  + A statement about who contributed what to the short answer questions (e.g. All three authors met as a working group to compile answers to the short answer questions.)

Submit your assignment via the Courseplus Dropbox. No late submissions will be accepted. **If you are submitting as a team, a single team member should submit the document.**

1. **The Data:**

Assessment data for the **2017-18** academic year are available from:

<http://reportcard.msde.maryland.gov/>

http://reportcard.msde.maryland.gov/Graphs/#/DataDownloads/datadownload/3/17/6/99/XXXX/3/17/6/99/XXXX

In addition to these data, we used the Baltimore City Schools website and the Maryland Public Charter School Association websites to identify which city schools are charter vs. traditional schools.

The dataset contains data from 121 elementary schools, 24 of these are charter schools. NOTE: There are more than 121 elementary schools in Baltimore City. A few schools were excluded because they either didn’t report any MSA data or we could not identify them as charter or traditional schools.

The data represent two-level data with students nested within schools.

You can download the dataset (**HW2 MSA 2017.csv**) from Courseplus.

The data contains the following variables:

School identifiers:

* school\_number: number assigned to school within county/city
* school\_name: name of school (character)

School-level variables:

* management\_type: character string describing the type of school (e.g. Traditional vs. charter).
* charter: indicator for whether a school is a charter school
* enrolled\_count: total number of students enrolled in the school
* entrants\_pct: percentage of students entering (transferring in or re-entering) school during the September to June school year after the first day of school.
* withdrawals\_pct: percentage of students withdrawing (transfers and terminations) for any reason during the September to June school year after the first day of school.
* farms\_pct: estimated percentage of students would qualify for free/reduced price meals based on family size and income guidelines (as promulgated annually by the U.S. Department of Agriculture). NOTE: In Baltimore City, breakfast and lunch is provided at school for all students; regardless of qualification for FARMS
* attend\_rate\_pct: percentage of student school days attended.

Student-level variables:

* grade: grade level of the students; “Grade 3”, “Grade 4”, “Grade 5”
* tested\_count: total number of students who completed the test at the given grade level
* Level\_1 … Level\_5: total number of students at the given grade level who scored in the respective category:
  + Level 1: did not yet meet expectations
  + Level 2: partially met expectations
  + Level 3: Approached\_expectations
  + Level 4: Met\_expectations
  + Level 5: Exceeded\_expectation
* pass: number of students who scored level 3 or higher among those who completed the test at the given grade level

# Important features of the data and analysis:

1. You can think of the data as having the following structure: child nested within school. So the data has two levels; i for school (i = 1, .., 121), j for student within school i (j = 1,

…, ni). The response is Yij, the indicator of “passing” the Math MSA. The primary exposure variables of interest are grade level (Xij=3,4,5), which is a student-level variable, and school type (Zi, charter (1) vs. traditional school (0)).

To protect the identity of the individual children, the MSA program only makes available aggregate data by grade level and school. Therefore, instead of a dataset with a row per child per grade level per school, i.e. (Yij, Xij,Zi), the public is provided with a row of data per grade level per school.

For each school *i* and grade level *k* = 3, 4 and 5 respectively,

* is the number of students in grade *k* from school *i* who took the mathematics MSA
* is the number of students who “pass” the mathematics MSA in grade *k* from school *i*
* Lastly, define a grade variable with values 3, 4 and 5.

Here is a listing of selected variables from the first 7 rows of available data. Notice that not all schools will have scores available for all grade levels. In the dataset, is “grade”, is “tested\_count”, is “pass” and is “charter” (the school level charter school indicator).

list school\_number school\_name testedcount grade pass charter in 1/7

+------------------------------------------------------------------------------------+

| school~r school\_name grade tested~t pass charter |

|------------------------------------------------------------------------------------|

1. | 314 SharpLeadenhall Elementary Grade 5 11 1 0 |

2. | 314 SharpLeadenhall Elementary Grade 4 13 1 0 |

3. | 371 Lillie May Carroll Jackson School Grade 5 13 3 1 |

4. | 322 New Song Academy Grade 5 14 5 0 |

5. | 89 Rognel Heights ElementaryMiddle Grade 3 14 3 0 |

|------------------------------------------------------------------------------------|

6. | 322 New Song Academy Grade 4 15 6 0 |

7. | 379 Roots and Branches School Grade 5 15 1 1 |

+------------------------------------------------------------------------------------+

Given that the data is aggregated at the school and grade level, how should you fit the models?

Ideally, if you had (Yij, Xij, Zi) then the model with main effects for grade level and school type may look like the following:

Without losing any information, the model for the aggregated data can be expressed as:

When we are fitting the models, we will need to specify both the number of students who “pass” (this is our outcome) and the number of students taking the test for a given grade level (this is “tested\_count” in the dataset).

This can be achieved in STATA using meqrlogit:

gen grade4 = grade == “Grade 4”

gen grade5 = grade == “Grade 5”

meqrlogit pass grade4 grade5 charter || school\_number: , /// binomial(tested\_count)

If you are an R-user:

library(lme4)

data$grade4 = ifelse(data$Grade==”Grade 4”,1,0)

data$grade5 = ifelse(data$Grade==”Grade 5”,1,0)

fit = glmer(cbind(pass,Tested\_Count-pass)

~grade4+grade5+(1|school\_number),data=data,family="binomial",nAGQ=7)

summary(fit)

1. You do not have to fit a random intercept model, you can use marginal models allowing for correlation among students from the same schools. This is your decision.

To fit the marginal model in this case, you can use:

STATA users:

xtset school\_number

xtgee pass grade4 grade5 charter, family(binomial testedcount) corr(exch) eform

R users:

library(geepack)

fit = geeglm(cbind(pass,Tested\_Count – pass)

~grade4+grade5+charter,family=”binomial”,corstr=”exchangeable”,

data=data,id=data$School\_Number)

summary(fit)

# Short answer questions for submission with Homework 2:

Run the code in the file “HW2 short answer” do-file (Stata users) or R script (R users).

Then answer the following questions:

**Question 1:** Using the results from Model 1, provide an interval that contains the school-specific proportion of students who “passed” the mathematics MSA for roughly 95% of Baltimore City Schools.

**Question 2:** Using the results from Model 2, interpret the exponentiated value of the sum of the regression coefficients assigned to the “charter” and “grade4charter” variables.

**Question 3:** Using the results from Model 2 and 3, separately for each grade level, provide evidence for whether or not the school enrollment, the proportion of students qualifying for FARMS, and attendance rate confound the relationship between “passing” the mathematics MSA and school type.

**Question 4:** Using the results from Model 4, interpret the exponentiated value of the sum of the regression coefficients assigned to the “charter” and “grade4charter” variables.