analysis

Anonymous

2025-10-25

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
# b) Multilevel model with random intercept for school, district, and student-level variance  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
# Running this multilevel model to account for:  
# 1. school variance - differences between schools  
# 2. district variance - differences between districts   
# 3. student-level variance - differences between students withing schools  
  
# load the data  
merged\_centered = file.path(out\_data\_fld,"df\_merged\_centered.xlsx")  
df\_merged\_centered = read\_excel(merged\_centered)   
  
# transform the data in numeric  
df\_merged\_centered <- df\_merged\_centered %>%  
 mutate(  
 reading\_z = as.numeric(reading\_z),  
 math\_z = as.numeric(math\_z)  
 )  
  
# Multilevel model on reading z-scores  
mixed\_reading <- lmer(  
 reading\_z ~ factor(household\_income\_quintile) + female + rural +  
 disability\_status + parent\_education\_level + attendance\_centered +  
 (1 | school\_id) + (1 | district\_id) + factor(year),  
 data = df\_merged\_centered,  
 weights = weight  
)

## boundary (singular) fit: see help('isSingular')

# summarize the results -READING  
summary(mixed\_reading)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']  
## Formula: reading\_z ~ factor(household\_income\_quintile) + female + rural + disability\_status + parent\_education\_level + attendance\_centered + (1 | school\_id) + (1 | district\_id) + factor(year)  
## Data: df\_merged\_centered  
## Weights: weight  
##   
## REML criterion at convergence: 522823.4  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.7497 -0.6529 0.0008 0.6487 4.5279   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## school\_id (Intercept) 0.0991 0.3148   
## district\_id (Intercept) 0.0000 0.0000   
## Residual 0.9124 0.9552   
## Number of obs: 192000, groups: school\_id, 9993; district\_id, 25  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) -0.267441 0.007122 73340.893518 -37.551 <0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)2 0.168922 0.005463 191981.314551 30.919 <0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)3 0.351070 0.006383 191819.831515 54.997 <0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)4 0.532559 0.007972 191081.130834 66.800 <0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)5 0.713616 0.012081 189548.308506 59.071 <0.0000000000000002 \*\*\*  
## female 0.006998 0.004324 191976.994578 1.619 0.106   
## rural -0.255341 0.004534 191704.428750 -56.318 <0.0000000000000002 \*\*\*  
## disability\_statusYes 0.017494 0.010723 191982.153351 1.631 0.103   
## parent\_education\_levelPrimary 0.006538 0.005186 191985.797204 1.261 0.207   
## parent\_education\_levelSecondary -0.004582 0.005653 191984.542600 -0.811 0.418   
## parent\_education\_levelTertiary -0.009528 0.010074 191985.435938 -0.946 0.344   
## attendance\_centered 0.817648 0.034595 187438.999391 23.635 <0.0000000000000002 \*\*\*  
## factor(year)2024 0.119088 0.005010 181880.435243 23.772 <0.0000000000000002 \*\*\*  
## factor(year)2025 0.239488 0.005016 182182.250749 47.747 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Correlation matrix not shown by default, as p = 14 > 12.  
## Use print(x, correlation=TRUE) or  
## vcov(x) if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')

# Multilevel model on math z-scores  
mixed\_math <- lmer(  
 math\_z ~ factor(household\_income\_quintile) + female + rural +  
 disability\_status + parent\_education\_level + attendance\_centered +  
 (1 | school\_id) + (1 | district\_id) + factor(year),  
 data = df\_merged\_centered,  
 weights = weight  
)

## boundary (singular) fit: see help('isSingular')

# summarize the results - MATH  
summary(mixed\_math)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']  
## Formula: math\_z ~ factor(household\_income\_quintile) + female + rural + disability\_status + parent\_education\_level + attendance\_centered + (1 | school\_id) + (1 | district\_id) + factor(year)  
## Data: df\_merged\_centered  
## Weights: weight  
##   
## REML criterion at convergence: 523501.9  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.4622 -0.6491 0.0014 0.6541 4.6891   
##   
## Random effects:  
## Groups Name Variance Std.Dev.   
## school\_id (Intercept) 0.103281091581 0.32137376  
## district\_id (Intercept) 0.000000000325 0.00001803  
## Residual 0.914315170047 0.95619829  
## Number of obs: 192000, groups: school\_id, 9993; district\_id, 25  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) -0.276347 0.007161 71899.513146 -38.590 < 0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)2 0.185627 0.005473 191964.362041 33.918 < 0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)3 0.354298 0.006394 191759.017935 55.409 < 0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)4 0.541457 0.007985 190974.746546 67.810 < 0.0000000000000002 \*\*\*  
## factor(household\_income\_quintile)5 0.718476 0.012098 189429.337429 59.388 < 0.0000000000000002 \*\*\*  
## female -0.002273 0.004331 191985.817593 -0.525 0.59979   
## rural -0.256621 0.004542 191793.490351 -56.495 < 0.0000000000000002 \*\*\*  
## disability\_statusYes 0.018136 0.010742 191985.656262 1.688 0.09135 .   
## parent\_education\_levelPrimary 0.016739 0.005195 191981.722363 3.222 0.00127 \*\*   
## parent\_education\_levelSecondary 0.002512 0.005663 191984.245778 0.444 0.65731   
## parent\_education\_levelTertiary -0.009176 0.010091 191983.014325 -0.909 0.36321   
## attendance\_centered 0.816741 0.034640 187320.249197 23.578 < 0.0000000000000002 \*\*\*  
## factor(year)2024 0.121560 0.005015 181873.319700 24.239 < 0.0000000000000002 \*\*\*  
## factor(year)2025 0.244321 0.005021 182168.696707 48.658 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Correlation matrix not shown by default, as p = 14 > 12.  
## Use print(x, correlation=TRUE) or  
## vcov(x) if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
# c) Prediction math and reading score given an increase in attendance due to the proposed intervention  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
# Define values for prediction  
sd\_increase = 0.3  
nbr\_school\_days = 190  
  
# Simulate an increase of 0.3 for low-income students living in rural settings  
df\_sim <- df\_merged\_centered %>%  
 mutate(attendance\_centered = if\_else(low\_income == 1 & rural == 1, # low income represents households in the first and second quantale  
 attendance\_centered + sd\_increase,  
 attendance\_centered))  
  
# Predict outcomes using the mixed model  
df\_sim$reading\_pred <- predict(mixed\_reading, newdata = df\_sim, re.form = NA)  
df\_sim$math\_pred <- predict(mixed\_math, newdata = df\_sim, re.form = NA)  
  
# Aggregate predicted outcomes by income quantile  
sim\_summary <- df\_sim %>%  
 group\_by(household\_income\_quintile) %>%  
 summarise(mean\_reading\_pred = mean(reading\_pred),  
 mean\_math\_pred = mean(math\_pred))  
  
print(sim\_summary)

## # A tibble: 5 × 3  
## household\_income\_quintile mean\_reading\_pred mean\_math\_pred  
## <dbl> <dbl> <dbl>  
## 1 1 -0.182 -0.188   
## 2 2 0.00862 0.0189  
## 3 3 0.0984 0.0952  
## 4 4 0.303 0.305   
## 5 5 0.512 0.510

# mean and standard deviation of attendance  
mean(df\_merged\_centered$attendance\_rate, na.rm = TRUE)

## [1] 0.686

sd(df\_merged\_centered$attendance\_rate, na.rm = TRUE)

## [1] 0.07350024

# presents results on school attendance  
  
# % increase in attendance:  
sd(df\_merged\_centered$attendance\_rate, na.rm = TRUE)\*sd\_increase

## [1] 0.02205007

# Increase in attendance days thanks to school feeding. Assuming 190 days of school:  
sd(df\_merged\_centered$attendance\_rate, na.rm = TRUE)\*sd\_increase\*nbr\_school\_days

## [1] 4.189514