

PSYC 5670: Homework 1

J DeVaney

Question 1:

```
ecls <- read_sas("datasets\\eclsk_thirds_within.sas7bdat")
```

Question 2:

Descriptives:

With Hmisc:

```
ecls.nomiss <- filter(ecls, !is.na(MTH_T))
ecls.nomiss$teacherid <- as.factor(ecls.nomiss$teacherid)

describe(select(ecls.nomiss, MTH_T, SES, teacherid))
```

```
## select(ecls.nomiss, MTH_T, SES, teacherid)
##
## 3 Variables      2961 Observations
## -----
## MTH_T : Math T-Score
##      n missing distinct      Info      Mean      Gmd      .05      .10
##  2961      0      2577        1  52.56    9.73   37.63   41.73
##    .25    .50    .75    .90    .95
##  47.23   52.77   58.33   63.42   66.44
##
## lowest : 18.732 23.192 24.672 24.932 27.026, highest: 76.412 77.793 77.839 78.844 80.691
## -----
## SES : Socioeconomic status composite
##      n missing distinct      Info      Mean      Gmd      .05      .10
##  2705    256      339        1  0.2187   0.823  -0.85  -0.66
##    .25    .50    .75    .90    .95
##  -0.31   0.14   0.72   1.20   1.48
##
## lowest : -2.21 -1.93 -1.89 -1.78 -1.73, highest:  2.24  2.30  2.33  2.43  2.58
## -----
## teacherid
##      n missing distinct
##  2961      0      300
##
## lowest : 0007T41 0011T41 0011T42 0015T43 0020T41
## highest: 6290T41 6290T47 6290T48 7054T44 7151T43
## -----
```

With stargazer:

```
stargazer(as.data.frame(select(ecls.nomiss, MTH_T, SES)), type="text")
```

```
##
## =====
## Statistic    N      Mean  St. Dev.  Min    Pctl(25)  Pctl(75)  Max
## -----
## MTH_T       2,961  52.557   8.632   18.732  47.232   58.334   80.691
## SES         2,705   0.219    0.730   -2.210  -0.310    0.720    2.580
## -----
```

Average class size:

```
#avg class size = number of obsv / number of distinct teacher IDs
2961 / 300
```

```
## [1] 9.87
```

Question 3:

Baseline Model

```
m1 <- lmer(MTH_T ~ 1 + (1|teacherid), data = ecls, na.action=na.omit)
tab_model(m1, show.aic = T, show.r2 = F, show.ci = F, show.se = T)
```

Math T-Score			
Predictors	Estimates	std. Error	p
(Intercept)	52.51	0.27	<0.001

Random Effects

σ^2	58.83
τ_{00} teacherid	15.76
ICC	0.21
N teacherid	300
Observations	2961
AIC	20855.496
Gamma00	= 52.51

Grand mean intercept: Overall intercept of the regression equation

τ_{00} = **15.76**

Intercept variance: Deviation between real and predicted cluster intercepts

σ^2 = **58.83**

Residual variance: Deviation between real and predicted outcomes within clusters

AIC = **20855.50**

ICC = sJPlot report of **0.21** model output manual calc below:

```
15.76/(15.76+58.83)
```

```
## [1] 0.2112884
```

The ICC is an estimation of the correlation among within-cluster observations.

DEFT:

```
# DEFT = sqrt(1 + ICC*(n-1))  
sqrt(1 + 0.21*(9.87-1))
```

```
## [1] 1.691952
```

Multilevel modeling would be required, as the DEFT indicates OLS would produced standard errors ~69% different than actual values.

Submodels and Reduced Form

Level 1: $Y_{ij} = \text{Beta0j} + 58.83$

Level 2: $\text{Beta0j} = 52.51 + 15.76$

Reduced: $Y_{ij} = 52.51 + 15.76 + 58.83$