PSYC 5670: Homework 1

J DeVaney

Question 1:

```
ecls <- read_sas("datasets\\eclsk_thirds_combined.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="html", digits = 2, summary.stat
 = c("n", "mean", "sd", "min", "max"))
```

Statistic	N	Mean	St. Dev.	Min	Max
MTH_T	2,961	52.56	8.63	18.73	80.69
SES	2,705	0.22	0.73	-2.21	2.58

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.nomiss)
summary(m1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: MTH_T ~ 1 + (1 | teacherid)
##    Data: ecls.nomiss
##
## REML criterion at convergence: 20849.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9090 -0.6298  0.0164  0.6626  3.7658
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  teacherid (Intercept) 15.76    3.97
##  Residual              58.83    7.67
## Number of obs: 2961, groups:  teacherid, 300
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  52.5068     0.2711 297.4715   193.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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		
$\tau_{00\ \text{teacherid}}$	15.76		
ICC	0.21		
$N_{\text{teacherid}}$	300		
Observations	2961		
AIC	20855.496		

Gamma00 = **52.51**

Grand mean intercept: Overall intercept of the regression equation

Tau00 = **15.76**

Intercept variance: Deviation between real and predicted cluster intercepts

sigma^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 describes what proportion of variance in the outcome is at the cluster level.

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 accounting for clustering would produce *correct* standard errors ~69% larger than standard errors calculated using OLS.

Submodels and Reduced Form

Level 1: Yij = B0j + rij

Level 2: B0j = Gam00 + Mu0j

Reduced: Yij = 52.51 + Mu0j + rij

Question 4

```
m2 <- lmer(MTH_T ~ SES_Mean + (1|teacherid), data = ecls.nomiss)
summary(m2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: MTH_T ~ SES_Mean + (1 | teacherid)
##    Data: ecls.nomiss
##
## REML criterion at convergence: 20706.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7583 -0.6394  0.0268  0.6686  3.9586
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  teacherid (Intercept)  7.349   2.711
##  Residual              58.873   7.673
## Number of obs: 2961, groups:  teacherid, 300
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  51.3857     0.2284 296.9543  225.00   <2e-16 ***
## SES_Mean      5.9221     0.4363 289.3579   13.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## SES_Mean -0.364
```

```
tab_model(m2, show.aic = T, show.r2 = F, show.ci = F, show.se = T)
```

	Math T-Score		
Predictors	*Estimates*	*std. Error*	*p*
(Intercept)	51.39	0.23	**<0.001**
Class mean Socioeconomic status composite	5.92	0.44	**<0.001**
Random Effects			
σ^2	58.87		
$\tau_{00\ \text{teacherid}}$	7.35		
ICC	0.11		
$N_{\text{teacherid}}$	300		
Observations	2961		
AIC	20714.430		

Gamma01 = **5.92**

With every one unit change in the L2 variable SES_Mean, there is a direct 5.92 unit change in Math T-Score.

AIC = **20714.43**

The interpretation of the grand mean intercept

Tau00 = **7.35**

sigma^2 = **58.87**

Intercept variance decreased, while there is no decrease in the residual variance. This is due to the addition of the SES_Mean variable at the cluster level explaining L2 variance, while being unable to explain L1 variance.

ICC = sJPlot report of **0.11** model output manual calc below:

```
7.35/(7.35+58.83)
```

```
## [1] 0.1110607
```

Submodels and Reduced Form

L1: Yij = B0j + rij

L2: B0j = Gam00 + Mu0j

Reduced: Yij = 51.39 + 5.92(Wj) + Mu0j + rij

Question 5

```
m3 <- lmer(MTH_T ~ SES_Mean + SES + (1|teacherid), data = ecls.nomiss)
summary(m3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: MTH_T ~ SES_Mean + SES + (1 | teacherid)
##    Data: ecls.nomiss
##
## REML criterion at convergence: 18814.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0151 -0.6366  0.0098  0.6785  3.8427
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  teacherid (Intercept)  7.294   2.701
##  Residual              56.484   7.516
## Number of obs: 2705, groups:  teacherid, 300
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   51.6038     0.2331  309.0632 221.418  < 2e-16 ***
## SES_Mean       2.8597     0.5175  539.3367   5.526 5.12e-08 ***
## SES            2.9925     0.2627 2413.9128  11.390  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) SES_Mn
## SES_Mean -0.331
## SES       0.001 -0.508
```

```
tab_model(m3, show.aic = T, show.r2 = F, show.ci = F, show.se = T)
```

	Math T-Score		
Predictors	*Estimates*	*std. Error*	*p*
(Intercept)	51.60	0.23	**<0.001**
Class mean Socioeconomic status composite	2.86	0.52	**<0.001**
Socioeconomic status composite	2.99	0.26	**<0.001**
Random Effects			
σ^2	56.48		
$\tau_{00\ \text{teacherid}}$	7.29		
ICC	0.11		
$N_{\text{teacherid}}$	300		
Observations	2705		

AIC	18824.373

Gamma10 = **2.99**

With every one unit change in the L1 variable SES, there is a direct 2.99 unit change in Math T-Score.

AIC = **18824.37**

The interpretation of the grand mean intercept

Tau00 = **7.29**

sigma^2 = **56.48**

Both the intercept and residual variance decreased. This was expected as a result of the SES_Mean variable at the cluster level explaining L2 variance, while the individual level variable SES explained L1 variance.

ICC = sJPlot report of **0.11** model output manual calc below:

```
7.29/(7.29+56.48)
```

```
## [1] 0.1143171
```

Submodels and Reduced Form

L1: Yij = B0j + B1j(Xij) + rij

L2: B0j = Gam00 + Gam01(Wj) + Mu0j

```
B1j = Gam10 + Mu1j
```

Reduced: Yij = 51.60 + 2.86(Wj) + 2.99(Xij) + Mu0j + Mu1j(Xij) + rij

Question 6

The third model, with both L1 and L2 SES predictors, best fits the data, as was indicated by the lowest AIC value across the set of those reported.