Module 5 Exercise - Part A

```
# Initialize libraries
library(tidyverse)
library(rio)
library(plm)
library(broom)
library(stargazer)
library(lmtest)
library(ggthemes)
```

1. Preliminaries

a. Create a new RMarkdown PDF document in a *module5* folder of your course repository.

- Be sure to commit and push changes to your course repository after each question at a minimum (so after Q1, Q2, Q3).
- Set the following output options:
  - Data frames should be printed using *kable*.
  - Code should be highlighted using the *zenburn* theme.

b. Initialize the Ecdat library

- For the first part of the exercise, you will be working with the Ecdat package, which has many economics datasets. Hence you will need to install and initialize the package.

```
library(Ecdat)
```

2. Educational outcomes using the Project STAR dataset

a. Create the dataset

For the first part of the exercise, you will be working with data from Project STAR, an educational experiment conducted in Tennesse that examine the effect of student-teacher ratio on outcomes. The experiment randomized students into three class types: (1) a "regular" class size with 22-25 students, (2) a small class size with 13-17 students, and (3) a regular class size where the teacher was assisted by a teacher's aide.

a. Create an "star_data" data frame from the "Star" data in Ecdat as follows:

```
star_data <- Star
```

b. Turn the data set into a tibble

c. Select all variables except for "treadssk" to keep.

d. Rename the variables as follows:

- Change "tmathssk" to "math_score"
- Change "classk" to "class_type"
- Change "totexpk" to "teacher_exper"
- Change "totexpk" to "teacher_exper"
- Change "freelunk" to "free_lunch"
- Change "schidkn" to "school"

e. Create the following new variables:

- "log_math_score", equal to the log of "math_score"
- "teacher_exper_sq", equal to the square of "teacher_exper"

```
star_data <- star_data %>% as_tibble() %>%
  select(-treadssk) %>%
  rename(math_score = tmathssk, class_type = classk,
         teacher_exper = totexpk, free_lunch = freelunk,
         school = schidkn) %>%
  mutate(log_math_score = log(math_score),
         teacher_exper_sq = teacher_exper^2)
```

f. Convert the school variable to a factor.

```
star_data$school <- as.factor(star_data$school)
```

2. Perform regression

a. Create a regression object that regresses *logged math scores* on *teacher experience*, *experience*^{2}, *class type*, *sex*, *race*, *free lunch status*, and fixed effects for each *school*.

- First, use the *teacher_exper_sq* and *log_math_score* variables in the regression
- Then rewrite the regression to appropriately express the log and squared variables using only *math_score* and *teacher_exper* in the regression.

```
star_lm_prew <- lm(log_math_score ~ class_type + teacher_exper +
                     I(teacher_exper^2) + sex + race + free_lunch +
                     as.factor(school), data = star_data)

star_lm <- lm(log(math_score) ~ class_type + teacher_exper +
                I(teacher_exper^2) + sex + race + free_lunch +
                as.factor(school), data = star_data)
```

b. Display the summary output of the regression.

```
summary(star_lm)
```

```
##
## Call:
## lm(formula = log(math_score) ~ class_type + teacher_exper + I(teacher_exper^2) +
##     sex + race + free_lunch + as.factor(school), data = star_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32182 -0.05593 -0.00244  0.05202  0.30832
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  6.211e+00  1.112e-02 558.455  < 2e-16 ***
## class_typesmall.class        1.894e-02  2.788e-03   6.794 1.20e-11 ***
## class_typeregular.with.aide  1.859e-03  2.694e-03   0.690 0.490181
## teacher_exper                2.150e-04  7.042e-04   0.305 0.760178
## I(teacher_exper^2)           2.647e-05  3.134e-05   0.844 0.398457
## sexboy                      -1.382e-02  2.222e-03  -6.222 5.27e-10 ***
## raceblack                   -3.556e-02  4.777e-03  -7.443 1.13e-13 ***
## raceother                   -1.982e-02  1.630e-02  -1.216 0.223927
```

```
## free_lunchyes              -4.118e-02  2.691e-03 -15.302  < 2e-16 ***
## as.factor(school)2         -6.090e-02  1.560e-02  -3.903 9.62e-05 ***
## as.factor(school)3         -6.844e-03  1.344e-02  -0.509 0.610533
## as.factor(school)4         -5.666e-02  1.510e-02  -3.753 0.000176 ***
## as.factor(school)5         -6.616e-02  1.500e-02  -4.412 1.04e-05 ***
## as.factor(school)6         -4.192e-02  1.541e-02  -2.721 0.006525 **
## as.factor(school)7          4.076e-02  1.294e-02   3.150 0.001644 **
## as.factor(school)8         -3.824e-02  1.330e-02  -2.876 0.004044 **
## as.factor(school)9         -1.860e-02  1.285e-02  -1.448 0.147784
## as.factor(school)10         3.655e-02  1.560e-02   2.343 0.019143 *
## as.factor(school)11         4.227e-02  1.520e-02   2.781 0.005429 **
## as.factor(school)12         2.942e-02  1.530e-02   1.923 0.054543 .
## as.factor(school)13         5.707e-02  1.509e-02   3.781 0.000158 ***
## as.factor(school)14         1.419e-02  1.824e-02   0.778 0.436549
## as.factor(school)15        -8.573e-03  1.584e-02  -0.541 0.588277
## as.factor(school)16        -3.896e-02  1.385e-02  -2.813 0.004927 **
## as.factor(school)17         2.404e-03  1.446e-02   0.166 0.868007
## as.factor(school)18        -2.436e-02  1.425e-02  -1.710 0.087259 .
## as.factor(school)19         2.352e-02  1.411e-02   1.667 0.095565 .
## as.factor(school)20         5.461e-03  1.463e-02   0.373 0.709029
## as.factor(school)21         5.568e-02  1.439e-02   3.870 0.000110 ***
## as.factor(school)22         4.309e-02  1.338e-02   3.221 0.001285 **
## as.factor(school)23         9.157e-02  1.477e-02   6.200 6.04e-10 ***
## as.factor(school)24        -7.046e-03  1.525e-02  -0.462 0.644162
## as.factor(school)25        -2.276e-03  1.539e-02  -0.148 0.882455
## as.factor(school)26        -3.852e-02  1.632e-02  -2.361 0.018268 *
## as.factor(school)27         7.444e-02  1.327e-02   5.610 2.12e-08 ***
## as.factor(school)28        -5.851e-03  1.331e-02  -0.440 0.660220
## as.factor(school)29         3.210e-02  1.605e-02   2.000 0.045598 *
## as.factor(school)30         1.592e-01  1.579e-02  10.088  < 2e-16 ***
## as.factor(school)31         8.011e-02  1.628e-02   4.920 8.91e-07 ***
## as.factor(school)32        -4.230e-02  1.395e-02  -3.032 0.002439 **
## as.factor(school)33        -4.618e-02  1.415e-02  -3.264 0.001104 **
## as.factor(school)34        -6.203e-02  1.444e-02  -4.294 1.78e-05 ***
## as.factor(school)35        -4.502e-02  1.493e-02  -3.015 0.002584 **
## as.factor(school)36        -4.376e-02  1.478e-02  -2.961 0.003077 **
## as.factor(school)37        -8.029e-03  1.360e-02  -0.590 0.555036
## as.factor(school)38        -1.825e-02  1.632e-02  -1.118 0.263674
## as.factor(school)39        -1.139e-03  1.506e-02  -0.076 0.939736
## as.factor(school)40         3.797e-02  1.499e-02   2.532 0.011360 *
## as.factor(school)41         4.262e-02  1.464e-02   2.912 0.003608 **
## as.factor(school)42        -2.610e-02  1.564e-02  -1.669 0.095142 .
## as.factor(school)43        -2.623e-02  1.459e-02  -1.798 0.072282 .
## as.factor(school)44         6.205e-02  1.487e-02   4.173 3.05e-05 ***
## as.factor(school)45        -8.240e-02  1.547e-02  -5.325 1.05e-07 ***
## as.factor(school)46        -3.135e-02  1.549e-02  -2.024 0.042980 *
## as.factor(school)47        -1.522e-02  1.528e-02  -0.996 0.319192
## as.factor(school)48         1.187e-02  1.476e-02   0.804 0.421256
## as.factor(school)49         1.450e-02  1.512e-02   0.959 0.337782
## as.factor(school)50         1.847e-02  1.434e-02   1.289 0.197561
## as.factor(school)51         2.689e-03  1.256e-02   0.214 0.830536
## as.factor(school)52         2.202e-02  1.593e-02   1.383 0.166799
## as.factor(school)53        -5.729e-02  1.550e-02  -3.696 0.000221 ***
## as.factor(school)54        -1.975e-02  1.542e-02  -1.280 0.200487
```

```
## as.factor(school)55         -3.667e-02  1.410e-02  -2.601 0.009312 **
## as.factor(school)56         -1.118e-01  1.343e-02  -8.323  < 2e-16 ***
## as.factor(school)57         -5.237e-02  1.493e-02  -3.509 0.000454 ***
## as.factor(school)58          1.433e-02  1.348e-02   1.063 0.287988
## as.factor(school)59          1.143e-02  1.531e-02   0.747 0.455372
## as.factor(school)60         -3.838e-02  1.460e-02  -2.629 0.008583 **
## as.factor(school)61         -3.468e-02  1.437e-02  -2.414 0.015811 *
## as.factor(school)62         -3.459e-02  1.563e-02  -2.213 0.026946 *
## as.factor(school)63          2.230e-02  1.298e-02   1.718 0.085886 .
## as.factor(school)64         -3.992e-02  1.372e-02  -2.908 0.003646 **
## as.factor(school)65          3.940e-03  1.664e-02   0.237 0.812832
## as.factor(school)66         -5.659e-02  1.361e-02  -4.156 3.28e-05 ***
## as.factor(school)67         -7.003e-04  1.697e-02  -0.041 0.967078
## as.factor(school)68          4.112e-02  1.339e-02   3.070 0.002150 **
## as.factor(school)69          3.072e-02  1.529e-02   2.009 0.044560 *
## as.factor(school)70         -1.065e-02  1.367e-02  -0.779 0.435978
## as.factor(school)71         -3.548e-02  1.407e-02  -2.521 0.011714 *
## as.factor(school)72          2.704e-02  1.313e-02   2.059 0.039577 *
## as.factor(school)73         -1.082e-02  1.477e-02  -0.733 0.463644
## as.factor(school)74          1.866e-02  1.413e-02   1.321 0.186479
## as.factor(school)75          1.021e-02  1.384e-02   0.738 0.460812
## as.factor(school)76         -1.269e-02  1.348e-02  -0.941 0.346732
## as.factor(school)78         -5.257e-02  1.502e-02  -3.499 0.000470 ***
## as.factor(school)79         -9.992e-03  1.482e-02  -0.674 0.500092
## as.factor(school)80         -1.047e-02  1.535e-02  -0.682 0.495326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08357 on 5661 degrees of freedom
## Multiple R-squared:  0.2734, Adjusted R-squared:  0.2624
## F-statistic: 24.77 on 86 and 5661 DF,  p-value: < 2.2e-16
```

c. Use **coeftest** to produce a regression summary with HC1 robust standard errors, saving this to *star_reg_robust* (do not display it).

```
star_reg_robust <- coeftest(star_lm, vcov = vcovHC(star_lm, type="HC1"))
```

d. Display the *star_reg_robust* regression output with **stargazer**, setting the following options:

- **omit="school"** to omit the school fixed effects.
- **type="latex"** to produce Latex-type output.
- **header=FALSE** to hide the autogenerated Header description.
- **float=FALSE** (a Latex option to ensure the tables stays in the position it was placed).
- You will also probably need to add the option **results="asis"** to your code chunk so that the Latex output from stargazer will render in the Rmarkdown document.

```
stargazer(star_reg_robust, header=FALSE, type="latex", omit="school",
            float = FALSE)
```

	Dependent variable:
class_typesmall.class	0.019***
	(0.003)
class_typeregular.with.aide	0.002
	(0.003)
teacher_exper	0.0002
	(0.001)
I(teacher_exper^2)	0.00003
	(0.00003)
sexboy	−0.014***
	(0.002)
raceblack	−0.036***
	(0.005)
raceother	−0.020
	(0.016)
free_lunchyes	−0.041***
	(0.003)
Constant	6.211***
	(0.010)
Note:	*p<0.1; **p<0.05; ***p<0.01

e. Display the regression table wit HC1 robust standard errors using the **tidy** function.

- Pipe the tidy regression output to the following **stringr** expression in to remove the school fixed effects from the output.

```
filter(!str_detect(term, "(school)"))
```

```
tidy_star_reg_rob <- tidy(star_reg_robust) %>%
    filter(!str_detect(term, "(school)"))
tidy_star_reg_rob
```

term	estimate	std.error	statistic	p.value
(Intercept)	6.2108206	0.0099695	622.9828786	0.0000000
class_typesmall.class	0.0189421	0.0028585	6.6266926	0.0000000
class_typeregular.with.aide	0.0018588	0.0026379	0.7046443	0.4810606
teacher_exper	0.0002150	0.0007172	0.2997196	0.7644020
I(teacher_exper^2)	0.0000265	0.0000322	0.8211528	0.4115938
sexboy	-0.0138231	0.0022235	-6.2166918	0.0000000
raceblack	-0.0355585	0.0046948	-7.5739864	0.0000000
raceother	-0.0198249	0.0159525	-1.2427457	0.2140130
free_lunchyes	-0.0411840	0.0026894	-15.3136948	0.0000000

- Bonus (ie not required): Replace the variable names in "term" with names of your own.

```
tidy_star_reg_rob$term <-  recode(tidy_star_reg_rob$term,
    "(Intercept)" = "Intercept",
    "teacher_exper" = "Teacher experience",
    "I(teacher_exper^2)" = "Teacher experience^2^",
    "class_typesmall.class" = "Small class size",
    "class_typeregular.with.aide" = "Regular class w/ teacher aide",
    "sexboy" = "Male",
    "raceblack" = "Black",
    "raceother" = "Other race",
    "free_lunchyes" = "Student on freelunch"
  )
tidy_star_reg_rob
```

term	estimate	std.error	statistic	p.value
Intercept	6.2108206	0.0099695	622.9828786	0.0000000
Small class size	0.0189421	0.0028585	6.6266926	0.0000000
Regular class w/ teacher aide	0.0018588	0.0026379	0.7046443	0.4810606
Teacher experience	0.0002150	0.0007172	0.2997196	0.7644020
Teacher experience^2	0.0000265	0.0000322	0.8211528	0.4115938
Male	-0.0138231	0.0022235	-6.2166918	0.0000000
Black	-0.0355585	0.0046948	-7.5739864	0.0000000
Other race	-0.0198249	0.0159525	-1.2427457	0.2140130
Student on freelunch	-0.0411840	0.0026894	-15.3136948	0.0000000

3. Perform model diagnostics

Write the statistical decision from each test in your RMarkdown report:

a. Test for heteroskestacity.

```
bptest(star_lm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  star_lm
## BP = 201.82, df = 86, p-value = 2.638e-11
```

b. Test for missing polynomial terms.

```
resettest(star_lm)
```

```
##
##  RESET test
##
## data:  star_lm
## RESET = 2.3989, df1 = 2, df2 = 5659, p-value = 0.09091
```

4. Visualize the relationship between math scores and class type with a box plot.

a. Re-rerun the regression above, but omit *class_type* from the regressors.

```
star_lm <- lm(log_math_score ~ teacher_exper + I(teacher_exper^2)  + sex +
          race + free_lunch + as.factor(school), data = star_data)
```

b. Create a tibble with *class_type* and the residuals from (a).

```
star_adj <- tibble(residuals(star_lm), star_data$class_type)
```

c. Appropriately rename the columns of the tibble.

```
colnames(star_adj) <- c("resid_math_score", "class_type")
```

d. Create a bar plot from the tibble, displaying the average residual math score by class type

- Since we are not interested in a count of values for our bar plots, but instead the average values for the residual math score, use the following options inside of *geom_bar*: stat = "summary", fun.y = "mean".
- Adding a graph title, axis labels, and style it like a Stata graph using **theme_stata()** from the ggthemes package.

```
ggplot(star_adj, aes(x=class_type, y=resid_math_score)) +
  geom_bar(stat = "summary", fun.y = "mean") +
  ggtitle("Residualized Log Math Score by Class Type") +
  xlab("Class Type") + ylab("Residualized Math Score") +
  theme_stata()
```