14.320 PSET 4

Sitara Kumbale

Part I: Regression Theory

1. For this question, we are considering the following Regression model:

$$lnY_i = \alpha + \rho C_i + \gamma X_i + \epsilon_i$$

where Y_i is worker i's weekly earnings are 40, C_i is a dummy indicating college graduation, and X_i is i's family income when they were 16.

(a) To prove: ρ can be interpreted as measuring the percentage change in Y_i as a function of C_i, conditional on X_i

$$lnY_i = \alpha + \rho C_i + \gamma X_i + \epsilon_i$$

Taking expectation of the equation and bringing constants out,

$$E[lnY_i|C_i, X_i] = \alpha + \rho E[C_i|C_i, X_i] + \gamma E[X_i|C_i, X_i] + E[\epsilon_i|C_i, X_i]$$

Since $E[\epsilon_i|C_i, X_i] = 0$, we can replace this in all the equations. Since C_i is a dummy variable, we substitute and find the expectation at each of its values.

$$C_i = 0 : E[lnY_i|C_i = 0, X_i] = \alpha + \rho E[C_i|C_i = 0, X_i] + \gamma E[X_i|C_i = 0, X_i]$$
$$E[lnY_i|C_i = 0, X_i] = \alpha + \rho.0 + \gamma X_i \ldots (eqn1)$$

$$C_i = 1 : E[lnY_i|C_i = 1, X_i] = \alpha + \rho E[C_i|C_i = 1, X_i] + \gamma E[X_i|C_i = 1, X_i]$$
$$E[lnY_i|C_i = 1, X_i] = \alpha + \rho.1 + \gamma X_i \ldots (eqn2)$$

Subtracting eqn1 from eqn2,

$$E[lnY_i|C_i = 1, X_i] - E[lnY_i|C_i = 0, X_i] = \alpha + \rho + \gamma X_i - \alpha - \gamma X_i$$
$$E[lnY_i|C_i = 1, X_i] - E[lnY_i|C_i = 0, X_i] = \rho$$

Rewriting the conditional expectations as potential outcomes,

$$\rho = E[lnY_{1i}|X_i] - E[lnY_{0i}|X_i]$$
$$\rho = E[lnY_{1i} - lnY_{0i}|X_i]$$
$$\rho = E[ln(\frac{Y_{1i}}{Y_{0i}})|X_i]$$
$$\rho = E[ln(1 + \frac{Y_{1i} - Y_{0i}}{Y_{0i}})|X_i]$$

Using the result $ln(1 + x) \approx x$

$$\rho = E[\frac{Y_{1i} - Y_{0i}}{Y_{0i}}|X_i]$$

Hence, we have proved that ρ can be interpreted as measuring the percentage change in Y_i as a function of C_i, conditional on X_i.

We might want to condition on family income when measuring economic returns to a college education because family income maybe an omitted variable and so may bias the estimates of returns to college. It may bias the coefficient in the following ways:

- Family income may have a significant effect on where a person chooses to go to college and if they go to college at all. A person from a lower income family may choose to forgo college in exchange for a job instantly after school. A person from a very wealthy family maybe able to go to a wealthy private university because their family may contribute to the college's endowment.
- Family income may directly influence a person's income later in life. People from wealthy families may own generational investments and assets that may significantly increase their wealth. People from lower-income families may not have access to higher paying opportunities due to nepotism, etc.

Hence, controlling for family income eliminates the Omitted Variable Bias that may arise.

(b) If X_i was replaced by $ln(X_i)$, the equation can be rewritten as:

$$lnY_i = \alpha + \rho C_i + \gamma lnX_i + \epsilon_i$$

Taking the derivative of both sides with respect to X_i,

$$\frac{\partial(lnY_i)}{\partial X_i} = \frac{\partial(\alpha + \rho C_i + \gamma lnX_i + \epsilon_i)}{\partial X_i}$$
$$\frac{\partial(lnY_i)}{\partial X_i} = \frac{\partial \alpha}{\partial X_i} + \frac{\partial(\rho C_i)}{\partial X_i} + \frac{\partial(\gamma lnX_i)}{\partial X_i} + \frac{\partial \epsilon_i}{\partial X_i}$$

Splitting up the differentials, $\frac{\partial \alpha}{\partial X_i} = 0$ as it is a constant, $\frac{\partial C_i}{\partial X_i} = 0$ as we are holding C_i constant, $\frac{\partial \epsilon_i}{\partial X_i} = 0$ as regressors are uncorrelated with residuals in a regression. Substituting in,

$$\frac{\partial(lnY_i)}{\partial X_i} = \gamma \frac{1}{X_i}$$

Using Chain Rule,

$$\frac{\partial(lnY_i)}{\partial Y_i} \frac{\partial Y_i}{\partial X_i} = \gamma \frac{1}{X_i}$$
$$\frac{1}{Y_i} \frac{\partial Y_i}{\partial X_i} = \gamma \frac{1}{X_i}$$
$$\frac{\partial Y_i}{Y_i} = \gamma \frac{\partial X_i}{X_i}$$

For small values, $\partial Y_i = \Delta Y_i$ and , $\partial X_i = \Delta X_i$. Substituting this in,

$$\frac{\Delta Y_i}{Y_i} = \gamma \frac{\Delta X_i}{X_i}$$

This can be interpreted as follows: A 1% change in X_i (parental income) is associated with a γ % change in Y_i (economic returns to college) holding C_i constant.

2. The given model is:

$$ln(Y_i) = \alpha + \beta_1 C_i + \beta_2 W_i + \beta_{12}(C_i \times W_i) + \gamma_1 X_i + \gamma_{12}(X_i \times W_i) + \epsilon_i$$

W_i is a dummy which is 1 when the individual is a women and 0 when the individual is a man.

To test whether the economic returns to college are the same for men and women, we need to test the hypothesis $H_0 : \beta_{12} = 0$.

The economic returns for men who did not go to college is α, for women who did not go to college is $\alpha + \beta_2$, for men who went to college is $\alpha + \beta_1$. The differential effect of going to college between women and men is β_{12}, so we test if this coefficient is equal to zero. One way to do this is by testing if the t-statistic for the value is greater than the threshold at 1%, 5% and 10% levels.

3. We are given the following regression: $\epsilon_i = Y_i - \alpha - \beta X_i$ where $\hat{Y_i} = \hat{\alpha} + \hat{\beta} X_i$

(a) To Prove: $\sum_{i=1}^{n} \hat{Y_i} e = 0$

Substituting for $\hat{Y_i} = \hat{\alpha} + \hat{\beta} X_i$,

$$\sum_{i=1}^{n} e(\hat{\alpha} + \hat{\beta} X_i)$$
$$\sum_{i=1}^{n} (\hat{\alpha} e + \hat{\beta} X_i e)$$

Using linearity of summation, we can distribute the summation over each of the terms. Bringing the constants out,

$$\hat{\alpha} \sum_{i=1}^{n} e + \hat{\beta} \sum_{i=1}^{n} X_i e \ldots (Eqn1)$$

Minimizing the Sample Standard Error $\sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$ gives us an estimated value of $\hat{\alpha} = \bar{Y_n} - \hat{\beta} \bar{X_n}$ and $\hat{\beta} = \frac{S_{Y_i, X_i}}{S^2_{X_i}}$. Using the value of $\hat{\alpha}$ and $\hat{\beta}$ below,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}X_i)$$
$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - (\bar{Y_n} - \hat{\beta}\bar{X_n}) - \hat{\beta}X_i)$$
$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \bar{Y_n}) - \hat{\beta}\sum_{i=1}^{n}(X_i - \bar{X_n})$$
$$\sum_{i=1}^{n} e_i = 0 \ldots (Eqn2)$$

$$\sum_{i=1}^{n} e_i X_i = \sum_{i=1}^{n} X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i)$$
$$\sum_{i=1}^{n} e_i X_i = \sum_{i=1}^{n}(Y_i - \bar{Y_n}) - \hat{\beta}(X_i - \bar{X_n})$$
$$\sum_{i=1}^{n} e_i X_i = \sum_{i=1}^{n}(Y_i - \bar{Y_n}) - \frac{\sum_{i=1}^{n} X_i(Y_i - \bar{Y_n})}{\sum_{i=1}^{n} X_i(X_i - \bar{X_n})}(X_i - \bar{X_n})$$
$$\sum_{i=1}^{n} e_i X_i = 0 \ldots (Eqn3)$$

Using Eqn2 and Eqn3 in Eqn1,

$$\hat{\alpha}\sum_{i=1}^{n} e + \hat{\beta}\sum_{i=1}^{n} X_i e$$
$$\hat{\alpha} \times 0 + \hat{\beta} \times 0$$
$$\sum_{i=1}^{n} \hat{Y}_i e = 0$$

Hence Proved.

(b) To Prove: $s_Y^2 = s_{\hat{Y}}^2 + s_e^2$

We are given: $Y_i = \hat{Y}_i + e_i$. So, taking variance of either side of the equation,

$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * C(\hat{Y}_i, e_i)$$
$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * \sum_{i=1}^{n} (\hat{Y}_i - \bar{\hat{Y}}_n)(e_i - \frac{1}{n} \sum_{i=1}^{n} e_i)$$

From Eqn2 in part (a) above, we know $\sum_{i=1}^{n} e_i = 0$.

$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * \sum_{i=1}^{n} (\hat{Y}_i - \bar{\hat{Y}}_n) e_i$$
$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * \sum_{i=1}^{n} (\hat{Y}_i e_i - \bar{\hat{Y}}_n e_i)$$
$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * (\sum_{i=1}^{n} \hat{Y}_i e_i - \bar{\hat{Y}}_n \sum_{i=1}^{n} e_i)$$

Using Eqn2 and the final result of part (a) above, we know $\sum_{i=1}^{n} e_i = 0$ and $\sum_{i=1}^{n} \hat{Y}_i e = 0$. Substituting,

$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2 + 2 * (0 - \bar{\hat{Y}}_n .0)$$
$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2$$

Hence Proved.

4. (a) To prove: $\hat{\beta_w} = \frac{\bar{y_1} - \bar{y_0}}{\bar{x_1} - \bar{x_0}}$ is an unbiased estimator of β

We know that the regression is of the form: $Y_i = \alpha + \beta X_i + \epsilon_i$. Let us define an indicator $D_i = 1$ when i is part of sample 1 and $D_i = 0$ if i is in sample 0.

Taking expectations of regression equation conditional on D_i and bringing constants out of the expectations,

$$E[Y_i|D_i] = \alpha + \beta E[X_i|D_i] + E[\epsilon_i|D_i]$$

D_i is a function of X_i, defined $D_i = X_i > median(X_i)$. Since the CEF is linear, ϵ_i is independent of X_i or any function of X_i. So, $E[\epsilon_i|D_i] = 0$. Substituting this in,

$$E[Y_i|D_i] = \alpha + \beta E[X_i|D_i]$$

Since D_i is an indicator, it takes two values.

$$\bar{y_1} = E[Y_i|D_i = 1] = \alpha + \beta E[X_i|D_i = 1]$$
$$\bar{y_1} = \alpha + \beta \bar{x_1}$$

$$\bar{y_0} = E[Y_i|D_i = 0] = \alpha + \beta E[X_i|D_i = 0]$$
$$\bar{y_0} = \alpha + \beta \bar{x_0}$$

Substituting these values into the Wald Estimator Equation,

$$\hat{\beta_w} = \frac{\bar{y_1} - \bar{y_0}}{\bar{x_1} - \bar{x_0}}$$
$$\hat{\beta_w} = \frac{\alpha + \beta\bar{x_1} - \alpha - \beta\bar{x_0}}{\bar{x_1} - \bar{x_0}}$$
$$\hat{\beta_w} = \beta\frac{\bar{x_1} - \bar{x_0}}{\bar{x_1} - \bar{x_0}}$$
$$\hat{\beta_w} = \beta$$

Hence Proved.

(b) To find a formula for sampling variance of $\hat{\beta_w}$

$$Var(\hat{\beta_w}) = Var(\frac{\bar{y_1} - \bar{y_0}}{\bar{x_1} - \bar{x_0}})$$

Since X_i is fixed in repeated samples, we can bring the denominator outside the variance term.

$$Var(\hat{\beta_w}) = \frac{1}{(\bar{x_1} - \bar{x_0})^2} Var(\bar{y_1} - \bar{y_0})$$
$$Var(\hat{\beta_w}) = \frac{1}{(\bar{x_1} - \bar{x_0})^2} Var(\frac{\sum D_i Y_i}{n/2} - \frac{\sum(1 - D_i)Y_i}{n/2})$$

Distributing the variances due to linearity and bringing the $n/2$ term outside,

$$Var(\hat{\beta_w}) = \frac{4}{n^2(\bar{x_1} - \bar{x_0})^2} (Var \sum D_i Y_i - Var \sum(1 - D_i)Y_i)$$

Since X_i is fixed in repeated samples, the only variance comes from ϵ_i. So, we replace it in the equation,

$$Var(\hat{\beta_w}) = \frac{4}{n^2(\bar{x_1} - \bar{x_0})^2}(Var \sum D_i \epsilon_i - Var \sum (1 - D_i)\epsilon_i)$$
$$Var(\hat{\beta_w}) = \frac{4}{n^2(\bar{x_1} - \bar{x_0})^2}(\frac{n\sigma_\epsilon^2}{2} + \frac{n\sigma_\epsilon^2}{2})$$
$$Var(\hat{\beta_w}) = \frac{4\sigma_\epsilon^2}{n(\bar{x_1} - \bar{x_0})^2}$$

(c) The Gauss-Markov Theorem states that the ordinary least squares (OLS) estimator of the coefficients of a linear regression model is the best linear unbiased estimator (BLUE), that is, the estimator that has the smallest variance. Hence, $\hat{\beta_{OLS}}$ is the most precise estimator of β.

(d) We know that: $Var(\hat{\beta_{OLS}}) = \frac{\sigma_\epsilon^2}{ns_x^2}$ and from part (b) above, $Var(\hat{\beta_w}) = \frac{4\sigma_\epsilon^2}{n(\bar{x_1} - \bar{x_0})^2}$.

To prove that: $Var(\hat{\beta_{OLS}}) < Var(\hat{\beta_w})$

Rewriting what we need to prove,

$$Var(\hat{\beta_{OLS}}) < Var(\hat{\beta_w})$$
$$\frac{\sigma_\epsilon^2}{ns_x^2} < \frac{4\sigma_\epsilon^2}{n(\bar{x_1} - \bar{x_0})^2}$$
$$\frac{1}{s_x^2} < \frac{4}{(\bar{x_1} - \bar{x_0})^2}$$
$$s_x^2 > \frac{1}{4}(\bar{x_1} - \bar{x_0})^2 \ldots (Eqn1)$$

The auxiliary regression of X_i can be written as: $X_i = \delta + \phi D_i + \nu_i = \hat{X}_i + \nu_i$. Using the result from 3(b) of this PSET, $s_x^2 = s_{\hat{x}}^2 + s_\nu^2$. Since $s_\nu^2 > 0$, it is sufficient to prove that $s_{\hat{x}}^2 > \frac{1}{4}(\bar{x_1} - \bar{x_0})^2$.

$$s_{\hat{x}}^2 = \sum(\hat{X}_i - \bar{X}_n)^2$$
$$s_{\hat{x}}^2 = \sum(\delta + \phi D_i - \bar{X}_n)^2$$

Replacing $\bar{X}_n = \delta + \phi \bar{D}_n$,

$$s^2_{\hat{x}} = \sum (\delta + \phi D_i - \delta - \phi \bar{D}_n)^2$$
$$s^2_{\hat{x}} = \sum (\phi D_i - \phi \bar{D}_n)^2$$
$$s^2_{\hat{x}} = \phi^2 \sum (D_i - \bar{D}_n)^2$$
$$s^2_{\hat{x}} = \phi^2 s^2_D$$

Since D_i is a dummy, ϕ is the difference in conditional means, that is $\bar{x_1} - \bar{x_0}$. $s^2_D = p(1 - p)$ as D_i is a dummy variable. Hence, $s^2_D = p(1-p) = 0.5 * 0.5 = 0.25 = \frac{1}{4}$.
Substituting these in,

$$s^2_{\hat{x}} = \phi^2 s^2_D$$
$$s^2_{\hat{x}} = \frac{\bar{x_1} - \bar{x_0}}{4}$$

As previously stated, $s^2_\nu > 0$, hence,

$$s^2_x = s^2_{\hat{x}} + s^2_\nu > \frac{1}{4}(\bar{x_1} - \bar{x_0})^2$$

This implies, $Var(\hat{\beta_{OLS}}) < Var(\hat{\beta_w})$ from Eqn1.

Part II: Empirical Work

1. **CPS Data**

(a) Creating a potex variable, checking distribution and setting implausible values to missing:

```
*Generating the potential years of experience
variable
gen potex = age - years_ed - 6
label var potex "Potential experience"
tab potex
sum potex
replace potex=0 if potex<0
```

(b) The estimated regression model of log weekly wages (ln uwe) on race, potential experience and its square, and years of schooling is shown below:

```
. *Regressing log weekly wages on race, potential experience, potential experience squared and years of schooling
. reg ln_uwe i.race potex potex_sqr years_ed
```

Source	SS	df	MS			
				Number of obs	=	14,493
				F(7, 14485)	=	303.53
Model	1215.32806	7	173.618294	Prob > F	=	0.0000
Residual	8285.4902	14,485	.572004847	R-squared	=	0.1279
				Adj R-squared	=	0.1275
Total	9500.81826	14,492	.655590551	Root MSE	=	.75631

ln_uwe	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
race						
2	-.1434447	.0222521	-6.45	0.000	-.1870616	-.0998278
3	-.1038155	.0705761	-1.47	0.141	-.2421536	.0345226
4	-.0214184	.0370177	-0.58	0.563	-.0939777	.051141
5	-.0449161	.0919973	-0.49	0.625	-.2252425	.1354103
potex	.0282141	.003682	7.66	0.000	.020997	.0354313
potex_sqr	-.0004194	.0001	-4.20	0.000	-.0006153	-.0002235
years_ed	.1067557	.0023947	44.58	0.000	.1020617	.1114497
_cons	4.192587	.0469542	89.29	0.000	4.100551	4.284623

(i) In this case, the quadratic term maybe useful because the relationship between log weekly wages and potential experience may not be linear. For example, wages maybe a concave function due to diminishing marginal returns of potential experience on wages.

The t-value on the quadratic term potex_sqr is -4.20 (>1.96), which means that the coefficient (-0.0004), though small, is significantly different from zero at the 5% level.

The small value on the quadratic term represents that the potential experience function is weakly concave. This means that there is diminishing marginal returns of potential experience on wages, but this effect is very small. So, at lower levels of potential experience, the function behaves almost linearly.

(ii) Re-estimating the model using robust SEs:

```
. reg ln_uwe i.race potex potex_sqr years_ed, robust

Linear regression                               Number of obs     =     14,493
                                                F(7, 14485)       =     291.30
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1279
                                                Root MSE          =     .75631
```

ln_uwe	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
race						
2	-.1434447	.0199481	-7.19	0.000	-.1825456	-.1043438
3	-.1038155	.0677606	-1.53	0.126	-.2366349	.029004
4	-.0214184	.037238	-0.58	0.565	-.0944097	.051573
5	-.0449161	.0945574	-0.48	0.635	-.2302607	.1404286
potex	.0282141	.0036024	7.83	0.000	.0211529	.0352754
potex_sqr	-.0004194	.0000998	-4.20	0.000	-.0006149	-.0002239
years_ed	.1067557	.002436	43.82	0.000	.1019808	.1115306
_cons	4.192587	.0476902	87.91	0.000	4.099108	4.286066

Estimating with robust SEs did not change the t-value on the potex or potex_sqr term, so it did not change my interpretation as well. The coefficients on potex and potex_sqr are still significant at the 5% and 1% levels.

(c) The derivative of the CEF with respect to the potential experience can be written as:

$$0.02821 - (2 * 0.00042) * potex$$
$$0.02821 - 0.00084 * potex$$

As potex increases by 1 year, the derivative of the CEF decreases by 0.00084.

(d) The average marginal effect of potential experience on wages as computed below is 0.0138 and the standard error for the same is 0.0009.

```
. local av_potex = r(mean)

. lincom _b[potex]+ 2*_b[potex_sqr]* `av_potex'

 ( 1)  potex + 34.27427*potex_sqr = 0
```

ln_uwe	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
(1)	.0138394	.0008778	15.77	0.000	.0121189	.01556

(e) To calculate the peak earnings for high school and college graduates, we first solve for model fit condition:

$$0.02821 - 0.00084 * potex = 0$$
$$potex = 33.5833 \approx 34$$

For high school graduates (12 years of schooling), the wages will peak at $34 + 12 + 6 = 52$ years.

For college graduates (16 years of schooling), wages will peak at $34 + 16 + 6 = 56$ years.

Wages peak first for high school graduates, presumably because they enter the job market earlier than the college graduates.

(f) (i) Calculating the standard errors for two of the estimated peak earnings computed above:

High School Graduates: As shown below the standard error for the peak earnings is 3.973

```
.
. *Calculating Std Errors for two estimated peak earnings
. nlcom _b[potex]/(-2*_b[potex_sqr]) + 12 + 6

       _nl_1:  _b[potex]/(-2*_b[potex_sqr]) + 12 + 6
```

ln_uwe	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_nl_1	51.63618	3.973448	13.00	0.000	43.84836	59.42399

College Graduates: As shown below the standard error for the peak earnings is 3.973

```
. nlcom _b[potex]/(-2*_b[potex_sqr]) + 16 + 6

       _nl_1: _b[potex]/(-2*_b[potex_sqr]) + 16 + 6
```

ln_uwe	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_nl_1	55.63618	3.973448	14.00	0.000	47.84836	63.42399

.

The standard error for both high school and college graduate peak earnings is 3.973. The estimated standard error is very small compared to the mean peak age calculated. However, we also see that the confidence intervals for the two estimates overlap so this may mean that our estimate is not as precise that we wished it was!

(ii) The standard error of the difference between peak earnings of high school and college graduates is 0. This is because the difference is a constant and hence not a random variable.

2. **CPS Data**

(a) The model explored in empirical Q1b allowing the relationship between schooling and ln(ahe) to differ for men and women, including a female main effect is estimated below:

```
. *Creating an interaction between years of education and sex
. gen female = (sex == 2)

. gen femed = female*years_ed

. reg ln_ahe i.race potex potex_sqr years_ed female femed, robust

Linear regression                               Number of obs     =     14,375
                                                F(9, 14365)       =     425.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2174
                                                Root MSE          =     .56072
```

ln_ahe	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
race						
2	-.1209689	.0159518	-7.58	0.000	-.1522366	-.0897012
3	-.098331	.0487101	-2.02	0.044	-.193809	-.002853
4	-.0072946	.0292066	-0.25	0.803	-.0645433	.0499542
5	-.0677022	.0719749	-0.94	0.347	-.2087823	.0733778
potex	.0263655	.0027428	9.61	0.000	.0209893	.0317417
potex_sqr	-.0003789	.000075	-5.05	0.000	-.0005259	-.0002318
years_ed	.0884981	.0023302	37.98	0.000	.0839306	.0930656
female	-.4935055	.0482303	-10.23	0.000	-.5880432	-.3989679
femed	.0146327	.0035709	4.10	0.000	.0076332	.0216322
_cons	.9660547	.0402954	23.97	0.000	.8870705	1.045039

Using the model, we see that the coefficient on the interaction term (femed) is 0.0146. This means that the college educated women have a 1.5% higher log average hourly earnings compared to college educated men. The t-value of the coefficient is 4.10, which means that the results is statistically significant at the 1% and 5% levels. Hence, the returns to schooling are 1.5% higher for women than for men.

(b) The model explored in Q1b that allows the relationship between potential experience and ln(ahe) to differ for men and women is estimated below:

```
. *Creating an interaction between potex and sex
. gen fempotex = female*potex

. gen fempotex_sqr = female*potex_sqr

. reg ln_ahe potex potex_sqr female fempotex i.race years_ed fempotex_sqr, robust

Linear regression                               Number of obs     =     14,375
                                                F(10, 14364)      =     397.30
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2228
                                                Root MSE          =      .5588
```

ln_ahe	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
potex	.0397857	.0036112	11.02	0.000	.0327072	.0468641
potex_sqr	-.0005768	.0000988	-5.84	0.000	-.0007705	-.0003832
female	.0319653	.0445509	0.72	0.473	-.0553603	.1192909
fempotex	-.0274865	.0054634	-5.03	0.000	-.0381955	-.0167774
race						
2	-.1227367	.015945	-7.70	0.000	-.1539909	-.0914824
3	-.0977571	.0488744	-2.00	0.046	-.1935572	-.0019571
4	-.00445	.0292122	-0.15	0.879	-.0617097	.0528098
5	-.0593745	.0721561	-0.82	0.411	-.2008097	.0820607
years_ed	.0948539	.0018479	51.33	0.000	.0912318	.098476
fempotex_sqr	.0004024	.0001504	2.68	0.007	.0001076	.0006972
_cons	.7209157	.0407512	17.69	0.000	.6410381	.8007933

(i) Under this model, we see that the female main effect is no longer statistically significant as demonstrated by the t-value of 0.72 (<1.96). This is of economic interest as we can see that the wage gap between men and women is through the channel of experience profile. That is, women earn lesser than men with the same potential experience. That is, women do not start out with an initial lower salary, but have more diminishing returns to potential experience.

(ii) The results of the F-test are as follows:

```
.
. *Creating the F-Test
. test fempotex fempotex_sqr

 ( 1)  fempotex = 0
 ( 2)  fempotex_sqr = 0

       F(  2, 14364) =   61.68
            Prob > F =    0.0000
```

The F-test evaluates two restrictions - the female interaction on potex and female interaction on potex_sqr.

3. **Angrist, Oreopoulos, Williams (2014) replication:**

(a)(i) The results of AOW (2014) Table 1 all-applicant balance estimates and replication data are displayed below in columns titled as 'Control Mean' and 'Treatment Difference' respectively.

	From the Paper		Replication		
	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Treatment Difference (no strata)
Age	18.7	-0.012	18.67	-0.013	0.072
	[0.757]	(0.036)	[0.757]	(0.036)	(0.046)
High school grade average	82.5	-0.024	82.5	0.028	-0.102
	[6.44]	(0.134)	[6.439]	(0.132)	(0.385)
First language is English	0.416	0.009	0.416	0.008	0.007
	[0.493]	(0.031)	[0.493]	(0.031)	(0.03)
Mother a college graduate	0.439	0.020	0.439	0.019	0.021
	[0.496]	(0.031)	[0.496]	(0.031)	(0.03)
Father a college graduate	0.532	0.049	0.532	0.045	0.056
	[0.499]	(0.031)	[0.499]	(0.031)	(0.03)
Correctly answered harder questions	0.666	-0.014	0.666	-0.013	-0.001
	[0.472]	(0.029)	[0.472]	(0.029)	(0.029)
Controls who would have earned some scholarship	0.923		0.943		
	[0.266]		[0.233]		
Hypothetical earnings for controls	1330		1333.23		
	[1190]		1190.64		

(ii) The covariates gender, year in school, and high school grade quartile play the role of strata controls needed to ensure balance. The balance estimates without strata controls is shown in the table under the column titled "Treatment Difference (no strata)". In this case, the omission of strata controls should not matter for covariate balance. This is because OK participants were randomly assigned to treatment. We see that none of the covariates in the "Treatment Difference (no strata)" are significant except "Father a college graduate" weakly at the 10% level. Hence, we can say that effectively the strata controls do not matter for covariate balance so the randomization has worked!

(iii) The replication of AOW (2014) Table 4 is shown in the table below. The rows "Control mean" and "Treatment effect" are the ones showing the estimates from the paper and the replication estimates

	From the Paper			Replication		
	First-Year	**Second-Year**	**Second-Year**	**First-Year**	**Second-Year**	**Second-Year**
Control mean	68.4	71.2	69.6	68.45	71.19	69.57
	[10.6]	[7.99]	[9.7]	[10.6]	[7.99]	[9.7]
Treatment effect	-0.458	0.614	-0.025	-0.458	0.614	-0.025
	(0.745)	(0.719)	(0.522)	(0.745)	(0.719)	(0.522)
Treatment effect (no covariates)				-0.598	0.621	-0.046
				(0.768)	(0.734)	(0.536)

(iv) The covariates included to increase precision are: average high school grades, if students' first language is English, parents' education, and if students answered questions on program rules correctly. The estimates after removing these covariates is found in the row "Treatment effect (no covariates)". We see that the standard errors do not increase by much even when these covariates are removed from the model. Moreover, the statistical significance does not change even after these covariates are removed, which leads to the conclusion that they do not have much of an effect on the model.

(b) **Claim 1:** *"We might expect OK incentives to have been more powerful for financially constrained students. But treatment effects come out similar in subgroups defined by whether students expressed concerns about funding."*

```
.inear regression                              Number of obs     =      1,188
                                               F(18, 1169)       =      38.19
                                               Prob > F          =     0.0000
                                               R-squared         =     0.3583
                                               Root MSE          =      7.996
```

avggrade2008	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.T	.2970179	.8223921	0.36	0.718	-1.316512	1.910547
l.s_highfundsconcern	-1.0623	.5477	-1.94	0.053	-2.136885	.0122848
T#s_highfundsconcern						
1 1	-.6350155	1.07158	-0.59	0.554	-2.73745	1.467419
s_male	1.412823	.4998898	2.83	0.005	.432041	2.393604
s_first_year	-3.539239	.4717119	-7.50	0.000	-4.464736	-2.613743
percent_hsgrade						
2	.3940821	.9579139	0.41	0.681	-1.48534	2.273505
3	1.582143	1.44201	1.10	0.273	-1.247073	4.41136
4	3.7435	2.058846	1.82	0.069	-.295946	7.782947
s_hsgrade3	.6532596	.1258557	5.19	0.000	.4063313	.9001878
s_mtongue_english	-.7857493	.4825088	-1.63	0.104	-1.732429	.1609308
s_mothergraddegree	-2.151645	.9282367	-2.32	0.021	-3.972841	-.3304492
s_mothercolldegree	-.0011967	.6018949	-0.00	0.998	-1.182112	1.179718
s_motherhsdegree	-.2176242	.6862256	-0.32	0.751	-1.563996	1.128747
s_fathergraddegree	.3253023	.7459154	0.44	0.663	-1.13818	1.788785
s_fathercolldegree	.185482	.6543383	0.28	0.777	-1.098327	1.469291
s_fatherhsdegree	-.1489447	.7228035	-0.21	0.837	-1.567082	1.269192
s_test1correct	1.307865	.7890773	1.66	0.098	-.2403014	2.856031
s_test2correct	.6902657	.5539526	1.25	0.213	-.3965866	1.777118
_cons	15.67195	9.358644	1.67	0.094	-2.689667	34.03357

From the regression results, we see that the t-value on the interaction term (T#s_highfundsconcern), that is the differential effect of treatment on financially constrained vs unconstrained students is -0.59. This value shows that the coefficient is not statistically significant at the 5% level. Further we also see that the effect of treatment on financially unconstrained (T)students has a t-value 0.36, which is also not statistically significant. This shows that the treatment had no significant effect in general in shifting the average grades of either financially constrained students or financially unconstrained students, and supports the authors claims.

Claim 2: *"Effects are somewhat larger in the subsample of students whose parents had not been to college than among those with college-educated parents, but the gap by parents' schooling is not large or precisely estimated"*

Linear regression			
	Number of obs	=	1,203
	F(18, 1184)	=	37.73
	Prob > F	=	0.0000
	R-squared	=	0.3526
	Root MSE	=	8.0019

avggrade2008	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.T	.8253333	1.088728	0.76	0.449	-1.310719	2.961386
1.s_parentcol	1.644771	1.463621	1.12	0.261	-1.22681	4.516351
T#s_parentcol						
1 1	-1.019191	1.23591	-0.82	0.410	-3.444009	1.405627
s_male	1.547464	.4974933	3.11	0.002	.5713971	2.523531
s_first_year	-3.496894	.4699235	-7.44	0.000	-4.418869	-2.574918
percent_hsgrade						
2	.3969197	.9571617	0.41	0.678	-1.481003	2.274842
3	1.727523	1.440838	1.20	0.231	-1.099358	4.554404
4	3.826504	2.068926	1.85	0.065	-.2326654	7.885673
s_hsgrade3	.6474738	.1269442	5.10	0.000	.3984131	.8965345
s_mtongue_english	-.6829843	.4802894	-1.42	0.155	-1.625297	.2593289
s_mothergraddegree	-2.18381	.9232871	-2.37	0.018	-3.995272	-.3723492
s_mothercolldegree	-.1328451	.6326026	-0.21	0.834	-1.373992	1.108302
s_motherhsdegree	-.1960917	.6874345	-0.29	0.776	-1.544817	1.152634
s_fathergraddegree	.4368845	.7447409	0.59	0.558	-1.024275	1.898043
s_fathercolldegree	.2078451	.6602632	0.31	0.753	-1.087571	1.503261
s_fatherhsdegree	-1.238241	1.350671	-0.92	0.359	-3.888216	1.411734
s_test1correct	1.249276	.7575078	1.65	0.099	-.2369312	2.735483
s_test2correct	.6726665	.5514078	1.22	0.223	-.4091789	1.754512
_cons	14.95561	9.471889	1.58	0.115	-3.627951	33.53916

From the regression results, we see that the t-value on the interaction term (T#s_parentcol), that is the differential effect of treatment on people who's parents had been to college and those whose parents had not is -0.82. This value shows that the coefficient is not statistically significant at the 5% level. Further we also see that the effect of treatment on students whose parents did not go to college (T) has a t-value 0.76, which is also not statistically significant. This shows that the treatment had no significant effect in general in shifting the average grades of both students whose parents had gone to college and those whose parents hadn't, and supports the authors claims.

```
                                                  ___  ____  ____  ____  ____(R)
                                                 /__    /   ____/   /   ____/
                                                ___/   /   /___/   /   /___/
                                                  Statistics/Data analysis

                                                    User: 2pdf
--------------------------------------------------------------------------------------------------------
> ------------------------------
      name:  <unnamed>
       log:  C:\Users\user\Documents\MIT\14.320\PSET 4\B1_B2.log
  log type:  text
 opened on:  11 Apr 2023, 21:40:56

 1 . do "C:\Users\user\Documents\MIT\14.320\PSET 4\PartB1.do"

 2 . *PSET4 Part B Q1
 3 . *Author: Sitara Kumbale
 4 .
 5 . clear all

 6 . import delimited "C:\Users\user\Documents\MIT\14.320\PSET 4\ps4-1.csv"
   (encoding automatically selected: ISO-8859-1)
   (6 vars, 14,493 obs)

 7 .
 8 . summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      ln_uwe |     14,493    5.939816    .8096855   .6539265   10.59663
      ln_ahe |     14,493    2.287855    .6908567  -3.034953   7.041286
         age |     14,493    36.48582    7.140522         25         50
         sex |     14,493    1.479887    .4996125          1          2
        race |     14,493    1.212516    .6472649          1          5
-------------+---------------------------------------------------------
    years_ed |     14,493     13.3491    2.789107          0         22

 9 .
10 . *Replacing Weekly and Hourly wages with missing values if it is <0
11 . replace ln_ahe=. if ln_ahe<0
   (118 real changes made, 118 to missing)

12 . replace ln_uwe=. if ln_uwe<0
   (0 real changes made)

13 .
14 . *Generating the potential years of experience variable
15 . gen potex = age - years_ed - 6

16 . label var potex "Potential experience"

17 . tab potex

  Potential |
 experience |      Freq.     Percent        Cum.
------------+-----------------------------------
         -3 |          1        0.01        0.01
         -1 |          3        0.02        0.03
          0 |         10        0.07        0.10
          1 |         21        0.14        0.24
          2 |         22        0.15        0.39
          3 |        153        1.06        1.45
          4 |        174        1.20        2.65
          5 |        215        1.48        4.13
          6 |        321        2.21        6.35
          7 |        524        3.62        9.96
        7.5 |          8        0.06       10.02
          8 |        592        4.08       14.10
        8.5 |         10        0.07       14.17
```

9	556	3.84	18.01
9.5	10	0.07	18.08
10	641	4.42	22.50
10.5	15	0.10	22.60
11	634	4.37	26.98
11.5	6	0.04	27.02
12	597	4.12	31.14
12.5	10	0.07	31.21
13	637	4.40	35.60
13.5	9	0.06	35.67
14	643	4.44	40.10
14.5	6	0.04	40.14
15	684	4.72	44.86
15.5	8	0.06	44.92
16	659	4.55	49.47
16.5	8	0.06	49.52
17	594	4.10	53.62
17.5	6	0.04	53.66
18	627	4.33	57.99
18.5	9	0.06	58.05
19	604	4.17	62.22
19.5	6	0.04	62.26
20	560	3.86	66.12
20.5	2	0.01	66.14
21	555	3.83	69.96
21.5	7	0.05	70.01
22	547	3.77	73.79
22.5	8	0.06	73.84
23	548	3.78	77.62
23.5	4	0.03	77.65
24	485	3.35	81.00
24.5	4	0.03	81.03
25	426	2.94	83.96
25.5	3	0.02	83.99
26	426	2.94	86.92
26.5	1	0.01	86.93
27	376	2.59	89.53
27.5	3	0.02	89.55
28	334	2.30	91.85
28.5	4	0.03	91.88
29	269	1.86	93.73
29.5	5	0.03	93.77
30	275	1.90	95.67
30.5	3	0.02	95.69
31	253	1.75	97.43
31.5	4	0.03	97.46
32	196	1.35	98.81
32.5	2	0.01	98.83
33	42	0.29	99.12
34	36	0.25	99.37
35	32	0.22	99.59
36	21	0.14	99.73
37	14	0.10	99.83
38	13	0.09	99.92
39	4	0.03	99.94
40	6	0.04	99.99
42	2	0.01	100.00
Total	14,493	100.00	

```
18 . sum potex

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       potex |     14,493    17.13672    7.547176         -3         42

19 . replace potex=0 if potex<0
(4 real changes made)

20 .
21 . *Generating square of potential experience term
22 . gen potex_sqr = potex^2

23 .
24 . *Regressing log weekly wages on race, potential experience, potential experience squared and years o
> f schooling
25 . reg ln_uwe i.race potex potex_sqr years_ed

      Source |       SS           df       MS      Number of obs   =    14,493
-------------+----------------------------------   F(7, 14485)     =    303.53
       Model |  1215.32806         7  173.618294   Prob > F        =    0.0000
    Residual |   8285.4902    14,485  .572004847   R-squared       =    0.1279
-------------+----------------------------------   Adj R-squared   =    0.1275
       Total |  9500.81826    14,492  .655590551   Root MSE        =    .75631

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
          2  |  -.1434447   .0222521    -6.45   0.000    -.1870616   -.0998278
          3  |  -.1038155   .0705761    -1.47   0.141    -.2421536    .0345226
          4  |  -.0214184   .0370177    -0.58   0.563    -.0939777     .051141
          5  |  -.0449161   .0919973    -0.49   0.625    -.2252425    .1354103
             |
       potex |   .0282141    .003682     7.66   0.000      .020997    .0354313
   potex_sqr |  -.0004194      .0001    -4.20   0.000    -.0006153   -.0002235
    years_ed |   .1067557   .0023947    44.58   0.000     .1020617    .1114497
       _cons |   4.192587   .0469542    89.29   0.000     4.100551    4.284623
------------------------------------------------------------------------------

26 . reg ln_uwe i.race potex potex_sqr years_ed, robust

Linear regression                               Number of obs     =     14,493
                                                F(7, 14485)       =     291.30
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1279
                                                Root MSE          =     .75631

------------------------------------------------------------------------------
             |               Robust
      ln_uwe | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
          2  |  -.1434447   .0199481    -7.19   0.000    -.1825456   -.1043438
          3  |  -.1038155   .0677606    -1.53   0.126    -.2366349     .029004
          4  |  -.0214184    .037238    -0.58   0.565    -.0944097     .051573
          5  |  -.0449161   .0945574    -0.48   0.635    -.2302607    .1404286
             |
       potex |   .0282141   .0036024     7.83   0.000     .0211529    .0352754
   potex_sqr |  -.0004194   .0000998    -4.20   0.000    -.0006149   -.0002239
    years_ed |   .1067557    .002436    43.82   0.000     .1019808    .1115306
       _cons |   4.192587   .0476902    87.91   0.000     4.099108    4.286066
------------------------------------------------------------------------------
```

```
27 .
28 . *Calculating average marginal effect of potential experience
29 . quietly summarize potex if e(sample)

30 . local av_potex = r(mean)

31 . lincom _b[potex]+ 2*_b[potex_sqr]* `av_potex'

 ( 1)  potex + 34.27427*potex_sqr = 0

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   .0138394   .0008778    15.77   0.000     .0121189      .01556
------------------------------------------------------------------------------

32 .
33 . nlcom _b[potex]/(-2*_b[potex_sqr])

       _nl_1: _b[potex]/(-2*_b[potex_sqr])

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |   33.63618   3.973448     8.47   0.000     25.84836    41.42399
------------------------------------------------------------------------------

34 .
35 . *Calculating Std Errors for two estimated peak earnings
36 . nlcom _b[potex]/(-2*_b[potex_sqr]) + 12 + 6

       _nl_1: _b[potex]/(-2*_b[potex_sqr]) + 12 + 6

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |   51.63618   3.973448    13.00   0.000     43.84836    59.42399
------------------------------------------------------------------------------

37 . nlcom _b[potex]/(-2*_b[potex_sqr]) + 16 + 6

       _nl_1: _b[potex]/(-2*_b[potex_sqr]) + 16 + 6

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |   55.63618   3.973448    14.00   0.000     47.84836    63.42399
------------------------------------------------------------------------------

38 .
```

```
39 . nlcom  _b[potex]/(-2*_b[potex_sqr]) + 16 + 6-_b[potex]/(-2*_b[potex_sqr])-12 -6

         _nl_1: _b[potex]/(-2*_b[potex_sqr]) + 16 + 6-_b[potex]/(-2*_b[potex_sqr])-12 -6

------------------------------------------------------------------------------
      ln_uwe | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _nl_1 |          4          .       .       .            .           .
------------------------------------------------------------------------------

40 .
41 . *PSET4 Part B Q2
42 .
43 . *Creating an interaction between years of education and sex
44 . gen female = (sex == 2)

45 . gen femed = female*years_ed

46 . reg ln_ahe i.race potex potex_sqr years_ed female femed, robust

Linear regression                               Number of obs     =     14,375
                                                F(9, 14365)       =     425.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2174
                                                Root MSE          =     .56072

------------------------------------------------------------------------------
             |               Robust
      ln_ahe | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
          2  |  -.1209689   .0159518    -7.58   0.000    -.1522366   -.0897012
          3  |   -.098331   .0487101    -2.02   0.044     -.193809    -.002853
          4  |  -.0072946   .0292066    -0.25   0.803    -.0645433    .0499542
          5  |  -.0677022   .0719749    -0.94   0.347    -.2087823    .0733778
             |
       potex |   .0263655   .0027428     9.61   0.000     .0209893    .0317417
   potex_sqr |  -.0003789    .000075    -5.05   0.000    -.0005259   -.0002318
    years_ed |   .0884981   .0023302    37.98   0.000     .0839306    .0930656
      female |  -.4935055   .0482303   -10.23   0.000    -.5880432   -.3989679
       femed |   .0146327   .0035709     4.10   0.000     .0076332    .0216322
       _cons |   .9660547   .0402954    23.97   0.000     .8870705    1.045039
------------------------------------------------------------------------------

47 .
48 .
49 . *Creating an interaction between potex and sex
50 . gen fempotex = female*potex

51 . gen fempotex_sqr = female*potex_sqr
```

```
52 . reg ln_ahe potex potex_sqr female fempotex i.race years_ed fempotex_sqr, robust

Linear regression                               Number of obs     =     14,375
                                                F(10, 14364)      =     397.30
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2228
                                                Root MSE          =      .5588

------------------------------------------------------------------------------
             |               Robust
      ln_ahe | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       potex |   .0397857   .0036112    11.02   0.000     .0327072    .0468641
   potex_sqr |  -.0005768   .0000988    -5.84   0.000    -.0007705   -.0003832
      female |   .0319653   .0445509     0.72   0.473    -.0553603    .1192909
    fempotex |  -.0274865   .0054634    -5.03   0.000    -.0381955   -.0167774
             |
        race |
          2  |  -.1227367    .015945    -7.70   0.000    -.1539909   -.0914824
          3  |  -.0977571   .0488744    -2.00   0.046    -.1935572   -.0019571
          4  |    -.00445   .0292122    -0.15   0.879    -.0617097    .0528098
          5  |  -.0593745   .0721561    -0.82   0.411    -.2008097    .0820607
             |
    years_ed |   .0948539   .0018479    51.33   0.000     .0912318     .098476
fempotex_sqr |   .0004024   .0001504     2.68   0.007     .0001076    .0006972
       _cons |   .7209157   .0407512    17.69   0.000     .6410381    .8007933
------------------------------------------------------------------------------

53 .
54 . *Creating the F-Test
55 . test fempotex fempotex_sqr

 ( 1)  fempotex = 0
 ( 2)  fempotex_sqr = 0

       F(  2, 14364) =   61.68
            Prob > F =    0.0000

56 .
57 .
end of do-file
```

```
                                                      ___  ____  ____  ____  ____(R)
                                                     /__    /   ____/   /   ____/
                                                    ___/   /   /___/   /   /___/
                                                      Statistics/Data analysis

                                                        User: 2pdf
-----------------------------------------------------------------------------------------------------
> -----------------------------
      name:  <unnamed>
       log:  C:\Users\user\Documents\MIT\14.320\PSET 4\Untitled.log
  log type:  text
 opened on:  11 Apr 2023, 20:42:09

1 . do "C:\Users\user\AppData\Local\Temp\STDa58_000000.tmp"

2 . *14.320 PSET4 Part B Q3
3 . *Author: Sitara Kumbale
4 .
5 . clear all

6 . use OKgrades, clear

7 . summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
avggrade2008 |      1,203    69.72694    9.870112   18.57143       94.8
  s_hsgrade3 |      1,271    82.47005    6.408013         54         98
       s_age |      1,271    18.69315    .7604032         17         21
      s_male |      1,271     .350118    .4771945          0          1
s_mtongue_~h |      1,271    .4177813    .4933879          0          1
-------------+---------------------------------------------------------
s_first_year |      1,271    .5468135    .4979996          0          1
           T |      1,271    .3147128    .4645842          0          1
s_mothergr~e |      1,271    .0857592    .2801184          0          1
s_motherco~e |      1,271    .4453186    .4971966          0          1
s_motherhs~e |      1,271    .7773407    .4161952          0          1
-------------+---------------------------------------------------------
s_mothered~s |      1,271    .0928403    .2903227          0          1
s_fathergr~e |      1,271    .1738788    .3791544          0          1
s_fatherco~e |      1,271    .5491739    .4977719          0          1
s_fatherhs~e |      1,271    .7844217    .4113848          0          1
s_fathered~s |      1,271    .1093627    .3122166          0          1
-------------+---------------------------------------------------------
s_highfund~n |      1,256    .5684713    .4954868          0          1
s_test1cor~t |      1,271    .8465775    .3605361          0          1
s_test2cor~t |      1,271    .6656176    .4719598          0          1
s_second_y~r |      1,271    .4531865    .4979996          0          1
    s_female |      1,271     .649882    .4771945          0          1
-------------+---------------------------------------------------------
  strata_num |      1,271     7.47915    4.464841          1         16
controls_w~d |        821    4.920828    3.136694          0         11
controls_h~s |        821    1333.228    1190.636          0       5960

8 .
9 . tab avggrade2008

avggrade200 |
          8 |      Freq.     Percent        Cum.
------------+-----------------------------------
   18.57143 |          1        0.08        0.08
   25.83333 |          1        0.08        0.17
   28.85714 |          1        0.08        0.25
   29.85714 |          1        0.08        0.33
       30.2 |          1        0.08        0.42
   30.66667 |          1        0.08        0.50
     34.375 |          1        0.08        0.58
   34.66667 |          1        0.08        0.67
         38 |          2        0.17        0.83
```

39.83333	1	0.08	0.91
40.5	1	0.08	1.00
41.11111	1	0.08	1.08
41.33333	1	0.08	1.16
42.8	1	0.08	1.25
43.85714	1	0.08	1.33
44.33333	1	0.08	1.41
44.42857	1	0.08	1.50
44.71429	1	0.08	1.58
45	1	0.08	1.66
45.25	1	0.08	1.75
45.42857	1	0.08	1.83
46.1	1	0.08	1.91
46.125	1	0.08	2.00
47	1	0.08	2.08
47.125	1	0.08	2.16
47.25	1	0.08	2.24
47.33333	1	0.08	2.33
47.85714	1	0.08	2.41
48.33333	1	0.08	2.49
48.55556	1	0.08	2.58
48.625	1	0.08	2.66
49.16667	1	0.08	2.74
49.25	1	0.08	2.83
49.57143	1	0.08	2.91
49.77778	1	0.08	2.99
49.85714	1	0.08	3.08
49.88889	1	0.08	3.16
50.125	2	0.17	3.33
50.83333	1	0.08	3.41
51	1	0.08	3.49
51.25	2	0.17	3.66
51.77778	1	0.08	3.74
51.85714	1	0.08	3.82
52.1	1	0.08	3.91
52.11111	1	0.08	3.99
52.125	3	0.25	4.24
52.16667	1	0.08	4.32
52.22222	1	0.08	4.41
52.3	1	0.08	4.49
52.33333	1	0.08	4.57
52.55556	1	0.08	4.66
52.57143	1	0.08	4.74
52.625	2	0.17	4.90
52.66667	1	0.08	4.99
53	2	0.17	5.15
53.25	1	0.08	5.24
53.33333	1	0.08	5.32
53.42857	1	0.08	5.40
53.5	1	0.08	5.49
53.57143	1	0.08	5.57
53.875	1	0.08	5.65
54	2	0.17	5.82
54.125	1	0.08	5.90
54.14286	1	0.08	5.99
54.28571	1	0.08	6.07
54.44444	1	0.08	6.15
54.45454	1	0.08	6.23
54.5	2	0.17	6.40
54.55556	1	0.08	6.48
54.57143	1	0.08	6.57
54.625	1	0.08	6.65
54.83333	1	0.08	6.73

54.85714	1	0.08	6.82
54.875	1	0.08	6.90
54.88889	1	0.08	6.98
55	1	0.08	7.07
55.14286	2	0.17	7.23
55.33333	1	0.08	7.32
55.42857	2	0.17	7.48
55.44444	1	0.08	7.56
55.625	1	0.08	7.65
55.7	1	0.08	7.73
55.75	2	0.17	7.90
55.77778	1	0.08	7.98
55.83333	1	0.08	8.06
55.88889	1	0.08	8.15
56	1	0.08	8.23
56.1	1	0.08	8.31
56.125	1	0.08	8.40
56.14286	1	0.08	8.48
56.22222	1	0.08	8.56
56.33333	2	0.17	8.73
56.375	1	0.08	8.81
56.42857	2	0.17	8.98
56.5	1	0.08	9.06
56.55556	1	0.08	9.14
56.57143	2	0.17	9.31
56.625	1	0.08	9.39
56.66667	1	0.08	9.48
56.75	1	0.08	9.56
56.85714	1	0.08	9.64
57	3	0.25	9.89
57.16667	1	0.08	9.98
57.2	1	0.08	10.06
57.22222	2	0.17	10.22
57.33333	2	0.17	10.39
57.42857	1	0.08	10.47
57.5	1	0.08	10.56
57.55556	1	0.08	10.64
57.7	1	0.08	10.72
57.77778	1	0.08	10.81
57.83333	1	0.08	10.89
57.875	2	0.17	11.06
58	4	0.33	11.39
58.11111	1	0.08	11.47
58.25	1	0.08	11.55
58.28571	1	0.08	11.64
58.375	1	0.08	11.72
58.4	1	0.08	11.80
58.42857	1	0.08	11.89
58.5	1	0.08	11.97
58.57143	1	0.08	12.05
58.625	1	0.08	12.14
58.71429	1	0.08	12.22
58.75	1	0.08	12.30
58.9	1	0.08	12.39
59	4	0.33	12.72
59.125	1	0.08	12.80
59.14286	1	0.08	12.88
59.2	1	0.08	12.97
59.22222	1	0.08	13.05
59.25	2	0.17	13.22
59.3	1	0.08	13.30
59.33333	1	0.08	13.38
59.5	3	0.25	13.63

59.55556	1	0.08	13.72
59.625	1	0.08	13.80
59.66667	1	0.08	13.88
59.75	1	0.08	13.97
60	2	0.17	14.13
60.11111	1	0.08	14.21
60.125	1	0.08	14.30
60.14286	1	0.08	14.38
60.22222	1	0.08	14.46
60.25	1	0.08	14.55
60.28571	2	0.17	14.71
60.3	1	0.08	14.80
60.33333	1	0.08	14.88
60.375	1	0.08	14.96
60.5	3	0.25	15.21
60.625	1	0.08	15.30
60.66667	4	0.33	15.63
60.75	1	0.08	15.71
60.85714	3	0.25	15.96
61	5	0.42	16.38
61.125	1	0.08	16.46
61.22222	1	0.08	16.54
61.25	1	0.08	16.63
61.5	2	0.17	16.79
61.55556	2	0.17	16.96
61.57143	1	0.08	17.04
61.6	1	0.08	17.12
61.625	2	0.17	17.29
61.7	1	0.08	17.37
61.75	1	0.08	17.46
61.8	1	0.08	17.54
61.85714	2	0.17	17.71
61.875	1	0.08	17.79
62	6	0.50	18.29
62.125	1	0.08	18.37
62.16667	1	0.08	18.45
62.22222	2	0.17	18.62
62.25	4	0.33	18.95
62.28571	1	0.08	19.04
62.3	1	0.08	19.12
62.33333	2	0.17	19.29
62.375	1	0.08	19.37
62.44444	2	0.17	19.53
62.5	5	0.42	19.95
62.57143	1	0.08	20.03
62.6	1	0.08	20.12
62.625	2	0.17	20.28
62.66667	1	0.08	20.37
62.75	1	0.08	20.45
62.77778	1	0.08	20.53
62.85714	1	0.08	20.62
62.875	2	0.17	20.78
62.88889	2	0.17	20.95
63	7	0.58	21.53
63.11111	1	0.08	21.61
63.125	3	0.25	21.86
63.14286	1	0.08	21.95
63.25	1	0.08	22.03
63.28571	2	0.17	22.19
63.3	1	0.08	22.28
63.33333	3	0.25	22.53
63.375	1	0.08	22.61
63.4	2	0.17	22.78

63.42857	1	0.08	22.86
63.44444	1	0.08	22.94
63.5	3	0.25	23.19
63.55556	2	0.17	23.36
63.57143	1	0.08	23.44
63.66667	1	0.08	23.52
63.75	4	0.33	23.86
63.8	3	0.25	24.11
63.85714	3	0.25	24.36
63.875	1	0.08	24.44
63.9	1	0.08	24.52
64	4	0.33	24.85
64.1	2	0.17	25.02
64.11111	3	0.25	25.27
64.2	2	0.17	25.44
64.22222	1	0.08	25.52
64.25	1	0.08	25.60
64.28571	2	0.17	25.77
64.33334	3	0.25	26.02
64.375	1	0.08	26.10
64.4	3	0.25	26.35
64.44444	3	0.25	26.60
64.5	1	0.08	26.68
64.55556	1	0.08	26.77
64.66666	4	0.33	27.10
64.7	1	0.08	27.18
64.75	1	0.08	27.27
64.85714	2	0.17	27.43
64.875	1	0.08	27.51
64.9	1	0.08	27.60
65	5	0.42	28.01
65.125	1	0.08	28.10
65.14286	1	0.08	28.18
65.16666	1	0.08	28.26
65.22222	2	0.17	28.43
65.25	2	0.17	28.60
65.3	3	0.25	28.84
65.33334	3	0.25	29.09
65.375	2	0.17	29.26
65.4	5	0.42	29.68
65.42857	1	0.08	29.76
65.55556	2	0.17	29.93
65.57143	3	0.25	30.17
65.6	1	0.08	30.26
65.625	1	0.08	30.34
65.66666	1	0.08	30.42
65.7	1	0.08	30.51
65.75	2	0.17	30.67
65.77778	3	0.25	30.92
65.8	2	0.17	31.09
65.83334	1	0.08	31.17
65.85714	1	0.08	31.26
65.88889	2	0.17	31.42
66	8	0.67	32.09
66.1	2	0.17	32.25
66.11111	1	0.08	32.34
66.125	5	0.42	32.75
66.16666	1	0.08	32.83
66.2	5	0.42	33.25
66.22222	2	0.17	33.42
66.25	1	0.08	33.50
66.28571	2	0.17	33.67
66.3	3	0.25	33.92

66.375	1	0.08	34.00
66.4	2	0.17	34.16
66.5	6	0.50	34.66
66.55556	1	0.08	34.75
66.57143	1	0.08	34.83
66.6	3	0.25	35.08
66.66666	1	0.08	35.16
66.7	1	0.08	35.25
66.71429	1	0.08	35.33
66.75	1	0.08	35.41
66.77778	1	0.08	35.49
66.8	1	0.08	35.58
66.83334	1	0.08	35.66
66.85714	1	0.08	35.74
66.875	1	0.08	35.83
66.88889	1	0.08	35.91
66.9	2	0.17	36.08
67	5	0.42	36.49
67.09091	1	0.08	36.58
67.1	2	0.17	36.74
67.11111	1	0.08	36.82
67.125	3	0.25	37.07
67.2	1	0.08	37.16
67.25	4	0.33	37.49
67.27273	1	0.08	37.57
67.3	1	0.08	37.66
67.33334	1	0.08	37.74
67.375	2	0.17	37.91
67.42857	1	0.08	37.99
67.44444	2	0.17	38.15
67.5	4	0.33	38.49
67.55556	2	0.17	38.65
67.57143	1	0.08	38.74
67.625	2	0.17	38.90
67.66666	3	0.25	39.15
67.75	2	0.17	39.32
67.77778	2	0.17	39.48
67.8	3	0.25	39.73
67.85714	1	0.08	39.82
67.875	1	0.08	39.90
67.88889	1	0.08	39.98
67.9	1	0.08	40.07
68	9	0.75	40.81
68.1	1	0.08	40.90
68.11111	1	0.08	40.98
68.125	1	0.08	41.06
68.22222	1	0.08	41.15
68.25	1	0.08	41.23
68.28571	1	0.08	41.31
68.3	1	0.08	41.40
68.375	1	0.08	41.48
68.42857	2	0.17	41.65
68.44444	5	0.42	42.06
68.5	3	0.25	42.31
68.55556	1	0.08	42.39
68.57143	2	0.17	42.56
68.6	3	0.25	42.81
68.625	3	0.25	43.06
68.7	4	0.33	43.39
68.71429	1	0.08	43.47
68.75	3	0.25	43.72
68.77778	2	0.17	43.89
68.8	2	0.17	44.06

68.875	3	0.25	44.31
68.88889	1	0.08	44.39
68.9	1	0.08	44.47
69	8	0.67	45.14
69.1	2	0.17	45.30
69.125	1	0.08	45.39
69.2	1	0.08	45.47
69.28571	1	0.08	45.55
69.33334	3	0.25	45.80
69.375	2	0.17	45.97
69.4	2	0.17	46.13
69.42857	1	0.08	46.22
69.44444	2	0.17	46.38
69.5	3	0.25	46.63
69.55556	1	0.08	46.72
69.57143	1	0.08	46.80
69.6	1	0.08	46.88
69.625	1	0.08	46.97
69.7	2	0.17	47.13
69.71429	1	0.08	47.22
69.77778	2	0.17	47.38
69.8	3	0.25	47.63
69.83334	1	0.08	47.71
69.85714	1	0.08	47.80
69.875	1	0.08	47.88
69.88889	1	0.08	47.96
69.9	2	0.17	48.13
70	4	0.33	48.46
70.1	2	0.17	48.63
70.11111	1	0.08	48.71
70.125	2	0.17	48.88
70.2	2	0.17	49.04
70.22222	1	0.08	49.13
70.25	1	0.08	49.21
70.3	1	0.08	49.29
70.375	1	0.08	49.38
70.4	1	0.08	49.46
70.44444	3	0.25	49.71
70.5	5	0.42	50.12
70.55556	1	0.08	50.21
70.57143	1	0.08	50.29
70.6	1	0.08	50.37
70.66666	2	0.17	50.54
70.7	1	0.08	50.62
70.71429	6	0.50	51.12
70.75	5	0.42	51.54
70.77778	5	0.42	51.95
70.8	4	0.33	52.29
70.85714	1	0.08	52.37
70.875	2	0.17	52.54
70.88889	3	0.25	52.78
70.9	1	0.08	52.87
71	7	0.58	53.45
71.1	3	0.25	53.70
71.11111	1	0.08	53.78
71.125	4	0.33	54.11
71.2	4	0.33	54.45
71.22222	3	0.25	54.70
71.25	2	0.17	54.86
71.28571	1	0.08	54.95
71.3	2	0.17	55.11
71.33334	1	0.08	55.20
71.375	2	0.17	55.36

71.4	4	0.33	55.69
71.44444	1	0.08	55.78
71.5	5	0.42	56.19
71.6	3	0.25	56.44
71.625	2	0.17	56.61
71.66666	2	0.17	56.77
71.7	2	0.17	56.94
71.71429	1	0.08	57.02
71.75	2	0.17	57.19
71.77778	3	0.25	57.44
71.8	2	0.17	57.61
71.83334	1	0.08	57.69
71.85714	2	0.17	57.86
71.875	3	0.25	58.10
71.9	1	0.08	58.19
72	13	1.08	59.27
72.1	2	0.17	59.43
72.11111	2	0.17	59.60
72.125	2	0.17	59.77
72.2	1	0.08	59.85
72.22222	1	0.08	59.93
72.28571	1	0.08	60.02
72.33334	8	0.67	60.68
72.375	5	0.42	61.10
72.4	1	0.08	61.18
72.5	1	0.08	61.26
72.55556	2	0.17	61.43
72.57143	1	0.08	61.51
72.625	1	0.08	61.60
72.66666	2	0.17	61.76
72.7	1	0.08	61.85
72.77778	1	0.08	61.93
72.8	2	0.17	62.09
72.85714	2	0.17	62.26
72.88889	1	0.08	62.34
72.9	3	0.25	62.59
72.90909	1	0.08	62.68
73	5	0.42	63.09
73.1	1	0.08	63.18
73.11111	2	0.17	63.34
73.14286	1	0.08	63.42
73.2	2	0.17	63.59
73.22222	2	0.17	63.76
73.33334	5	0.42	64.17
73.375	2	0.17	64.34
73.4	2	0.17	64.51
73.5	6	0.50	65.00
73.55556	2	0.17	65.17
73.57143	1	0.08	65.25
73.625	2	0.17	65.42
73.66666	3	0.25	65.67
73.7	3	0.25	65.92
73.71429	1	0.08	66.00
73.77778	1	0.08	66.08
73.8	3	0.25	66.33
73.85714	1	0.08	66.42
73.875	1	0.08	66.50
73.88889	1	0.08	66.58
73.9	1	0.08	66.67
73.90909	1	0.08	66.75
74	4	0.33	67.08
74.11111	1	0.08	67.17
74.14286	1	0.08	67.25

74.22222	|	2	0.17	67.41
74.25	|	2	0.17	67.58
74.33334	|	1	0.08	67.66
74.375	|	1	0.08	67.75
74.4	|	6	0.50	68.25
74.42857	|	1	0.08	68.33
74.44444	|	1	0.08	68.41
74.5	|	4	0.33	68.74
74.55556	|	1	0.08	68.83
74.625	|	1	0.08	68.91
74.66666	|	1	0.08	68.99
74.7	|	1	0.08	69.08
74.75	|	1	0.08	69.16
74.77778	|	1	0.08	69.24
74.875	|	2	0.17	69.41
74.9	|	1	0.08	69.49
75	|	3	0.25	69.74
75.1	|	2	0.17	69.91
75.11111	|	3	0.25	70.16
75.125	|	2	0.17	70.32
75.2	|	4	0.33	70.66
75.28571	|	2	0.17	70.82
75.3	|	1	0.08	70.91
75.33334	|	2	0.17	71.07
75.375	|	2	0.17	71.24
75.4	|	1	0.08	71.32
75.41666	|	1	0.08	71.40
75.5	|	2	0.17	71.57
75.54546	|	1	0.08	71.65
75.55556	|	1	0.08	71.74
75.6	|	2	0.17	71.90
75.625	|	1	0.08	71.99
75.7	|	1	0.08	72.07
75.75	|	1	0.08	72.15
75.77778	|	2	0.17	72.32
75.875	|	2	0.17	72.49
75.88889	|	1	0.08	72.57
75.9	|	3	0.25	72.82
76	|	5	0.42	73.23
76.1	|	4	0.33	73.57
76.11111	|	2	0.17	73.73
76.14286	|	1	0.08	73.82
76.22222	|	2	0.17	73.98
76.25	|	1	0.08	74.06
76.33334	|	2	0.17	74.23
76.375	|	1	0.08	74.31
76.4	|	1	0.08	74.40
76.5	|	5	0.42	74.81
76.55556	|	1	0.08	74.90
76.57143	|	1	0.08	74.98
76.6	|	4	0.33	75.31
76.66666	|	2	0.17	75.48
76.7	|	1	0.08	75.56
76.77778	|	2	0.17	75.73
76.8	|	3	0.25	75.98
76.85714	|	1	0.08	76.06
76.88889	|	4	0.33	76.39
76.9	|	1	0.08	76.48
77	|	4	0.33	76.81
77.1	|	3	0.25	77.06
77.125	|	1	0.08	77.14
77.22222	|	2	0.17	77.31
77.25	|	1	0.08	77.39

77.3	1	0.08	77.47
77.375	2	0.17	77.64
77.4	1	0.08	77.72
77.44444	1	0.08	77.81
77.5	2	0.17	77.97
77.55556	1	0.08	78.05
77.625	1	0.08	78.14
77.66666	1	0.08	78.22
77.7	5	0.42	78.64
77.75	3	0.25	78.89
77.875	2	0.17	79.05
77.88889	1	0.08	79.14
78	5	0.42	79.55
78.1	3	0.25	79.80
78.11111	1	0.08	79.88
78.125	2	0.17	80.05
78.16666	1	0.08	80.13
78.18182	1	0.08	80.22
78.2	2	0.17	80.38
78.22222	1	0.08	80.47
78.25	2	0.17	80.63
78.3	1	0.08	80.71
78.33334	1	0.08	80.80
78.375	1	0.08	80.88
78.5	2	0.17	81.05
78.55556	2	0.17	81.21
78.6	1	0.08	81.30
78.625	1	0.08	81.38
78.66666	1	0.08	81.46
78.7	3	0.25	81.71
78.77778	1	0.08	81.80
78.8	2	0.17	81.96
78.88889	2	0.17	82.13
78.9	1	0.08	82.21
79	1	0.08	82.29
79.1	3	0.25	82.54
79.11111	1	0.08	82.63
79.125	3	0.25	82.88
79.2	7	0.58	83.46
79.28571	1	0.08	83.54
79.33334	3	0.25	83.79
79.375	2	0.17	83.96
79.4	2	0.17	84.12
79.42857	1	0.08	84.21
79.5	2	0.17	84.37
79.6	4	0.33	84.70
79.7	1	0.08	84.79
79.77778	3	0.25	85.04
79.8	2	0.17	85.20
79.875	1	0.08	85.29
79.88889	1	0.08	85.37
79.9	2	0.17	85.54
80	2	0.17	85.70
80.1	3	0.25	85.95
80.11111	1	0.08	86.03
80.125	1	0.08	86.12
80.2	2	0.17	86.28
80.22222	1	0.08	86.37
80.25	1	0.08	86.45
80.28571	1	0.08	86.53
80.375	1	0.08	86.62
80.4	3	0.25	86.87
80.5	1	0.08	86.95

80.57143	|	1	0.08	87.03
80.6	|	2	0.17	87.20
80.77778	|	1	0.08	87.28
80.9	|	1	0.08	87.36
81	|	3	0.25	87.61
81.1	|	2	0.17	87.78
81.125	|	2	0.17	87.95
81.2	|	2	0.17	88.11
81.25	|	1	0.08	88.20
81.3	|	2	0.17	88.36
81.375	|	1	0.08	88.45
81.4	|	1	0.08	88.53
81.44444	|	1	0.08	88.61
81.5	|	2	0.17	88.78
81.55556	|	1	0.08	88.86
81.6	|	2	0.17	89.03
81.7	|	2	0.17	89.19
81.75	|	1	0.08	89.28
81.77778	|	1	0.08	89.36
81.8	|	2	0.17	89.53
81.85714	|	1	0.08	89.61
81.88889	|	1	0.08	89.69
82	|	6	0.50	90.19
82.1	|	1	0.08	90.27
82.11111	|	1	0.08	90.36
82.125	|	2	0.17	90.52
82.2	|	2	0.17	90.69
82.28571	|	1	0.08	90.77
82.3	|	2	0.17	90.94
82.33334	|	2	0.17	91.11
82.375	|	2	0.17	91.27
82.4	|	1	0.08	91.35
82.5	|	4	0.33	91.69
82.55556	|	2	0.17	91.85
82.57143	|	1	0.08	91.94
82.625	|	1	0.08	92.02
82.66666	|	1	0.08	92.10
82.7	|	1	0.08	92.19
82.72727	|	1	0.08	92.27
82.75	|	2	0.17	92.44
82.9	|	1	0.08	92.52
83	|	5	0.42	92.93
83.1	|	1	0.08	93.02
83.14286	|	1	0.08	93.10
83.2	|	1	0.08	93.18
83.22222	|	2	0.17	93.35
83.4	|	1	0.08	93.43
83.5	|	1	0.08	93.52
83.66666	|	1	0.08	93.60
83.7	|	3	0.25	93.85
83.9	|	3	0.25	94.10
84	|	3	0.25	94.35
84.1	|	2	0.17	94.51
84.22222	|	1	0.08	94.60
84.3	|	1	0.08	94.68
84.375	|	1	0.08	94.76
84.4	|	2	0.17	94.93
84.42857	|	1	0.08	95.01
84.44444	|	1	0.08	95.10
84.5	|	1	0.08	95.18
84.55556	|	2	0.17	95.34
84.6	|	1	0.08	95.43
84.7	|	2	0.17	95.59

```
   84.71429 |          1        0.08       95.68
       84.8 |          1        0.08       95.76
       84.9 |          2        0.17       95.93
         85 |          2        0.17       96.09
       85.1 |          3        0.25       96.34
   85.14286 |          1        0.08       96.43
       85.3 |          2        0.17       96.59
   85.33334 |          2        0.17       96.76
       85.4 |          1        0.08       96.84
       85.5 |          2        0.17       97.01
       85.6 |          1        0.08       97.09
   85.66666 |          1        0.08       97.17
       85.8 |          1        0.08       97.26
         86 |          2        0.17       97.42
     86.125 |          1        0.08       97.51
   86.33334 |          1        0.08       97.59
     86.375 |          1        0.08       97.67
       86.4 |          1        0.08       97.76
       86.7 |          2        0.17       97.92
         87 |          1        0.08       98.00
     87.125 |          1        0.08       98.09
       87.4 |          2        0.17       98.25
       87.5 |          1        0.08       98.34
       87.6 |          1        0.08       98.42
       87.8 |          2        0.17       98.59
   87.88889 |          1        0.08       98.67
         88 |          3        0.25       98.92
   88.66666 |          1        0.08       99.00
         89 |          1        0.08       99.09
       89.1 |          1        0.08       99.17
       89.7 |          1        0.08       99.25
      89.75 |          1        0.08       99.33
       90.4 |          1        0.08       99.42
   90.83334 |          1        0.08       99.50
     91.875 |          1        0.08       99.58
       93.1 |          1        0.08       99.67
       93.4 |          1        0.08       99.75
       93.9 |          1        0.08       99.83
   94.66666 |          1        0.08       99.92
       94.8 |          1        0.08      100.00
------------+-----------------------------------
      Total |      1,203      100.00

10 .
11 . *Since only 1271 values exist as in the paper, this must be the full sample
12 .
13 . *Questions a-i and a-ii
14 . *Creating replication Table Structure
15 . matrix table1 = J(19,3,.)

16 . matrix rownames table1 = "Age" "" "High school grade average" "" "First language is English" "" "Mot
> her a college graduate" "" "Fa
> ther a college graduate" "" "Harder Qn correct " "" "Controls earn scholarship" "" "Hypothetical ear
> nings controls" ""
```

```
17 . matrix colnames table1 = "Control Mean" "Treatment Difference" "Treatment Difference (no strata)"

18 .
19 . *generating a dummy for the control of people earning some scholarship atleast
20 . gen scholarship_d = 1

21 . replace scholarship_d = 0 if controls_whoearned == .
   (450 real changes made)

22 . replace scholarship_d= . if T == 1
   (400 real changes made, 400 to missing)

23 .
24 . *Adding the Control Column stats to the table
25 . local i = 1

26 .
27 . foreach var in s_age s_hsgrade3 s_mtongue_englis s_mothercolldegree s_fathercolldegree s_test2correc
   > t scholarship_d controls_hyp_e
   > arnings {
     2.
28 . quietly summarize `var' if T==0
     3. matrix table1[`i',1] = r(mean)
     4. local i = `i' + 1
     5. matrix table1[`i',1] = r(sd)
     6. local i = `i' + 1
     7. }

29 .
30 . *Creating dummies for high school grade quartile
31 . xtile percent_hsgrade = s_hsgrade3, nquantiles(4)

32 .
33 . *Adding the Treatement Column stats to the table
34 . local i = 1

35 .
36 . foreach var in s_age s_hsgrade3 s_mtongue_englis s_mothercolldegree s_fathercolldegree s_test2correc
   > t {
     2. quietly reg `var' T s_male s_first_year i.percent_hsgrade, robust
     3. matrix table1[`i',2] = _b[T]
     4. local i = `i' + 1
     5. matrix table1[`i',2] = _se[T]
     6. local i = `i' + 1
     7. }

37 .
38 . *Adding the Treatment (no strata) Column stats to the table
39 . local i = 1

40 .
```

```
41 . foreach var in s_age s_hsgrade3 s_mtongue_englis s_mothercolldegree s_fathercolldegree s_test2correc
   > t {
     2. quietly reg `var' T, robust
     3. matrix table1[`i',3] = _b[T]
     4. local i = `i' + 1
     5. matrix table1[`i',3] = _se[T]
     6. local i = `i' + 1
     7. }

42 .
43 . matrix list table1, format(%8.3f)

   table1[19,3]
                 Control Mean  Treatment ~e  Treatment ~)
            Age        18.670        -0.013         0.072
             r2         0.757         0.036         0.046
   High schoo~e        82.502         0.028        -0.102
             r4         6.439         0.132         0.385
   First lang~h         0.416         0.008         0.007
             r6         0.493         0.031         0.030
   Mother a c~e         0.439         0.019         0.021
             r8         0.496         0.031         0.030
   Father a c~e         0.532         0.045         0.056
            r10         0.499         0.031         0.030
   Harder Qn ~          0.666        -0.013        -0.001
            r12         0.472         0.029         0.029
   Controls e~p         0.943             .             .
            r14         0.233             .             .
   Hypothetic~s      1333.228             .             .
            r16      1190.636             .             .
            r16             .             .             .
            r16             .             .             .
            r16             .             .             .

44 .
45 . *Exporting the Generated Table to Excel
46 . putexcel set Table1, modify

47 . putexcel A1 = matrix(table1), names nformat (number_d3)
   file Table1.xlsx saved

48 .
49 . *Questions a-iii and a-iv
50 . *Creating replication Table Structure
51 . matrix table4 = J(6,3,.)

52 . matrix rownames table4 = "Control mean" "" "Treatment effect" "" "Treatment effect (no covariates)"
   > ""

53 . matrix colnames table4 = "First-Year" "Second-Year" "Second-Year"
```

```
54 .
55 . *Adding in Control Column Stats to the table
56 . quietly summarize avggrade2008 if T == 0 & s_first_year == 1

57 . matrix table4[1,1] = r(mean)

58 . matrix table4[2,1] = r(sd)

59 .
60 . quietly summarize avggrade2008 if T == 0 & s_second_year == 1

61 . matrix table4[1,2] = r(mean)

62 . matrix table4[2,2] = r(sd)

63 .
64 . quietly summarize avggrade2008 if T == 0

65 . matrix table4[1,3] = r(mean)

66 . matrix table4[2,3] = r(sd)

67 .
68 . *Adding in Treatment Column Stats into the table
69 . quietly reg avggrade2008 T s_male i.percent_hsgrade s_hsgrade3 s_mtongue_englis s_mothergraddegree s
   > _mothercolldegree s_motherhsde
   > gree s_fathergraddegree s_fathercolldegree s_fatherhsdegree s_test1correct s_test2correct if s_first
   > _year == 1, robust

70 . matrix table4[3,1] = _b[T]

71 . matrix table4[4,1] = _se[T]

72 .
73 . quietly reg avggrade2008 T s_male i.percent_hsgrade s_hsgrade3 s_mtongue_englis s_mothergraddegree s
   > _mothercolldegree s_motherhsde
   > gree s_fathergraddegree s_fathercolldegree s_fatherhsdegree s_test1correct s_test2correct if s_secon
   > d_year == 1, robust

74 . matrix table4[3,2] = _b[T]

75 . matrix table4[4,2] = _se[T]

76 .
77 . quietly reg avggrade2008 T s_male s_first_year i.percent_hsgrade s_hsgrade3 s_mtongue_englis s_mothe
   > rgraddegree s_mothercolldegree
   >  s_motherhsdegree s_fathergraddegree s_fathercolldegree s_fatherhsdegree s_test1correct s_test2corre
   > ct, robust

78 . matrix table4[3,3] = _b[T]
```

```
 79 . matrix table4[4,3] = _se[T]

 80 .
 81 . *Adding in Treatment without covariates column stats into the table
 82 . quietly reg avggrade2008 T s_male i.percent_hsgrade if s_first_year == 1, robust

 83 . matrix table4[5,1] = _b[T]

 84 . matrix table4[6,1] = _se[T]

 85 .
 86 . quietly reg avggrade2008 T s_male i.percent_hsgrade if s_second_year == 1, robust

 87 . matrix table4[5,2] = _b[T]

 88 . matrix table4[6,2] = _se[T]

 89 .
 90 . quietly reg avggrade2008 T s_male s_first_year i.percent_hsgrade , robust

 91 . matrix table4[5,3] = _b[T]

 92 . matrix table4[6,3] = _se[T]

 93 .
 94 . matrix list table4, format(%8.3f)

    table4[6,3]
                   First-Year  Second-Year  Second-Year
    Control mean       68.445       71.191       69.572
              r2       10.602        7.988        9.704
    Treatment ~t       -0.391        0.636       -0.022
              r4        0.747        0.716        0.521
    Treatment ~)       -0.435        0.676        0.019
              r6        0.771        0.728        0.534

 95 .
 96 . *Exporting the generated table to Excel
 97 . putexcel set Table4, modify

 98 . putexcel A1 = matrix(table4), names nformat(number_d3)
    file Table4.xlsx saved

 99 .
100 . *Question b
101 . *Checking the claim for financially constained students
102 . reg avggrade2008 T##s_highfundsconcern s_male s_first_year i.percent_hsgrade s_hsgrade3 s_mtongue_en
    > glis s_mothergraddegree s_moth
    > ercolldegree s_motherhsdegree s_fathergraddegree s_fathercolldegree s_fatherhsdegree s_test1correct
    > s_test2correct, robust

    Linear regression                               Number of obs     =      1,188
                                                    F(18, 1169)       =      38.19
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.3583
                                                    Root MSE          =      7.996
```

```
------------------------------------------------------------------------------------
                     |               Robust
        avggrade2008 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
---------------------+--------------------------------------------------------------
                 1.T |   .2970179   .8223921     0.36   0.718    -1.316512    1.910547
1.s_highfundsconcern |    -1.0623      .5477    -1.94   0.053    -2.136885    .0122848
                     |
T#s_highfundsconcern |
                 1 1 |  -.6350155    1.07158    -0.59   0.554     -2.73745    1.467419
                     |
              s_male |   1.412823   .4998898     2.83   0.005      .432041    2.393604
        s_first_year |  -3.539239   .4717119    -7.50   0.000    -4.464736   -2.613743
                     |
     percent_hsgrade |
                   2 |   .3940821   .9579139     0.41   0.681     -1.48534    2.273505
                   3 |   1.582143    1.44201     1.10   0.273    -1.247073     4.41136
                   4 |     3.7435   2.058846     1.82   0.069     -.295946    7.782947
                     |
           s_hsgrade3 |   .6532596   .1258557     5.19   0.000     .4063313    .9001878
   s_mtongue_english |  -.7857493   .4825088    -1.63   0.104    -1.732429    .1609308
  s_mothergraddegree |  -2.151645   .9282367    -2.32   0.021    -3.972841   -.3304492
  s_mothercolldegree |  -.0011967   .6018949    -0.00   0.998    -1.182112    1.179718
    s_motherhsdegree |  -.2176242   .6862256    -0.32   0.751    -1.563996    1.128747
  s_fathergraddegree |   .3253023   .7459154     0.44   0.663     -1.13818    1.788785
  s_fathercolldegree |    .185482   .6543383     0.28   0.777    -1.098327    1.469291
    s_fatherhsdegree |  -.1489447   .7228035    -0.21   0.837    -1.567082    1.269192
      s_test1correct |   1.307865   .7890773     1.66   0.098    -.2403014    2.856031
      s_test2correct |   .6902657   .5539526     1.25   0.213    -.3965866    1.777118
               _cons |   15.67195   9.358644     1.67   0.094    -2.689667    34.03357
------------------------------------------------------------------------------------

103 .
104 . *Checking the claim for children whose parents had not been to college
105 . gen s_parentcol = 0

106 . replace s_parentcol = 1 if (s_mothercolldegree == 1 | s_fatherhsdegree == 1)
    (1,036 real changes made)

107 . reg avggrade2008 T##s_parentcol s_male s_first_year i.percent_hsgrade s_hsgrade3 s_mtongue_englis s_
    > mothergraddegree s_mothercolld
    > egree s_motherhsdegree s_fathergraddegree s_fathercolldegree s_fatherhsdegree s_test1correct s_test2
    > correct, robust

    Linear regression                               Number of obs     =      1,203
                                                    F(18, 1184)       =      37.73
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.3526
                                                    Root MSE          =     8.0019

------------------------------------------------------------------------------------
                   |               Robust
      avggrade2008 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------+----------------------------------------------------------------
               1.T |   .8253333   1.088728     0.76   0.449    -1.310719    2.961386
     1.s_parentcol |   1.644771   1.463621     1.12   0.261     -1.22681    4.516351
                   |
     T#s_parentcol |
               1 1 |  -1.019191    1.23591    -0.82   0.410    -3.444009    1.405627
                   |
            s_male |   1.547464   .4974933     3.11   0.002     .5713971    2.523531
      s_first_year |  -3.496894   .4699235    -7.44   0.000    -4.418869   -2.574918
                   |
   percent_hsgrade |
```

```
                 2  |   .3969197   .9571617     0.41   0.678    -1.481003    2.274842
                 3  |   1.727523   1.440838     1.20   0.231    -1.099358    4.554404
                 4  |   3.826504   2.068926     1.85   0.065    -.2326654    7.885673
                    |
       s_hsgrade3   |   .6474738   .1269442     5.10   0.000     .3984131    .8965345
  s_mtongue_english |  -.6829843   .4802894    -1.42   0.155    -1.625297    .2593289
 s_mothergraddegree |   -2.18381   .9232871    -2.37   0.018    -3.995272   -.3723492
 s_mothercolldegree |  -.1328451   .6326026    -0.21   0.834    -1.373992    1.108302
   s_motherhsdegree |  -.1960917   .6874345    -0.29   0.776    -1.544817    1.152634
 s_fathergraddegree |   .4368845   .7447409     0.59   0.558    -1.024275    1.898043
 s_fathercolldegree |   .2078451   .6602632     0.31   0.753    -1.087571    1.503261
   s_fatherhsdegree |  -1.238241   1.350671    -0.92   0.359    -3.888216    1.411734
     s_test1correct |   1.249276   .7575078     1.65   0.099    -.2369312    2.735483
     s_test2correct |   .6726665   .5514078     1.22   0.223    -.4091789    1.754512
              _cons |   14.95561   9.471889     1.58   0.115    -3.627951    33.53916
------------------------------------------------------------------------------------

108 .
    end of do-file
```