Question I

a) Regression model:

$$\widehat{lwage}_i = 0.4436 + 0.0709\, educ_i + 0.0389\, exper_i - 0.0006\, expersq_i + 0.1925\, union_i - 0.4421\, female_i$$

Implication on $educ_i$ is that when the year in education increases by one unit, hourly wage will increase by 0.0709 percent on average.

Test if education has impact on logarithm of hourly wage at $\alpha = 0.05$:

(i) $H_0: \hat{\beta}_{educ_i} = 0$, $H_1: \hat{\beta}_{educ_i} \neq 0$

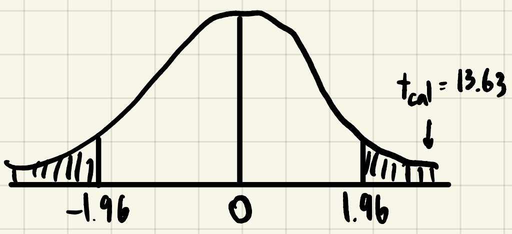
(ii) test statistics

$$t_{cal} = \frac{0.0709 - 0}{0.0052} = 13.6346$$

Critical value

$$t_{cri} = \pm 1.96$$

(iii)



$\because t_{cal} > 1.96$

$\therefore H_0$ rejected.

(iv) At 0.05 evel of significant, we can say for sure that β_{educ_i} is not zero i.e. number of year in education has impact on logarithm of hourly wage 95 times of 100 times we sample because the test statistics lies beyond the boundary.

b) Test overall significant of the model: F-test

(i) H_0: Every β_k are simultaneously equal to zero

H_1: otherwise

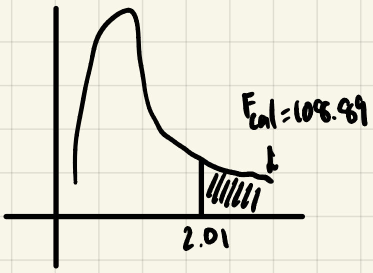
(ii) Test Statistics

$$F_{cal} = \frac{\frac{168.6972}{7}}{\frac{276.2828}{1,252}} = 109.2095$$

Critical value

$F_{cri} = 2.01$

(iii)



$\because F_{cal} > F_{cri}$

$\therefore H_0$ rejected

(iv) At $\alpha = 0.05$, we can say for sure that all the statistics are not simultaneously zero.

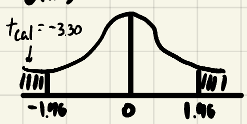
c) Test below average attractive:

(i) $H_0: \beta_{belavg} = 0$, $H_1: \beta_{belavg} \neq 0$

(ii) Test Statistics

$$t_{cal} = \frac{-0.1388 - 0}{0.0418} = -3.3014$$

(iii)



$t_{cri} = \pm 1.96$

(iv) Since $t_{cal} < t_{cri}$, we can say for sure that below average physically attractive has impact at $\alpha = 0.05$

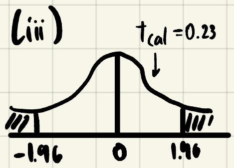
Test above average attractive:

(i) $H_0: \beta_{abvavg} = 0$, $H_1: \beta_{abvavg} \neq 0$

(ii) Test Statistics

$$t_{cal} = \frac{0.0070 - 0}{0.0303} = 0.2310$$

(iii)



$t_{cri} = \pm 1.96$

(iv) Since t_{cal} lies within the boundary, at $\alpha = 0.05$, we cannot say for sure that being above average attractive has impact.

d) No. According to the t-test of the previous question, it stated that being above average physical attractive has no impact on the natural logarithm of hourly wage at 0.05 level of significant. For example, picking up 2 randomly, one with above average attractiveness, and another women with below average attractiveness, both of them will receive the same natural logarithm of hourly wage.

Question II

a) Yes. Because of two reasons. First is living outside the town is usually cheaper than living inside the town. Thus, having more child age under 15 who cannot work will increase household expenditure.

b) Test for significant of area:

(i) $H_0: \beta_{area_i} = 0$, $H_1: \beta_{area_i} \neq 0$

(ii) Test statistics

$t_{cal} = -15.8$

Critical value

$t_{cri} = \pm 2.576$

(iii) $t_{cal} > t_{cri}$

$\therefore H_0$ rejected

(iv) At 0.01 level of significant, we can say for sure that area is statistically significant.

Test for significant of child:

(i) $H_0: \beta_{child_i} = 0$, $H_1: \beta_{child_i} \neq 0$

(ii) Test Statistics

$t_{cal} = 6.82$

Critical Value

$t_{cri} = \pm 2.576$

(iii) $t_{cal} > t_{cri}$

$\therefore H_0$ rejected

(iv) At 0.01 level of significant, we can say for sure that number of children age below 15 is statistically significant.

c)
$$\begin{aligned} \widehat{hhexp_i} &= 9{,}736 - 2{,}835\,area_i + 881\,child_i \\ &= 9{,}736 - 2{,}835(1) + 881(3) \\ &= 9{,}544 \end{aligned}$$

d) <u>Test significance</u>

Let $\beta_{area_i} = \beta_2$, $\beta_{child_i} = \beta_3$, $(area * child) = \beta_4$

(i) H_0: $\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$

H_1: $\beta_2 \neq 0$, $\beta_3 \neq 0$, $\beta_4 \neq 0$

(ii) Test Statistics

$t_{cal\beta_2} = -6.55$, $t_{cal\beta_3} = 5.17$, $t_{cal\beta_4} = -0.25$

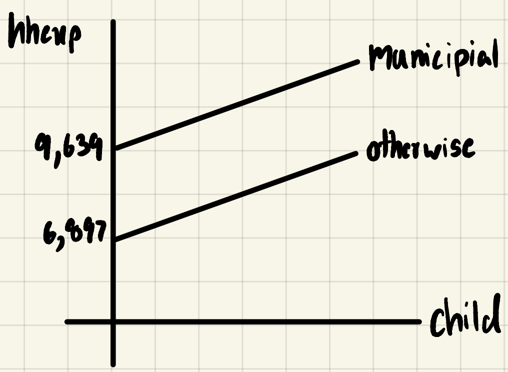
$t_{cri} = \pm 2.576$

(iii) $t_{cal\beta_2} < t_{cri}$ $\therefore$ H_0 rejected.

$t_{cal\beta_3} > t_{cri}$ $\therefore$ H_0 rejected.

$t_{cal\beta_4}$ lies within the boundary $\therefore$ H_0 is not rejected.

(iv) At 0.01 level of significant, we can say for sure that area and number of children age under 15 are significant while the interaction term is not significant.



SRF of otherwise

$hhexp_i = 9{,}639 - 2{,}742(1) + 910\,child_i$
$= 6{,}897 + 910\,child_i$

Question III

a) If there is some pair of varibles that have VIF exceed 10. These pair of variable could be troublesome. In this case a pair of age and age squared may have multicolinearity problem as VIF of this pair is about 50.61 and 50.68

b) No. Since both age and age square provide economic sense for the model. Working hour per week should increase as a person gets older. Then, when he reaches a certain age the value of age square becomes higher and causes working hours per week to decrease.

c) No. This model has no observable relationship betwen $weekot_i$ and the error term squared i.e. the size of error square does not relatively change corresponding to the change in $weekot_i$.

d) <u>Test for heteroscedasticity</u>

(i) H_0: homoscedasticity , H_1: heteroscedasticity

(ii) Test Statistics

$$F_{cal} = \frac{\frac{R^2_{aux}}{k}}{\frac{(1-R^2_{aux})}{n-k-1}} = \frac{\frac{0.184}{5}}{\frac{1-0.184}{2{,}032-5-1}} = 7.5954$$

$F_{cri} = 2.2141$

(iii) $F_{cal} > F_{cri}$

$\therefore H_0$ rejected

(iv) At 0.05 level of significant, we can reject homoscedasticity for this model. In other words, heteroscedasticity is present at 95 percent confidence interval.