1.a) ① Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). ② Interpret the estimated coefficients associated with $educ_i$. Based on Model (1.1), ③ test whether education has an impact on logarithm of hourly wage. Show your work. (Use α = 0.05)

① $\log(wage_i) = 0.4436 + 0.0708\,educ_i + 0.03898\,exper_i$
$- 0.000598\,expersq + 0.1925\,union_i - 0.44216\,female_i$

② if years of school increase by 1 year, logarithm of hourly wage will increase by 0.0708 dollar.

③ • H_0 ; $\beta_2 = 0$; null hypothesis
H_a , $\beta_2 \neq 0$

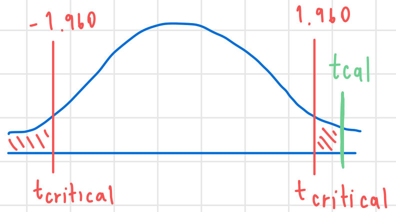
• $t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2}$

$= \dfrac{0.07085 - 0}{0.00523}$

$= 13.5468$ //

• $\alpha = 0.05$ d.f = 1254

$t_{critical} = 1.960$



∴ We can reject null hypothesis and we can make sure 95% that education impact on logarithm of wage.

1.b) What is the overall significance of the regression from Model (1.2)? What test do you use? (Use α = 0.05)

used : F-test

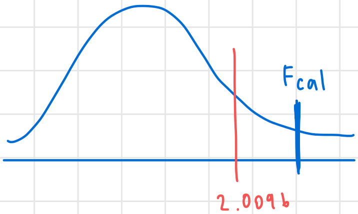
H_0 ; $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$

H_a ; otherwise

$$F_{cal} : \frac{Ess/k-1}{Rss/n-k} = \frac{168.69715/7}{276.282816/1252} = 109.5436$$

α : 0.05

$F_{cri}(7, 1252) = 2.0096$



∴ We can reject null hypothesis and we can make sure 95% that overall variables are significant in 1.2 model.

1.c) If we are interested in testing whether "physical attractiveness" has an impact on logarithm of hourly wage at all, what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use $\alpha = 0.05$)

: F-test ; marginal contribution

H_0 : physical attractiveness has no impact on logarithm of hourly wage ;

H_a : otherwise

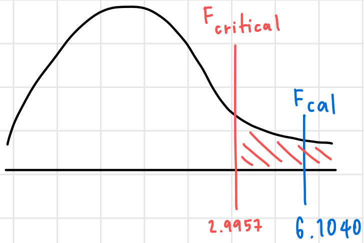
$$F_{cal} = \frac{Ess_{1.2} - Ess_{1.1} / 2}{Rss_{1.2} / n - u_{1.2}}$$

$$= \frac{(168.697151) - (166.011412) / 2}{(276.282816) / 1260 - 8}$$

$$= \frac{1.3429}{0.22} = 6.1040$$

$\alpha = 0.05$

$F_{cri}(2, 1252) = 2.9957$



∴ We can reject H_0

1.d) Is there convincing evidence that women with above average looks earn more than women with average looks? Explain.

as (1.2) $\log(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female_i + \beta_7 \underbrace{belavg_i}_{\text{below}} + \beta_8 \underbrace{abvavg_i}_{\text{above}} + u_i$

- woman, average looks → $belavg_i = 0$ $abvag_i = 0$

  $\log(wage) = 0.4737 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i - 0.4388235(1) + 0 + 0$

  $\log(wage) = 0.034877 + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union$ - woman with average looks

- woman (1) above average (1)

  $\log(wage) = 0.4737 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i - 0.4388235(1) + 0.0070104(1)$

  $\log(wage) = 0.48071 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i$ - woman with above rage looks.

∴ Explain: the intercept of 2 groups are different, which woman above rate looks have higher value of intercept.

2.a) Do all the signs for each coefficient make economic sense? Explain.

$$\underset{\beta_1}{\widehat{hhexp_i}} = \overset{\beta_1}{9{,}736} - \overset{\beta_2}{2{,}835} area_i + \overset{\beta_3}{881} child_i + \hat{u}_i$$

$$(43.83)\quad(-15.8)\qquad(6.82)$$

- β_2 is nagative value, this means not living in municiple have lower expense by 2,835, this make economic sense because people who living in municiple area are likely to have less expense due to lower income and lower cost of living.
- β_3 is positive, this means the more children family have, the more expense in that house. This is obvious that more people consume food and money for living

2.b) Test each parameter separately if they are significantly different from zero or not. (Use α = 0.01)

$$\widehat{hhexp}_i = 9{,}693 - 2{,}742area_i + 910child_i - 64(area_i * child_i) + \hat{u}_i$$
$$(34.38) \quad (-6.55) \qquad (5.17) \qquad (-0.25)$$

$H_0 : \beta_2 = 0$ - null hypothesis

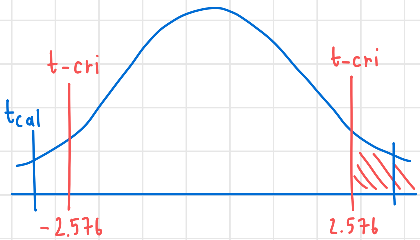
$H_1 : \beta_2 \neq 0$

$\alpha : 0.01 \rightarrow \frac{\alpha}{2} = 0.005$

$t_{cri} = 2.576$

$df = 14908 - 3$

$= 14905$



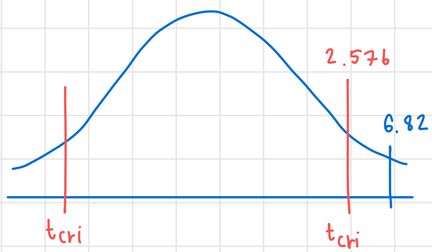
$\therefore$ We can reject H_0

$\alpha = 0.01 \rightarrow \frac{\alpha}{2} = 0.005$

$t_{cri} = 2.576$

$df = 14908 - 3$

$= 14905$



$\therefore$ We can reject H_0

2.c) Find the expected value of a household expenditure not living in a municipal area with 3 children aged under 15.

- HH expenditure municipal area = 1 , no. children = 3

$$\widehat{hhexp}_i = 9{,}736 - 2{,}835area_i + 881child_i + \hat{u}_i$$

$\widehat{hhexp}_i = 9{,}736 - 2{,}835 + 881(3)$

$= 9{,}544$ //

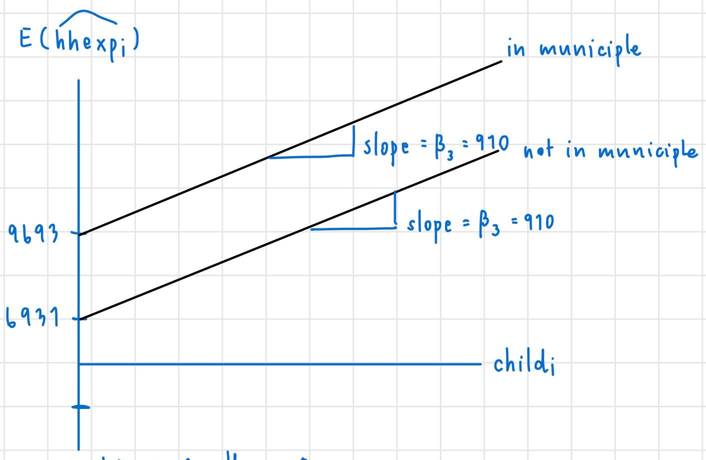
2.d) When an interaction term is included in this model, the result becomes with **t value in parentheses.**

$$\widehat{hhexp_i} = 9{,}693 - 2{,}742area_i + 910child_i - 64(area_i * child_i) + \hat{u}_i$$
$$(34.38) \quad (-6.55) \qquad (5.17) \qquad (-0.25)$$

Draw a diagram for this model displaying sampled regression functions (SRF) with expected value of household expenditure on the vertical axis and number of children on the horizontal axis, taking **only significant parameter(s)** into account. Indicate the intercept and slope for each SRF where applicable. Testing of significance can be shortened.

$$\widehat{hhexp_i} = 9{,}693 - 2{,}742area_i + 910child_i - 64(area_i * child_i) + \hat{u}_i$$
$$(34.38) \quad (-6.55) \qquad (5.17) \qquad (-0.25)$$

$H_0 ; \hat{\beta}_2 = 0$	$H_0 ; \beta_3 = 0$	$H_0 ; \beta_4 = 0$
$H_a ; \hat{\beta}_2 \neq 0$	$H_a ; \beta_3 \neq 0$	$H_a ; \beta_4 \neq 0$
$t_{cal} = -6.55$	$t_{cal} = 5.17$	$t_{cal} = -0.25$
$t_{cri} = 2.576$	$t_{cri} = 2.576$	$t_{cri} = 2.576$
$\alpha = 0.01$	can reject H_0	cannot reject H_0
$\frac{\alpha}{2} = 0.005$		
can reject H_0		



mun = 1 other = 0

$$\widehat{hhexp_i} = \overset{\beta_1}{9{,}693} - \overset{\beta_2}{2{,}742}area_i + \overset{\beta_3}{910}child_i - \overset{\beta_4}{64}(area_i * child_i) + \hat{u}_i$$

$\widehat{hhexp_i} = 9693 - 2742(0) + 910(0) = 9693$; base case : in munuciple

$\widehat{hhexp_i} = 9693 - 2742(1) + 910(0) = 6951$; no kid : not in munuciple

$\widehat{hhexp_i} = 9693 - 2742(0) + 910(1) = 10603$ or $\beta_1 + \beta_3(child_i)$

$\widehat{hhexp_i} = 9693 - 2742(1) + 910(1) = 7797$

$= (\beta_1 + \beta_2) + (\beta_3)(child_i)$

3.a) A VIF and tolerance table (postestimation) is given below

Variable	VIF	1/VIF
2.sex	1.02	0.979129
age	50.61	0.019759
agesq	50.68	0.019731
weekot	1.01	0.985618
Mean VIF	25.83	

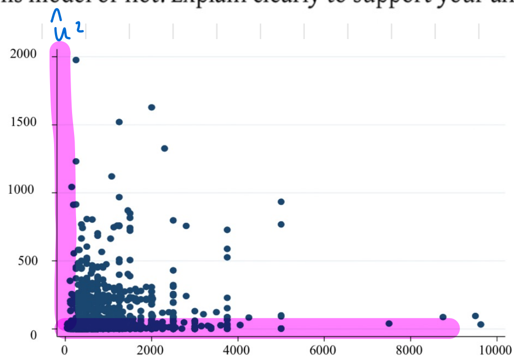
Given that you are exploring multicollinearity assumption, which pair of variables that you suspect they might be linearly correlated? Provide clear explanation what criteria (ion) that you rely on making that judgement.

: Age and ageeq are suspected to be linearly correlation.
because the VIF value of these 2 variables are exceed 10 and TOL value or $\frac{1}{VIF}$ are nearly 0, this means r^2 between these 2 independences are high.

3.b) From **(3.a)**, do you consider removing one of the variables from the model? Why or why not and which one that you choose to remove, if that is the case?

: No, we don't have other data to make sure that which variable should be eliminated.

3.c) The graph provided below is a scatter plot between $\hat{u}_i^2$ (vertical axis) and $weekot_i$ (horizontal axis). Using the graphical method, do you conclude that heteroscedasticity is present in this model or not. Explain clearly to support your answer.



: Yes, heteroscedas is detected in this model because as weekot increase, there are increase in $\hat{u}^2$ as well.

3.d) An auxiliary model here is estimated and the result is given in the table below.

$$\hat{u}_i^2 = \beta_1 + \beta_2 sex_i + \beta_3 age_i + \beta_4 agesq_i + \beta_5 weekot_i + v_i$$

Source	SS	df	MS
Model	829063.863	4	207265.966
Residual	44148135	2,027	21780.037
Total	44977198.8	2,031	22145.3465

Number of obs	=	2,032
F(4, 2027)	=	9.52
Prob > F	=	0.0000
R-squared	=	0.0184
Adj R-squared	=	0.0165
Root MSE	=	147.58

uhat2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
2.sex	-5.648899	6.630832	-0.85	0.394	-18.65286	7.355058
age	-2.490434	2.37094	-1.05	0.294	-7.140168	2.1593
age2	.044175	.0301279	1.47	0.143	-.0149098	.1032599
weekot	.0229916	.0043502	5.29	0.000	.0144603	.0315229
_cons	83.8484	44.4418	1.89	0.059	-3.307973	171.0048

From the table, setup the hypotheses and perform the Breusch-Pagan test to check that heteroscedasticity is present in the original model or not.

H_0 : the model is Homoscedasticity
H_a : otherwise

$$F_{cal} = \frac{R^2_{\hat{u}^2_i} / (k)}{(1-R^2_{\hat{u}^2_i})/(n-k-1)} = \frac{(0.0184)/5}{(1-0.0184)/(2032-5-1)} = \frac{0.00368}{0.00048} = 7.6667$$

$F_{cri}(5,2026) = 2.2141$

$\alpha = 0.05$

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

We can reject H_0, this means that 95 times of observation, Heteroscedast is present.