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

$$\log(wage_i) = 0.4436 + 0.0709\, educ_i + 0.3898\, exper_i - 0.06\, exper_i^2 + 0.1925\, union_i - 0.4422\, female_i + U_i$$

Based on the coefficient of educi, if a person has one more year of schooling, his/her wage would increase by 100×0.0709% or 7.09%

Test whether 0.0709 = 0 at $\alpha = 0.05$

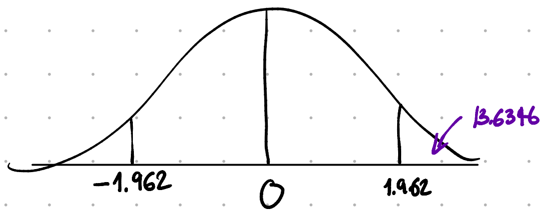
$H_0: 0.0709 = 0$

$H_1: 0.0709 \neq 0$

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

$df = 1260 - 6 = 1254$ $\quad T_{upper} = 1.962$

$T_{lower} = -1.962$



(Reject H_0)

Ans: Education has a significant impact on log of wage at 95% significant level.

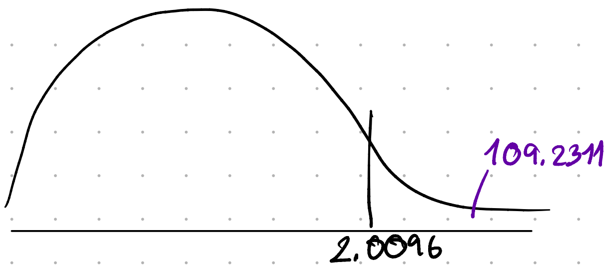
1.b) What is the overall significance of the regression from Model (1.2)? What test do you use? (Use α = 0.05)

use f test

$H_0 \rightarrow B_2 = B_3 = B_4 = B_5 = B_6 = B_7 = B_8 = 0$

$H_1 \rightarrow$ otherwise

$$F_{cal} = \frac{ESS/k-1}{RSS/n-k} = \frac{168.6972/7}{276.2828/1252} = 109.2311$$



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

Ans: The regression $\alpha = 0.05$

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 α = 0.05)

use t test $\alpha = 0.05$ $df = 1252$ $t_{crit} = \pm 1.962$

$H_0 \rightarrow B_7 = 0$

$H_1 \rightarrow B_7 \neq 0$

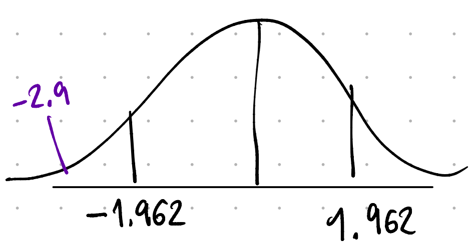
$$t_{cal}(B_7) = \frac{-0.1388 - 0}{0.0478} = -2.9$$

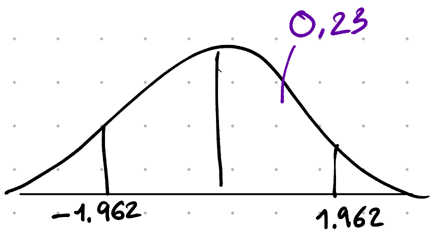
$H_0 \rightarrow B_8 = 0$

$H_1 \rightarrow B_8 \neq 0$

$$t_{cal}(B_8) = \frac{0.007 - 0}{0.0303} = 0.23$$

ANS: physical attractiveness has an impact on log of hourly wage but only for below average attractiveness. above average physical attractiveness has no impact on log of wage/hr





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

No, based on question (1.C), above average physical attractiveness has no impact on log of wage/hr
Let's say that these are 2 women, one with average attractiveness and another with above average, both of them will receive the same log of wage/hr at $\alpha = 0.05$

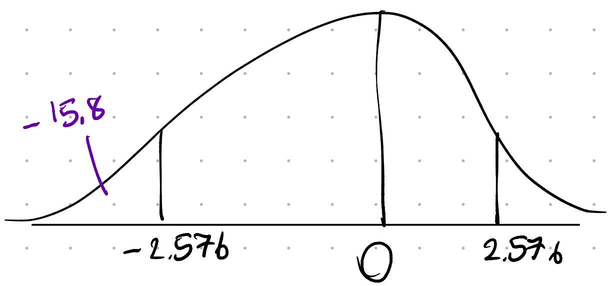
2.a) Do all the signs for each coefficient make economic sense? Explain.

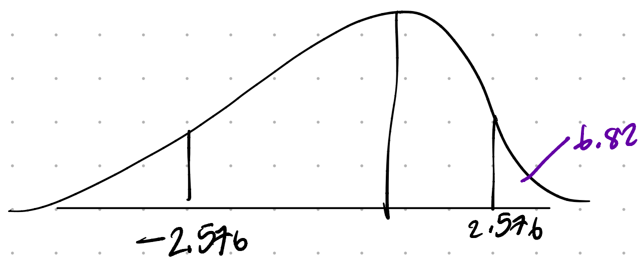
Yes, incase of -2,835 area; (area (municipality) = 0, otherwise = 1), meaning that households outside municipal areas will have less expenditure, which make sense because the cost of living outside municipal areas is cheaper. Also, +881 child; make sense because more children always correlated to more expenditure

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

use t test $\quad df = 14{,}908 - 3 = 14{,}905 \quad \alpha = 0.01 \quad t_{crit} = \pm 2.576$

$H_0 \rightarrow 2835 = 0$	$H_0 \rightarrow 881 = 0$
$H_1 \rightarrow 2835 \neq 0$	$H_1 \rightarrow 881 \neq 0$
$t_{cal} = -15.8$	$t_{cal} = 6.82$





Ans: B_2 or 2835 is significantly different from 0 at $\alpha = 0.01$ //

B_3 or 881 " ______________________

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

$$h\,hexp = 9736 - 2835\,area_i + 881\,child_i$$

$$h\,hexp = 9736 - 2835(1) + 811(3)$$

$$hhexp = 9334$$

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

$\hat{B}_1$ $\hat{B}_2$ $\hat{B}_3$ $\hat{B}_4$

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

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.

Test significant

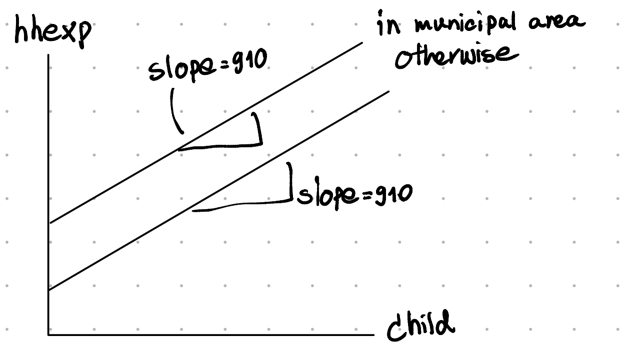
$t_{crit} = \pm 2.576$ (the same because a decrease in df by 1 won't make a difference)

$t_{cal}(B_2) = -6.55$ $t_{cal}(B_3) = 5.17$ $t_{cal}(B_4) = -0.25$

$\therefore B_4$ is insignificant from zero, the interaction term is insignificant

$\because B_2$ & B_3 are significant from zero

Diagram



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.

$VIF = \frac{1}{1-r^2}$ VIF high $\rightarrow$ r^2 closer to 1

$\downarrow$

more multicolinearity

ANS: age & agesq due to high r^2, which means higher coefficient of correlation between the 2 variables and higher multicolinearity

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, because we need both age and agesq for the model to make economic sense, since hours worked/week should increase with age at younger age, and you get older, the value of agesq become larger and caused hours worked/week to decrease

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.

No because there is not a clear correlation between $weekot_i$ and $\hat{u}_i^2$, ie when weekot increase, $\hat{u}_i^2$ doesn't increase with it

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	Number of obs	=	2,032
Residual	44148135	2,027	21780.037	F(4, 2027)	=	9.52
				Prob > F	=	0.0000
				R-squared	=	0.0184
				Adj R-squared	=	0.0165
Total	44977198.8	2,031	22145.3465	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.

$H_0 \rightarrow$ Homoscedaticity

$H_1 \rightarrow$ Heteroscedaticity

$$F_{cal} = \frac{R^2_{\hat{u}_i^2}/K}{(1-R^2_{\hat{u}_i^2})/n-K-1)} = \frac{0.0814/5}{(1-0.0184/2082-5-1)} = 7.6$$

$F_{crit} = 2.2141$

$F_{cal} > F_{crit} \rightarrow$ Reject homascedaticity

Ans: Heteroticity is present