6304641894 FERNC

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)

$$l(\hat{wage}_i) = \hat{\beta}_1 + \hat{\beta}_2 educ_i + \hat{\beta}_3 exper_i + \hat{\beta}_4 expersq_i + \hat{\beta}_5 union_i + \hat{\beta}_6 female_i + \hat{u}_i$$

use t-test : $H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$$t_{calc}(\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.7085 - 0}{0.0052} = 136.25$$

$n-k = 1260-6$

$= 1254$

$$t_{(\frac{\alpha}{2} = 0.025,\ 1254)} = 1.96$$

$\therefore t_{calc} > t_{cri} \rightarrow$ reject H_0

We are sure that β_2 is not zero 95 out of 100 times when we sample.

Hence, education has an impact on logarithm of hourly wage.

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

H_1 : otherwise

$$F_{calc} = \frac{ESS / df}{RSS / df} = \frac{168.6972 / 7}{276.2828 / 1252} = 108.5621$$

$$F_{cri};\ \alpha = 0.05\ (7, 1252) = 2.01$$

$F_{calc} > F_{cri}$ $\therefore$ reject H_0

$\therefore$ we can make sure that 95 out of 100 times

$\beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$, and β_8 are not simultaneously equal to zero.

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) β_7 & β_8

use F-test : $H_0 : \beta_7 = \beta_8 = 0$ (Physical attractiveness has no marginal contribution to the model)

H_1 : otherwise

$$F_{calc} = \frac{(\underbrace{ESS_{new}}_{\text{model 2}} - \underbrace{ESS_{old}}_{\text{model 1}}) / \#\text{ of new regressors}}{RSS_{new} / (n - k_{new})}$$

$$= \frac{(168.6972 - 166.0114) / 2}{276.2828 / 1252} = 6.0855$$

$$F_{cri};\ \alpha = 0.05\ (2, 1252) = 3$$

$F_{calc} > F_{cri}$ $\therefore$ Reject H_0

$\therefore$ we can make sure that 95 out of 100 times

Physical attractiveness has marginal contribution to the model

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

Since 1.c) we found that physical attractiveness has marginal contribution to the model and from 1.b) that $\beta_1 - \beta_8$ are not simultaneously equal to zero at 95% CI. Hence, we can interpret that when $abvag_i$ increases 1 unit, $l\hat{wage}_i$ will increase by 0.007 unit as the data given in table 1.2)

$$\widehat{hhexp}_i = 9{,}736 - 2{,}835area_i + 881child_i + \hat{u}_i$$
$$(43.83) \quad (-15.8) \qquad (6.82)$$

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

Yes, since households in municipal area tend to have higher expenditure than other area and the larger number of children (age under 15) may lead to higher household expenditure as well due to the fact they cannot take care and work for themselves. Hence, the signs for each coefficient make economic sense.

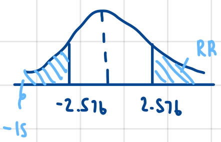
2.b) Test each parameter separately if they are significantly different from zero or not. (Use α = 0.01)

t - test :

d.f. : n − k
= 14908 − 3
= 14905

β_1 ; $H_0 : \beta_1 = 0$
$H_a : \beta_1 \neq 0$
$t_{calc}(\beta_1) = 43.83$
$t_{cri}(\frac{\alpha}{2} = 0.005, 14905) = 2.576$
$t_{calc} > t_{cri}$ ∴ Reject H_0

∴ We can say for sure that β_1 is significantly different from zero at 99 out of 100 times when we sample.



β_2 ; $H_0 : \beta_2 = 0$
$H_a : \beta_2 \neq 0$
$t_{calc}(\beta_2) = -15.8$
$t_{cri}(\frac{\alpha}{2} = 0.005, 14905) = -2.576$
$t_{calc} < t_{cri}$ ∴ Reject H_0

∴ We can say for sure that β_2 is significantly different from zero at 99 out of 100 times when we sample.

β_3 ; $H_0 : \beta_3 = 0$
$H_a : \beta_3 \neq 0$
$t_{calc}(\beta_3) = 6.82$
$t_{cri}(\frac{\alpha}{2} = 0.005, 14905) = 2.576$
$t_{calc} > t_{cri}$ ∴ Reject H_0

∴ We can say for sure that β_3 is significantly different from zero at 99 out of 100 times when we sample.

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

$E(\widehat{hhexp}_i \mid \widehat{area}_i = 1, child_i = 3) = 9{,}736 - 2{,}835(1) + 881(3) = 9{,}544$ #

The average value of a household expenditure not living in a municipal area with 3 children aged under 15 is 9544 unit per month.

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

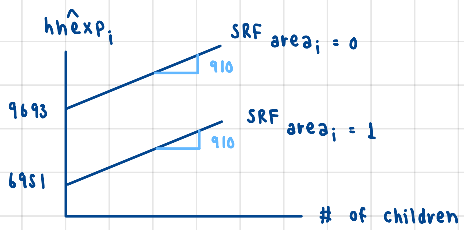
Draw a diagram for this model displaying sampled regression functions (SRF) with expected value of household expenditure on the vertical axis and number of children on the horizontal axis, taking **only significant parameter(s)** into account. Indicate the intercept and slope for each SRF where applicable. Testing of significance can be shortened. assume that level of significant is 0.01

d.f. = n − k
= 14,908 − 4
= 14,904

$t_{cri}(\frac{\alpha}{2} = 0.005, 14{,}904) = \pm 2.576$

β_1 : $t_{calc} > 2.576$ → reject H_0
β_2 : $t_{calc} < -2.576$ → reject H_0
β_3 : $t_{calc} > 2.576$ → reject H_0
β_4 : $t_{calc} < 2.576$ → fail to reject H_0

only $\beta_1, \beta_2, \beta_3$ that are significant



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 - exceed 10	

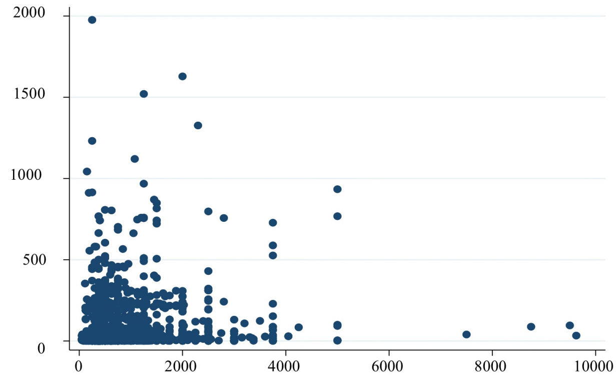
Given that you are exploring multicollinearity assumption, which pair of variables that you suspect they might be linearly correlated? Provide clear explanation what criteria (ion) that you rely on making that judgement.

Age_i and $Agesq_i$ might be linearly correlated due to the value of VIF of these two are high (exceed 10) and TOL values are close to 1, leading to higher variance and high correlation and is cause for concern.

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?

$Agesq_i$ since it is the square of age_i variable and both have high value of VIF and low value of TOL, removing $agesq_i$ might be more efficient and help fix the problem of multicollinearity.

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.



Heteroscedasticity is present since there is no statistical significant relationship between $\hat{u}_i^2$ and $weekot_i$.

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		
				Number of obs =	2,032
				F(4, 2027) =	9.52
Model	829063.863	4	207265.966	Prob > F =	0.0000
Residual	44148135	2,027	21780.037	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 : homoscedasticity

H_a : otherwise

$$F_{calc} = \frac{R^2_{\hat{u}_i^2} / k}{(1 - R^2_{\hat{u}_i^2})/(n-k-1)}$$

$$= \frac{0.0184 / 5}{(1-0.0184)/(2032-5-1)}$$

$$= 7.5954$$

Assume $\alpha = 0.05$

$F_{cri} = F_{0.05\,(2,\,2026)} = 2.21$

$\therefore F_{calc} > F_{cri} \rightarrow$ Reject H_0

We can make sure that heteroscedasticity is present in the model at 95% confidence interval.