WARISSARIN JANTRON
ID : 6304641803

Question 1

Table 1.1

Source	SS	df	MS			
Model	166.011417	5	33.2022834	Number of obs	=	1,260
Residual	278.96855	1,254	.222462959	F(5, 1254)	=	149.25
				Prob > F	=	0.0000
				R-squared	=	0.3731
				Adj R-squared	=	0.3706
Total	444.979967	1,259	.353439211	Root MSE	=	.47166

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
educ	.0708503	.0052325	Omitted for the purpose of this exam		
exper	.0389808	.0043524			
expersq	-.0005986	.0000975			
union	.1924593	.0301994			
female	-.4421609	.0289766			
_cons	.443611	.078859			

Table 1.2

Source	SS	df	MS			
				Number of obs	=	1,260
				F(7, 1252)	=	109.21
Model	168.697151	7	24.099593	Prob > F	=	0.0000
Residual	276.282816	1,252	.220673176	R-squared	=	0.3791
				Adj R-squared	=	0.3756
Total	444.979967	1,259	.353439211	Root MSE	=	.46976

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
educ	.0691306	.00525	Omitted for the purpose of this exam		
exper	.0395785	.0043428			
expersq	-.0006081	.0000971			
union	.1884632	.0301843			
female	-.4388235	.028877			
belavg	-.1388291	.0417749			
abvavg	.0070104	.0302809			
_cons	.4737302	.0795614			

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)

$$\log(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female_i + u_i$$

$$\log(wage_i) = 0.4436 + 0.07 educ_i + 0.039 exper_i - 0.006 expersq_i + 0.192 union_i - 0.442 female_i$$

T - test , $\alpha = 0.05$

$H_0 : \hat{\beta}_2 = 0$

$H_a : \hat{\beta}_2 \neq 0$

$t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)}$

$= \dfrac{0.0709 - 0}{0.0052}$

$= 13.6346$

$t_{cri} = t_{\frac{\alpha}{2}}, df = n-k = t_{\frac{0.05}{2}}, df = 1260-6 = t_{0.025}, df = 1254$

$= \pm 1.96$

$\therefore t_{cal} > t_{cri}$ So, We enable to reject H_0 (Null-Hypothesis) or We can conclude that $\hat{\beta}_2$ is significantly different from zero implying that education variable has impact on logarithm of hourly wage.

WARISSARIN JANTRON
ID : 6304641303

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

⇒ Find the overall significant by using F-test

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

H_a : Otherwise

$\alpha = 0.05$

F-Stats = 109.21

p-value = 0.000 , p-valve < α
0.0000 < 0.05 ⇒ ∴ We can reject H_0 implying that there is (at least 1) variable affects wage vorible.

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)

$H_0 : \beta_{bel} = \beta_{abv} = 0$

H_a : otherwise

$\alpha = 0.05$

F restrict / Unrestrick
(R) (UR)

$$F_{cal} = \frac{(R^2_{UR} - R^2_{R})/q}{1-R^2_{UR}/(n-k-1)}$$

$$= \frac{(0.3791 - 0.3731)/2}{(1-0.3791)/(1260-8-1)}$$

$$= 6.048$$

F_{cri} : $df_1 = q = 2$
$df_2 = n-1-k = 1251$, $\alpha = 0.05$

$F_{cri} = 2.996$, $F_{cri} < F_{cal}$
$2.996 < 6.048$ ⇒ ∴ We can reject H_0 (Null-Hypothesis) implying that physical attractiveness has an impact on logarithm of hourly wage.

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

⇒ According to the result from 1.c) , we found that physical attractiveness has no impact on logarithm of hourly wage. There fore, there is no any evidence to convince that the woman with above average look will earn more wage than the woman with average looks.

WARISSARIN JANTRON
ID : 6304641303

Question 2

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

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

→ The providing model represents household expenditure which the family who live in municipal area mostly spend more than the family who live in other locations. As well as having more children, there is more household expense to pay. These two variables ($area_i$ and $child_i$) affect the household expenditure. Therefore, 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)

→ $H_0 : \beta_k = 0$
$H_a : \beta_k \neq 0$; for every coefficient
$\alpha = 0.01$

$t_{cri} = t_{\alpha}, df = n-k$
$= t_{0.01}, df = 14908 - 3$
$= t_{0.01}, df = 14905$
$= 2.576$

$t_{cal} = \dfrac{\hat{\beta} - \beta}{se_{\beta}}$

$\hat{\beta}_1 : \dfrac{9{,}736 - 0}{43.83} = 222.0296 > t_{cri}$ → Reject H_0, significant

$\hat{\beta}_2 : \dfrac{-2{,}835}{-15.8} = 179.4304 > t_{cri}$ → Reject H_0, significant

$\hat{\beta}_3 : \dfrac{881}{6.81} = 129.3686 > t_{cri}$ → Reject H_0, significant

∴ Constant, $area_i$, and $child_i$ are statistically significant from zero.

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

=1   child=3

→ Not living in municipal ; $area_i = 1$
Having 3 children aged under 15 ; $= 3$

$$E(\widehat{hhexp_i}) = 9{,}736 - 2{,}835(1) + 881(3)$$
$$= 9{,}544$$

∴ Expected value of $hhexp_i$ is 9,544

WARISSARIN JANTRON
ID: 6304641803

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

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

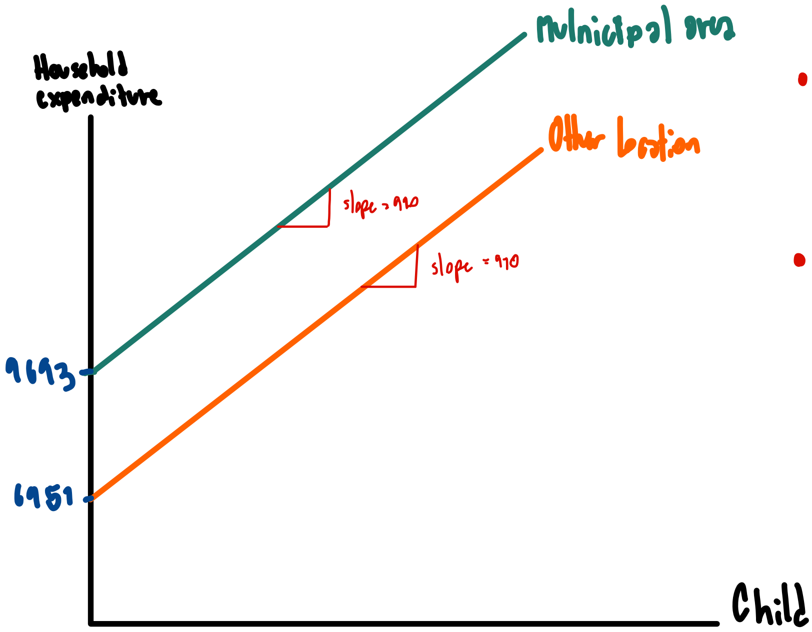
Draw a diagram for this model displaying sampled regression functions (SRF) with expected value of household expenditure on the vertical axis and number of children on the horizontal axis, taking **only significant parameter(s)** into account. Indicate the intercept and slope for each SRF where applicable.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

⇒ To test significance, $|t\text{-test}| < t_{cri}$

$t_{cri} = 1.96$ at $\alpha = 0.05$

The parameter of $area_i \times child_i$ is -0.25 which $|-0.25| < 1.96$

Therefore, $area_i \times child_i$ is not significant. → $\widehat{hhexp} = 9693 - 2742area_i + 910\,child_i$



- Municipal area ; (area = 0) ~→ put in the model

  ⇒ $\widehat{hhexp} = 9693 - 910\,child$

  slope = 910 , intercept = 9693

- Other location , (area = 1)

  ⇒ $\widehat{hhexp} = 9693 - 2742(1) + 910\,child_i$

  $\widehat{hhexp} = 6951 + 910\,child_i$

  slope = 910 , intercept = 6951

Question 3

WARISSARIN JANTRON
ID : 6304641803

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	

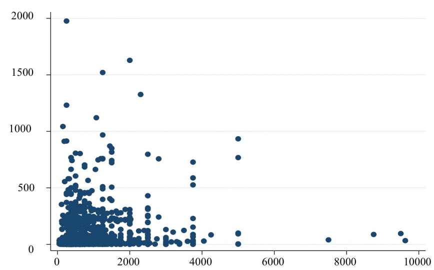
Given that you are exploring multicollinearity assumption, which pair of variables that you suspect they might be linearly correlated? Provide clear explanation what criteria (ion) that you rely on making that judgement.

→ According to the table, the value of VIF of age and agesq are very high but actually it supposed to lower than 10. The value of their $\frac{1}{VIF}$ should be closer to 1 but the real values are only 1.01 and 1.02 which is very low comparing to the values of $\frac{1}{VIF}$ of other variables. By putting all values to plot the graph, we will find that these pair variables are linearly correlated. From the reasons providing above, age and agesq might be linearly correlated.

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?

→ According to the value of VIF of age and agesq which not exceed 10, they are multicolinearity. So we can elaborate 1 variable and the rest are the same.

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 in this model since there is no relationship between $\hat{u}_i^2$ and weekot. Most of the data closes to zero and some are spreaded out. It is difficult to find the slope.

WARISSARIN JANTRON
ID : 6304641803

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
Total	44977198.8	2,031	22145.3465	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 : Homoscedasticity

H_a : Heteroscedasticity

$\alpha = 0.05$

$LM = N \times R^2$

$= 2032 \times 0.0184$

$= 37.3888$

Chi-square, $df = k - 1$ (5)

Chi-square, $df = 4$: $\chi^2 = 9.488$

Since $LM > \chi^2$

$37.3888 > 9.488$

∴ We can reject Ho, and we are sure that heteroscedasticity is present in the model.