6304640086
Aurasa Rueanjaiman

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)

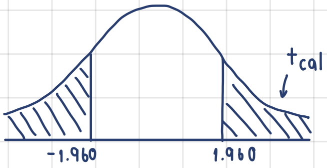
1) $\log(wage_i) = 0.4436 + 0.0708\, educ_i + 0.03898\, exper_i - 0.000598\, expersq + 0.1925\, union_i - 0.44216\, female_i$

2) If years of school increase by 1 year, logarithm of hourly wage will increase by 0.0708 dollars

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 $= 1.960$



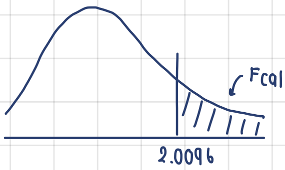
∴ we can reject null hypothesis and we can make sure 95% that education impact on logarithm of wage.

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

• $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$
H_a : otherwise

• $F_{cal} = \dfrac{ESS/df}{RSS/df} = \dfrac{ESS/K-1}{RSS/n-K} = 109.2094$

• $\alpha = 0.05$
$F_{upper,\alpha}(7, 1252) = 2.0096$



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

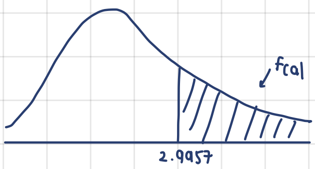
1.c) If we are interested in testing whether "physical attractiveness" has an impact on logarithm of hourly wage at all, what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

→ F - test

- H_0 : physical attractiveness has no impact on logarithm of hourly wage
  H_a : otherwise

- $F_{cal} = \dfrac{ESS_{1.2} - ESS_{1.1} / 7-5}{RSS_{1.2} / 1260 - 8} = 6.0853$

- $\alpha = 0.05$
  $F_{upper} = 2.9957$



∴ we can reject null hypothesis and we can make sure 95% 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.

as $\log(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female_i + \beta_7 belavg_i + \beta_8 abvg_i + u_i$

- woman with average looks $belavg_i = 0$, $abvavg_i = 0$

$\log(wage_i) = 0.4737302 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i - 0.4388235(1)$

$= 0.0349067 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i$

- woman with above average looks $abovg_i = 1$

$\log(wage_i) = 0.4737302 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i - 0.4388235(1) + 0.0070104(1)$

$= 0.0419171 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i$

∴ The intercept (β_1) of woman with average looks is lower than the intercept (β_1) of woman with above average looks.

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

β_2 is negative value which means that numbers not living in municipal have lower expense which make economic sense because actually, people who live in non-municipal have less expense according to lower cost of living.

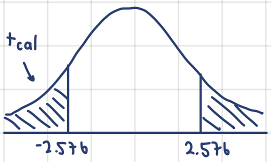
β_3 is positive value which means that when family has more childrend, expense will definitely higher due to cost of living or education.

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

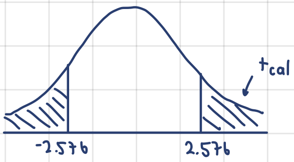
- β_1 $H_0: \beta_1 = 0$
  $H_a: \beta_1 \neq 0$
- $t_{cal}(\beta_1) = \dfrac{\hat{\beta}_1 - \beta_1}{se\,\hat{\beta}_1} = 43.83$
- $\alpha = 0.01$ $\alpha/2 = 0.005$
  $df = 14905$
  $t_{cri} = 2.576$

β_2 $H_0: \beta_2 = 0$
$H_a: \beta_2 \neq 0$

- $t_{cal}(\beta_2) = \dfrac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = -15.8$
- $\alpha = 0.01$ $\alpha/2 = 0.005$
  $df = 14905$
  $t_{cri} = 2.576$



- β_3 $H_0: \beta_3 = 0$
  $H_a: \beta_3 \neq 0$
- $t_{cal}(\beta_3) = \dfrac{\hat{\beta}_3 - \beta_3}{se\,\hat{\beta}_3} = 6.89$
- $\alpha = 0.01$ $\alpha/2 = 0.005$
  $df = 14905$
  $t_{cri} = 2.576$



∴ we can reject all these null hypothesis and we can make sure 99% that they are significantly different from 0.

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

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

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

$= 9544$

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.

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

- $H_0 : \hat{\beta}_2 = 0$
  $H_a : \hat{\beta}_2 \neq 0$
- $t_{cal} = -6.55$
- $\alpha = 0.01$; $\alpha/2 = 0.005$
  $t_{cri} = 2.576$

$\therefore$ can reject H_0

- $H_0 : \beta_3 = 0$
  $H_a : \beta_3 \neq 0$
- $t_{cal} = 5.17$
- $\alpha = 0.01$; $\alpha/2 = 0.005$
  $t_{cri} = 2.576$

$\therefore$ can reject H_0

- $H_0 : \beta_4 = 0$
  $H_a : \beta_4 \neq 0$
- $t_{cal} = -0.25$
- $\alpha = 0.01$; $\alpha/2 = 0.005$
  $t_{cri} = 2.576$

$\therefore$ cannot reject H_0

$\widehat{hhexp_i} = 9693 - 2742\ area_i + 910\ child_i$

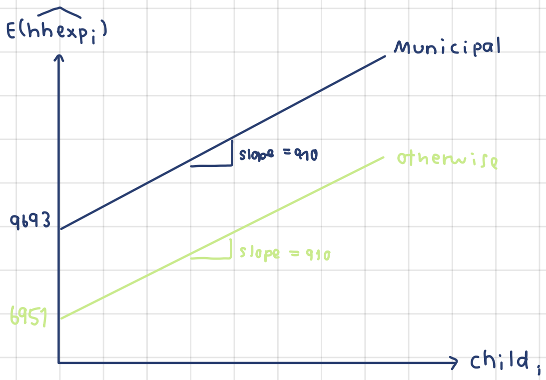
- municipal area = 0

$\widehat{hhexp_i} = 9693 + 910\ child_i$

- otherwise area = 1

$\widehat{hhexp_i} = 9693 - 2742 + 910\ child_i$

$= 6951 + 910\ child$



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.

As VIF > 5 ; have a chance multicolinearity
VIF > 10 ; multicolinearity

∴ age and agesq are suspected to 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?

No, I don't consider removing variables from the model because we don't have enough evidence to make sure

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, we conclude that heteroscedasticity is present in this model because when $weekot_i$ increases, $\hat{u}_i^2$ increases 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

- $F_{cal} = \dfrac{R^2_{\hat{u}_i^2} / k}{(1 - R^2_{\hat{u}_i^2}) / (n-k-1)} = \dfrac{0.0184/5}{(1-0.0184)/(2,032-4)} = 7\ 6029$

- $\alpha = 0.05$

  $F_{cri\,(5,2028)} = 2.2141$

→ $F_{cal} > F_{cri}$ ∴ we can reject the null hypothesis at 95% , heteroscedasticity is present.