6304640821
Penguin

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

1.a) $Log(wage_i) = 0.4436 + 0.0709\, educ_i + 0.0398\, exper_i - 0.0005\, expersq_i + 0.1925\, union_i - 0.4422\, female_i + u_i$

To test how education has an impact on logarithm of hourly wage or not, we use t-test to test statistical significance.

(1) state hypothesis

$H_0 = \beta_2 = 0$

$H_1 = \beta_2 \neq 0$

(2) Calculate test statistics

$$t_{cal}(\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{se\hat{\beta}_2} = \frac{0.0708505 - 0}{0.0052325} = 13.54043$$

(3) Pick an α and state decision rule

$\alpha = 0.05$

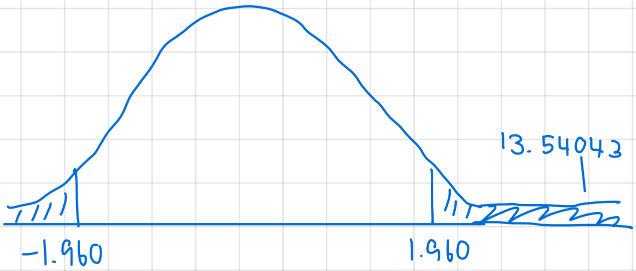
$n = 1260$; $k = 6$

d.f. = 1,254

$t_{lower} = -1.960$

$t_{upper} = 1.960$

(4) Concluding the test result



∴ When education go up, we can make sure that your wage will go up as well.

1.b) To testing the overall significance, we relies on an F-test which the hypothesis is

H_0 : All the β_k are simultaneously equal to zero

H_1 : Otherwise

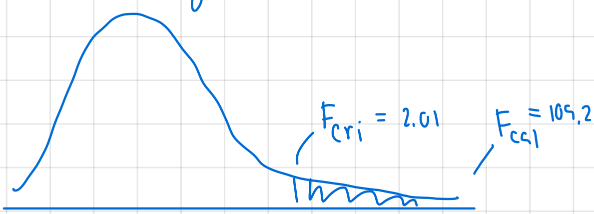
(2) Calculate test statistic

$$F_{cal} = \frac{ESS/d.f.}{RSS/d.f.} = \frac{168.697151 / 7}{276.282816 / 1252} = \frac{24.099593}{0.220673176} = 109.2094356$$

(3) Pick an α and state decision rules

$\alpha = 0.05$ $\quad F_{upper,\alpha(7,1252)} = 2.01$

(4) concluding the test result



∴ Since $F_{cal} > F_{cri}$, therefore we can reject the null hypothesis or we can say that all the coefficients are not simultaneously equal to zero.

1. C) To testing impact of physical attractiveness to logarithm of hourly wage, we should test the marginal contribution of new regressors which is $belavg_i$ and $abavg_i$ with the F-test. For hypothesis, it should be

H_0: physical attractiveness has no marginal contribution to the model $\beta_7 = \beta_8 = 0$

H_1: otherwise

(2) Calculate the test statistics

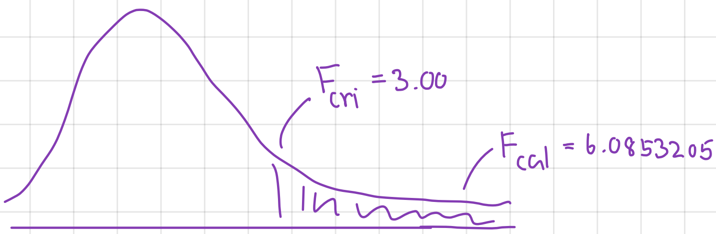
$$F_{cal} = \frac{ESS_{new} - ESS_{old} / \#\text{of new regressors}}{RSS_{new} \,(n-k_{new})} = \frac{(168.697151 - 166.011417)/2}{276.282816/(1260-8)}$$

$$= \frac{1.342867}{0.220673176} = 6.0853205$$

(3) Pick an α and state decision rule

$\alpha = 0.05$ $\quad F_{upper,\alpha(2,1252)} = 3.00$

(4) Conclude the test result



∴ reject null hypothesis
physical attractiveness has a marginal contribution to the model

2.a) Yes. The reason is, first, people who lived in municipal area tend to spend money more than people who lived in rural area, for example, rental expenses in middle of the town is higher than suburb area. Second is chid aged under 15 have to financial independence to their parent, so expenditure of parent will be higher.

2.b) Testing the individual significance relies on t-test.

(1) $H_0 : \beta_k = 0$

$H_1 : \beta_k \neq 0$ for every coefficient

(2) $\alpha = 0.01$ $\quad$ $d.f = n-k = 14905$

$t_{cri} = \pm 2.576$

(3) $t_{cal}(\beta_1) = 43.98$

$t_{cal} > t_{cri}$

$\therefore$ Reject H_0

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

$t_{cal}(\beta_2) = -15.$

$t_{cal} < t_{cri}$

$\therefore$ Reject H_0

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

$t_{cal}(\beta_3) = 6.82$

$t_{cal} > t_{cri}$

$\therefore$ Reject H_0

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

2.c) $E(\widehat{hhexp_i} \mid area_i = 1, child_i = 3) = 9{,}736 - 2{,}835(1) + 881(3) = 9{,}544$ #

2.d) assume $\alpha = 0.01$; $d.f = n-k = 14{,}908 - 4 = 14{,}904$

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

$t_{cal}(\beta_1) : 34.38 > 2.576$ $\quad$ $\therefore$ reject H_0

$t_{cal}(\beta_2) : -6.55 < -2.576$ $\quad$ $\therefore$ reject H_0

$t_{cal}(\beta_3) \cdot 5.17 > 2.576$ $\quad$ $\therefore$ reject H_0

$t_{cal}(\beta_4) : -0.25 < 2.576$ $\quad$ $\therefore$ cannot reject H_0

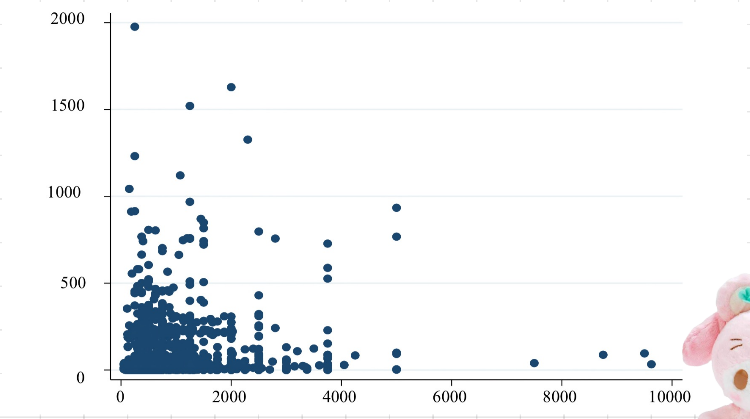
3.a)

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	

Age, and Agesq, might be Linearly correlated because the value of VIF is higher than 10 and tolerance (TOL) values are close to 1.

3.b) The variable that I will remove is agesq, because age and square of age are having high VIF and tolerance. So remove one of them should help we solve the multicollinearity problem.

3.C)



There is heteroscedasticity in this model because statistical significant relationship is not exists

3.d)

(1) H_0 : homoscedasticity

H_a : otherwise.

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

$$= \frac{0.018415}{(1-0.0184)/(2032-5-1)}$$

$$= 7.5954$$

Assume $\alpha = 0.05$

$F_{cri} = 2.21$

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

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