Wasunan Seeduangmaneechai
6204640319

Assignment #2

Instructions:

- For all questions, answer up to 4 decimal places.
- This assignment is due on **Thursday, May 20, 2021 before 23.59.**
- Write your answer in either digital or ordinary paper. For digital paper, export pages into a single PDF file. For ordinary paper, take photos of your writing and convert them into a single PDF file as well.
- There is no need to rewrite the question. Assign number item, i.e. 1 a., clearly before your answer is sufficient.
- Submit your assignment into Moodle.
- Name your file as StudentID_Nickname (in Thai) such as 123456789_ไม้. **Please follow this instruction strictly since it will help me a lot with file management.**

---

dummy var.

Question 1. The data set CEOSAL1.DTA contains information on 209 CEOs for the year 1990; these data were obtained from Business Week (5/6/1991). To study effect of firm performances and types of industry where CEOs work on CEO compensation, the CEO salary regression is proposed as follows:

= dummy

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + \beta_2 \text{ROE}_i + \beta_3 finance_i + \beta_4 consprod_i + \beta_5 utility_i + u_i$$

where $\log(salary_i)$ = logarithm of CEO annual salary (in 1,000 USD)
$\log(sales_i)$ = logarithm of firms' sale (in 1 million USD)
ROE_i = average return on equity for the CEO's firm for the previous three years
(Return on equity is defined in terms of net income as a percentage of common equity)
$finance_i$ = 1 if in financial industry, = 0 otherwise
$consprod_i$ = 1 if in consumer product industry, = 0 otherwise
$utility_i$ = 1 if in utility industry, = 0 otherwise

= dummy

($finance_i$, $consprod_i$, and $utility_i$ are binary variables indicating the financial, consumer products, and utilities industries. The omitted industry is transportation.

Using STATA, the estimation result is shown below. Answer the following questions.

Source	SS	df	MS
Model	23.8109943	5	4.76219887
Residual	42.9111689	203	.211385068
Total	66.7221632	208	.320779631

Number of obs	= 209
F(5, 203)	= 22.53
Prob > F	= 0.0000
R-squared	= 0.3569
Adj R-squared	= 0.3410
Root MSE	= .45977

	lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1	lsales	.2571917	.0320348	8.03	0.000	.0194282	.3203553
2	roe	.0111517	.3342996	2.59	0.010	.0026742	.0196293
3	finance	.1579564	.0890017	1.77	0.077	-.0175299	.3334426
4	consprod	.1808917	.0847683	2.13	0.034	.0137524	.3480311
5	utility	-.2830015	.0992337	-2.85	0.005	-.4786624	-.0873405
0	_cons	4.588101	.2950221	15.55	0.000	4.0064	5.169801

a. Write out the estimated regression equation for $\log(salary_i)$. Interpret the estimated coefficient associated with $\log(sales_i)$.

b. What is the overall significance of the regression? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State **the critical value** for hypothesis testing to receive full points.

c. Compute the approximate percentage difference in estimated salary between the utility and transportation sector, holding $sales_i$ and ROE_i fixed.

d. Why can't we put all the sector dummies (i.e. $finance_i$, $consprod_i$, $utility_i$ and $transport_i$) in the equation? What would happen if we put all the sector dummies in the equation and use STATA run the regression anyway?

e. In the above model, is there any benefit if we add interaction terms between roe and sector dummies, i.e. $\text{ROE}_i * finance_i$ and/or $\text{ROE}_i * consprod_i$ and/or $\text{ROE}_i * utility_i$?

a. Write out the estimated regression equation for $\log(salary_i)$. Interpret the estimated coefficient associated with $\log(sales_i)$.

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + \beta_2 \mathrm{ROE}_i + \beta_3 finance_i + \beta_4 consprod_i + \beta_5 utility_i + u_i$$

$$\log(salary_i) = 4.5881 + 0.2572 \log(sale_i) + 0.0112\ \mathrm{ROE}_i + 0.1580\ finance_i + 0.1809\ conprod_i - 0.2834\ utility_i + u_i$$

$\hat{\beta}_1 = 0.2572$; As firm's sale increase by 1%, on average, the annual salary of the CEO increase by about 0.2572%, holding other variables constant.

b. What is the overall significance of the regression? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State **the critical value** for hypothesis testing to receive full points.

Overall significance of the regression: F-test

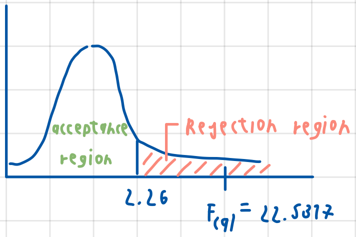
1. $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

   H_a: otherwise

2. $$F_{cal} = \frac{R^2/(k-1)}{1-R^2/(n-k)} = \frac{0.3569/(6-1)}{1-0.3569/(209-6)} = \frac{0.3569/5}{1-0.3569/203} = 22.5817$$

3. $\alpha = 0.05$

   $F_{upper,\,0.05}\,(5, 203) = 2.26$

4. 

$F_{cal} > F_{cri(0.05)}\,(5, 203)$

$\therefore$ We can reject H_0. We can make sure that $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 are not simultaneously zero, at α equal to 0.05.

Testing individual regression coefficients: t-test

1.
For β_0: $H_0: \beta_0 = 0$, $H_a: \beta_0 \neq 0$

For β_1: $H_0: \beta_1 = 0$, $H_a: \beta_1 \neq 0$

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

For β_3: $H_0: \beta_3 = 0$, $H_a: \beta_3 \neq 0$

For β_4: $H_0: \beta_4 = 0$, $H_a: \beta_4 \neq 0$

For β_5: $H_0: \beta_5 = 0$, $H_a: \beta_5 \neq 0$

2.
For β_0: $t_{cal}(\beta_0) = \dfrac{\hat{\beta}_0 - \beta_0}{se_{\hat{\beta}_0}} = \dfrac{4.5881 - 0}{0.2950} = 15.5529$

For β_1: $t_{cal}(\beta_1) = \dfrac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}} = \dfrac{0.2572 - 0}{0.0320} = 8.0375$

For β_2: $t_{cal}(\beta_2) = \dfrac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \dfrac{0.0112 - 0}{0.3343} = 0.0335$

For β_3: $t_{cal}(\beta_3) = \dfrac{\hat{\beta}_3 - \beta_3}{se_{\hat{\beta}_3}} = \dfrac{0.1580 - 0}{0.0890} = 1.7753$

For β_4: $t_{cal}(\beta_4) = \dfrac{\hat{\beta}_4 - \beta_4}{se_{\hat{\beta}_4}} = \dfrac{0.1809 - 0}{0.0848} = 2.1333$

For β_5: $t_{cal}(\beta_5) = \dfrac{\hat{\beta}_5 - \beta_5}{se_{\hat{\beta}_5}} = \dfrac{-0.2830 - 0}{0.0992} = -2.8528$

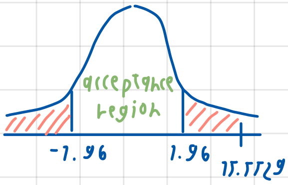
3. $\alpha = 0.05$

   $df = 203$

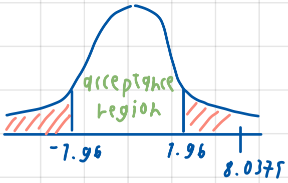
   $t_{upper} = 1.96$

   $t_{lower} = -1.96$

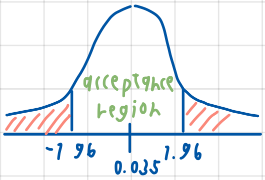
4. For β_0, reject H_0



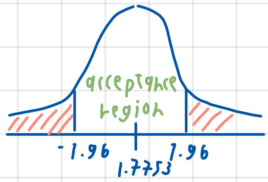
For β_1, reject H_0



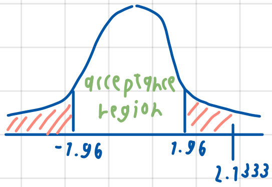
For β_2, can't reject H_0



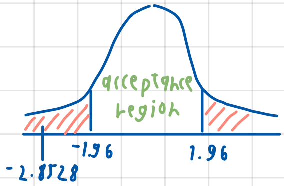
For β_3, can't reject H_0



For β_4, reject H_0



For β_5, reject H_0



$\therefore$ Reject H_0 for $\beta_0, \beta_1, \beta_4, \beta_5$.
$\therefore$ There's enough evidence to say that $\beta_0, \beta_1, \beta_4, \beta_5$ are individually significant, at α equal to 0.05.

c. Compute the approximate percentage difference in estimated salary between the utility and transportation sector, holding $sales_i$ and ROE_i fixed.

$\hat{\beta}_5 = -0.283$, $100\,(e^{(-0.283)} - 1) = -24.648\,\%$

On average, the CEO annual salary are lower than the transportation sector by about 24.648%, holding other variables fixed.

d. Why can't we put all the sector dummies (i.e. $finance_i$, $consprod_i$, $utility_i$ and $transport_i$) in the equation? What would happen if we put all the sector dummies in the equation and use STATA run the regression anyway?

In this case, we omitted the transportation because transportation is a base group, the group against which the comparisons are made. So that's why the transportation is taken as a base group category. Thus, in the case that more than 1 qualitative variable is included in the regression, it's important that the omitted category should be chosen as a base category, and all comparisons will be made in relation to that category.

e. In the above model, is there any benefit if we add interaction terms between roe and sector dummies, i.e. $\text{ROE}_i * finance_i$ and/or $\text{ROE}_i * consprod_i$ and/or $\text{ROE}_i * utility_i$?

└ slope dummy

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + \beta_2 \text{ROE}_i + \beta_3 finance_i + \beta_4 consprod_i + \beta_5 utility_i + u_i$$

If we add another slope dummy:

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + \beta_2 \text{ROE}_i + \beta_3 \, finance_i + \beta_4 \, consprod_i + \beta_5 \, utility + \beta_6 \, \text{ROE} \cdot finance_i$$

$$E(\log(salary_i) \mid finance = 1) = (\beta_0 + \beta_3) + \beta_1 \log(sales) + (\beta_2 + \beta_6)\,\text{ROE} + \beta_4\, consprod + \beta_5\, utility$$

$$E(\log(salary_i) \mid finance = 0) = \beta_0 + \beta_1 \log(sales) + \beta_2\, \text{ROE} + \beta_4\, consprod + \beta_5\, utility$$

$\therefore$ For financial industry, if ROE increase by 1%, CEO annual wage will increase equal to $\beta_2 + \beta_6$, higher than an increase in non financial industry, increase equal to β_2.
$\therefore$ By adding a slope dummy, it can help to distinguish to effect between each group.

Question 2. Birth weight has been used by officials as one of the main determinants of health. Data set BWGHT.DTA contains data on infant birth weights in ounces ($bwght_i$), average number of cigarettes mother smoked per day during pregnancy (cigs), family income ($faminc_i$), father's year of education ($fathereduc_i$), and mother's year of education ($mothereduc_i$). The following two regressions were estimated using data on n = 1191 births:

Model 2.1: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + u_i$ » constrained

```
regress  bwght  cigs   faminc

      Source |       SS       df       MS              Number of obs =    1191
-------------+------------------------------           F(  2,  1188) =   18.44
       Model |  14536.9538     2  7268.47691           Prob > F      =  0.0000
    Residual |  468209.738  1188  394.115941           R-squared     =  0.0301
-------------+------------------------------           Adj R-squared =  0.0285
       Total |  482746.692  1190  405.669489           Root MSE      =  19.852

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5876985   .1090181
      faminc |   .0624684   .0324438     Omitted for the purpose of this exam.
       _cons |   118.5568   1.234278
------------------------------------------------------------------------------
```

Model 2.2: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + \beta_3 fathereduc_i + \beta_4 mothereduc_i + u_i$ » unconstrained

```
regress  bwght  cigs   faminc fatheduc motheduc

      Source |       SS       df       MS              Number of obs =    1191
-------------+------------------------------           F(  4,  1186) =   10.05
       Model |  15827.6593     4  3956.91482           Prob > F      =  0.0000
    Residual |  466919.033  1186  393.69227            R-squared     =  0.0328
-------------+------------------------------           Adj R-squared =  0.0295
       Total |  482746.692  1190  405.669489           Root MSE      =  19.842

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5894954   .1106172     Omitted for the purpose of this exam.
      faminc |   .0538254   .0366502
    fatheduc |   .4936695   .2832896
    motheduc |  -.4379234   .3197377
       _cons |   118.0741   3.500291
------------------------------------------------------------------------------
```

where $bwght_i$ = birth weight, ounces
$cigs_i$ = average number of cigarettes the mother smoked per day while pregnant
$faminc_i$ = 1988 family income, \$1000s
$fathereduc_i$ = father's years of education
$mothereduc_i$ = mother's years of education

Answer the following questions.

a. Based on **Model 2.1**, test whether smoking has an impact on birth weight. Show your work. (use $\alpha = 0.05$)

b. Based on **Model 2.1**, construct a 99% confidence interval for β_2.

c. Would your conclusion in a) change if you use the result from **Model 2.2**? Show your work. (use $\alpha = 0.05$)

d. What is the overall significance of the regression from **Model 2.2**? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State the critical value for hypothesis testing to receive full points.

e. If we are interested in testing whether "**parents' education**" has an impact on birth weight at all, what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (use $\alpha = 0.05$)

a. Based on **Model 2.1**, test whether smoking has an impact on birth weight. Show your work. (use $\alpha = 0.05$)

Test β_1 individually : t-test

1. $H_0: \beta_1 = 0$

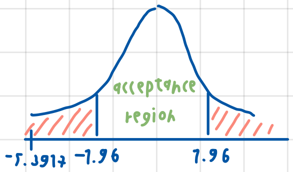
   $H_a: \beta_1 \neq 0$

2. $t_{cal}(\beta_1) = \dfrac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}} = \dfrac{-0.5877 - 0}{0.109} = -5.3917$

3. $\alpha = 0.05$

   $t_{lower} = t_{n-k} = t_{1191-3} = -1.96$

   $t_{upper} = t_{n-k} = t_{1191-3} = 1.96$

4. 

∴ Reject H_0

∴ There's enough evidence that smoking has an impact on birth weight, at $\alpha = 0.05$.

b. Based on **Model 2.1**, construct a 99% confidence interval for β_2.

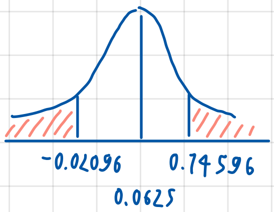
Confidence Interval :

1. $\alpha = 0.01$

2. $t_{\alpha/2} = t_{0.005} = 2.576$

3. Lower limit: $\hat{\beta}_2 - t_{\alpha/2} \cdot \hat{\sigma}_{\hat{\beta}_2} = 0.0625 - (2.576)(0.0324)$
   $= -0.02096$ #

   Upper limit: $\hat{\beta}_2 + t_{\alpha/2} \cdot \hat{\sigma}_{\hat{\beta}_2} = 0.0625 + (2.576)(0.0324)$
   $= 0.14596$ #



c. Would your conclusion in a) change if you use the result from **Model 2.2**? Show your work. (use $\alpha = 0.05$)

Test β_1 individually : t-test

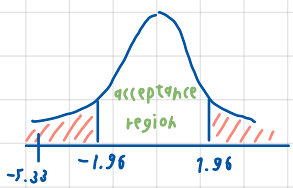
1. $H_0: \beta_1 = 0$

   $H_a: \beta_1 \neq 0$

2. $t_{cal}(\beta_1) = \dfrac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}} = \dfrac{-0.5895 - 0}{0.1106} = -5.3300$

3. $\alpha = 0.05$

   $t_{lower} = -1.96$ $\quad t_{upper} = 1.96$

4. 

∴ Reject H_0

∴ Conclusion will still be the same as a.)

d. What is the overall significance of the regression from **Model 2.2**? What test do you use? Which of the coefficients are individually statistically significant at the 5 percent level? State the critical value for hypothesis testing to receive full points.

Overall significance of the regression: F-test

1. $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

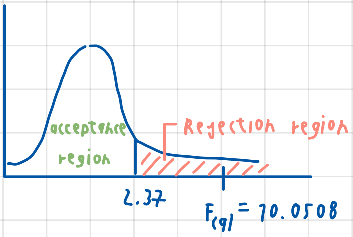
H_a: otherwise

2. $F_{cal} = \dfrac{ESS/k-1}{RSS/n-k} = \dfrac{MS(E)}{MS(R)} = \dfrac{3956.9148}{393.6928} = 10.0508$

3. $\alpha = 0.05$

$F_{upper,\alpha}(4, 1186) = 2.37$

4.



$\therefore F_{cal} > F_{cri(0.05)}(4, 1186)$

$\therefore$ We can reject H_0. We can make sure that $\beta_1, \beta_2, \beta_3$, and β_4 are not simultaneously zero. #

Testing individual regression coefficients: t-test

1. For β_0 $H_0: \beta_0 = 0$
$H_a: \beta_0 \neq 0$

For β_1 $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

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

For β_3 $H_0: \beta_3 = 0$
$H_a: \beta_3 \neq 0$

For β_4 $H_0: \beta_4 = 0$
$H_a: \beta_4 \neq 0$

2. For β_0 $t_{cal}(\beta_0) = \dfrac{\hat{\beta}_0 - \beta_0}{se_{\hat{\beta}_0}} = \dfrac{118.0741 - 0}{3.50029} = 33.7327$

For β_1 $t_{cal}(\beta_1) = \dfrac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}} = \dfrac{-0.5895 - 0}{0.1106} = -5.3300$

For β_2 $t_{cal}(\beta_2) = \dfrac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \dfrac{0.0538 - 0}{0.0367} = 1.4659$

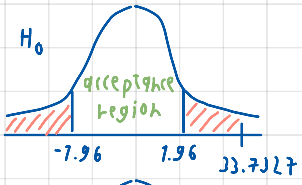
For β_3 $t_{cal}(\beta_3) = \dfrac{\hat{\beta}_3 - \beta_3}{se_{\hat{\beta}_3}} = \dfrac{0.4937 - 0}{0.2833} = 1.7427$

For β_4 $t_{cal}(\beta_4) = \dfrac{\hat{\beta}_4 - \beta_4}{se_{\hat{\beta}_4}} = \dfrac{-0.4379 - 0}{0.3197} = -1.3697$

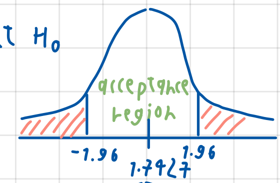
3. $\alpha = 0.05$

$t_{n-k} = t_{1191-5} = t_\infty$

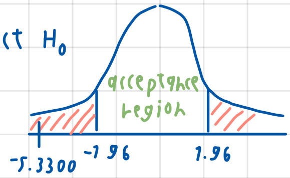
4. for β_0, Reject H_0



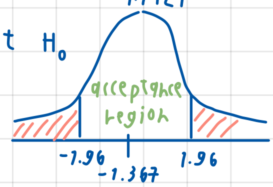
for β_3, can't reject H_0



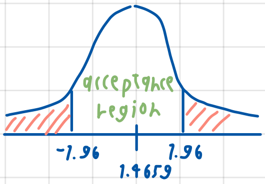
for β_1, Reject H_0



for β_4, can't reject H_0



for β_2, can't reject H_0



$\therefore$ Can reject H_0 for β_0 & β_1 only, at α equal to 0.05.

$\therefore$ There's enough evidence that β_0 & β_1 have an impact on birth weight.

e. If we are interested in testing whether "**parents' education**" has an impact on birth weight at all, what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (use $\alpha = 0.05$)

$$bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + \beta_3 fathereduc_i + \beta_4 mothereduc_i + u_i$$

Check that whether "parents' education" has any impact on birth wage:

1. $E(bwght \mid fathereduc = 0, mothereduc = 0) = \beta_0 + \beta_1 cigs + \beta_2 faminc + \beta_3(0) + \beta_4(0)$
$= \beta_0 + \beta_1 cigs + \beta_2 faminc$

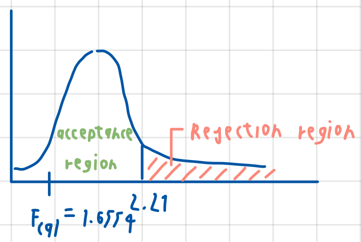
2. $E(bwght \mid fathereduc = 1, mothereduc = 0) = \beta_0 + \beta_1 cigs + \beta_2 faminc + \beta_3(1) + \beta_4(0)$
$= \beta_0 + \beta_1 cigs + \beta_2 faminc + \beta_3$
$= (\beta_0 + \beta_3) + \beta_1 cigs + \beta_2 famin$

3. $E(bwght \mid fathereduc = 0, mothereduc = 1) = \beta_0 + \beta_1 cigs + \beta_2 faminc + \beta_3(0) + \beta_4(1)$
$= \beta_0 + \beta_1 cigs + \beta_2 faminc + \beta_4$
$= (\beta_0 + \beta_4) + \beta_1 cigs + \beta_2 famin$

1. $H_0: \beta_3 = \beta_4 = 0$
H_a: otherwise

2. $F_{cal} = \dfrac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k_{UR})} = \dfrac{(0.0328 - 0.0301)/2}{(1-0.0328)/(1191-5)} = 1.6554$

3. $\alpha = 0.05$
$F_{upper,\alpha}(2, 1186) = 2.21$



∴ We can't reject H_0
∴ "parents' education" has no impact on birth weight, at α equal to 0.05

Question 3. A model of wage equation is given by

$$lawge_i = \beta_1 + \beta_2 exp_i + +\beta_3 expsq_i + \beta_4 educ_i + \beta_5 age_i + \beta_6 kid6_i + \beta_7 kid18_i + u_i$$

where $lwage_i$ = natural log of hourly wage of married women
exp_i = years of experience
$expsq_i$ = years of experience squared
$educ_i$ = years of education
age_i = age
$kid6_i$ = number of children aged 0-6 in a household
$kid18_i$ = number of children aged 6-18 in a household

The regression result from OLS is shown in the table below and answer the following questions.

	Source	SS (1.)	df (2.)	MS (3. = 1./2.)
ESS	Model	35.3398098	6	5.8899683
RSS	Residual	187.9876312	421	.446526442
TSS	Total	223.327441	427	0.5230150843

Number of obs =	428
F(6 , 421) =	13.19
Prob > F =	0.0000
R-squared =	0.1582
Adj R-squared =	0.1462
Root MSE =	.66823

	lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1	exper	.039819	.013393	2.97	0.003	.0134936	.0661444
2	expersq	-.0007812	.0004022	-1.94	0.053	-.0015718	9.37e-06
3	educ	.1078319	.0144021	7.49	0.000	.079523	.1361409
4	age	-.0014653	.0052925	-0.28	0.782	-.0118682	.0089377
5	kidslt6	-.0607106	.0887626	-0.68	0.494	-.2351836	.1137625
	kidsge6	-.014591	.0278981	-0.52	0.601	-.069428	.0402459
6 7	_cons	-.4209078	.316905	-1.33	0.185	-1.043821	.2020053

a) Figure out all the degrees of freedom in this model.

b) Figure out all the sum of squares (ESS and RSS) and mean squares in this model.

c) Figure out the adjusted R-squared ($\bar{R}^2$)

d) Given that the model above is called '**Model 3.1**', there is another competing model called '**Model 3.2**' which **an explanatory variable is excluded**, compared to '**Model 3.1**'. Though the result of estimating '**Model 3.2**' is not shown here, **what is the maximum value of R^2 from 'Model 3.2'** which will make you conclude that the excluded variable has a significant contribution in '**Model 3.1**', at the significance level of 0.05. (**Hint:** the critical value of the F-test at the significance level of 0.05 is $F_{1,421} = 3.84$)

e) As you can see from the result, age is not significantly different from zero. In other words, age does not determine how much hourly wage would be. Does this make economic sense in your opinion? What do you think cause this insignificance?

a) Figure out all the degrees of freedom in this model.

$df(\text{model}) = k-1 = 7-1 = 6$ #
$df(\text{Residual}) = n-k = 428-7 = 421$ #
$\Big\} \; df(\text{total}) = 421+6 = 427$ #

b) Figure out all the sum of squares (ESS and RSS) and mean squares in this model.

$\frac{RSS}{421} = 0.446526442$

$RSS = 187.9876312$

$223.327441 = ESS + 187.98763208$

$ESS = 35.33980892$

$mss \text{ of } ESS = \frac{ESS}{df} = \frac{35.33980892}{6} = 5.8899681533$

$mss \text{ of } TSS = \frac{TSS}{df} = \frac{223.327441}{427} = 0.5230150843$

c) Figure out the adjusted R-squared ($\bar{R}^2$)

$$\bar{R}^2 = 1-(1-R^2)\frac{(n-1)}{(n-k)} = 1-(1-0.1582)\frac{(428-1)}{(428-7)} = 0.1462$$

∴ Adjusted here means that the R^2 is adjusted by the df. associated with the sum of the squares entering the specification

∴ In this case, $k>1$, $\bar{R}^2 < R^2$, which implies that when the number of X variables increases, the adjusted R^2 increases less than the unadjusted R^2.

d) Given that the model above is called '**Model 3.1**', there is another competing model called '**Model 3.2**' which **an explanatory variable is excluded**, compared to '**Model 3.1**'. Though the result of estimating '**Model 3.2**' is not shown here, **what is the maximum value of R^2 from 'Model 3.2'** which will make you conclude that the excluded variable has a significant contribution in '**Model 3.1**', at the significance level of 0.05. (**Hint:** the critical value of the F-test at the significance level of 0.05 is $F_{1,421} = 3.84$)

model 3.2: old
model 3.1: new
$R^2_{old} = 0.1582$
$R^2_{old} = ?$
$K_{old} = 6$
$K_{new} = 7$
$F_{upper,\alpha}(1,421) = 3.84$

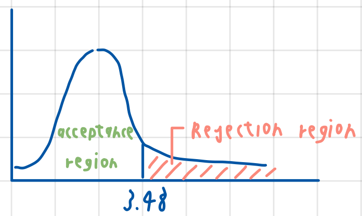
1. H_0: excluded variable has no significant contribution
H_a: otherwise

2. $F_{cal} = \frac{R^2_{new} - R^2_{old} / (\text{number of new regressors})}{1-R^2_{new} / (n-k_{new})}$

$F_{cal} = \frac{0.1582 - R^2_{old} / 1}{1-R^2_{old} / 428-7} = \frac{0.1582 - R^2_{old}/1}{1-R^2_{old}/421}$

3. $\alpha = 0.05$
$F_{upper,\alpha}(1,421) = 3.84$

4.



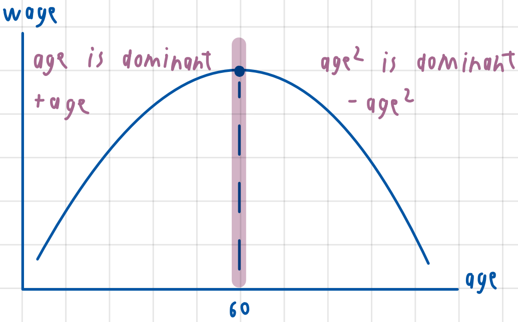
$3.84 = \frac{0.1582 - R^2_{old}/1}{1-R^2_{old}/421}$

$R^2_{old} = 0.1504$ #

max R^2, so that we can reject H_0

e) As you can see from the result, age is not significantly different from zero. In other words, age does not determine how much hourly wage would be. Does this make economic sense in your opinion? What do you think cause this insignificance?

Yes, in a sense of economic form, age is a variable that can be used to estimate wage. However, the effect of age to wage is taking in the quadratic form. This is due to the law of deminishing return that once age is reaching the level that slope equal to zero, the wage will decrease after that point.



As age² becomes higher, the square (neg.) effect will be dominant.
e.g. A retired professor at BE who are still teaching has lower wage.