Phutanate Techatanates
6204640814

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_โบ๊. **Please follow this instruction strictly since it will help me a lot with file management.**

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:

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

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

($finance_i$, $consprod_i$, 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.

Phutanate Techatanates

6204640814

```
    Source |       SS       df       MS              Number of obs =      209
-----------+------------------------------           F(  5,   203) =    22.53
     Model |  23.8109943     5  4.76219887           Prob > F      =   0.0000
  Residual |  42.9111689   203  .211385068           R-squared     =   0.3569
-----------+------------------------------           Adj R-squared =   0.3410
     Total |  66.7221632   208  .320779631           Root MSE      =   .45977

------------------------------------------------------------------------------
   lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
    lsales |   .2571917   .0320348     8.03   0.000     .0194282    .3203553
       roe |   .0111517   .3342996     2.59   0.010     .0026742    .0196293
   finance |   .1579564   .0890017     1.77   0.077    -.0175299    .3334426
  consprod |   .1808917   .0847683     2.13   0.034     .0137524    .3480311
   utility |  -.2830015   .0992337    -2.85   0.005    -.4786624   -.0873405
     _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 $\text{ROE}_i * utility_i$?

Phutanate Techatanates

6204640814

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

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

$\log(salary) = 4.588101 + 0.2571917 \, lsales + 0.0111517 \, roe + 0.1579564 \, finance_i + 0.1808917 \, consprod - 0.2830015 \, utility_i$

```
      Source |       SS           df       MS      Number of obs   =       209
-------------+----------------------------------   F(  5,   203)   =     22.53
       Model |  23.8109943         5  4.76219887   Prob > F        =    0.0000
    Residual |  42.9111689       203  .211385068   R-squared       =    0.3569
-------------+----------------------------------   Adj R-squared   =    0.3410
       Total |  66.7221632       208  .320779631   Root MSE        =    .45977

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

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. $t_{\frac{\alpha}{2}, n-k} = t_{0.025, 203} \approx 1.96$

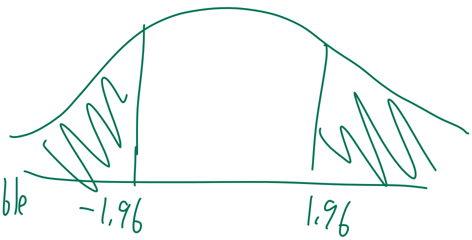
the overall significance of the regression at 95% is greater than 0

Use F test $\therefore$ $P \sim 0.00 < 0.05$

$\therefore$ H_0: All $\beta_i = 0$

H_a: All $\beta_i \neq 0$

$\therefore$ Reject H_0 $\therefore$ The model is suitable



for the coefficients to be individually significant $t_{\beta_i} > t_{\frac{\alpha}{2}, n-k}$

β_1 $t = 8.03 > 1.96$ $\therefore$ Reject H_0

β_2 $t = 2.59 > 1.96$ $\therefore$ Reject H_0

β_3 $t = 1.77 < 1.96$ $\therefore$ do not Reject H_0

β_4 $t = 2.13 > 1.96$ $\therefore$ reject H_0

β_5 $t = -2.85 < -1.96$ $\therefore$ reject H_0

β_6 $t = 15.55 > 1.96$ $\therefore$ reject H_0

i. all variables are significiant except finance

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

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

need to take parital derivatives of log(salary)

$$\frac{\log(salary)}{utility} \approx \beta_5\, utility = -0{,}2830015$$

transportation → baseline → 4.588101

including utility → 4.588101 − 0.2830015 = 4.3050995

$$e^{(4.588101)} = 98.3075672$$

$$e^{(4.3050995)} = 74.07658571$$

$$\text{Percentage} = \frac{98.30756672 - 74.07658571}{98.30756672}$$

$$= 0.24648 = 24.65\%$$

Phutanate Techatanates
6204640814

D2 D3 D4

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?

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

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

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

Transportation is already the baseline. If it was added back then that would cause an error. Its also unlikely that all dummy variable would be equal to 1.

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

D4 D5

It would depend on the importance of the interactions if they would benefit the model or not. In another words, we would need the P and t values of the interactions. If P value is less than 0.05 then it would be beneficial to add to the model

Phutanate Techatanates

6204640814

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: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + u_i$

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

```
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.

Phutanate Techatanates

6204640814

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 $\alpha = 0.05$)

Phutanate Techatanates
6204640814

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.

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

```
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   Omitted for the purpose of this exam.
      faminc |   .0624684   .0324438
       _cons |   118.5568   1.234278
------------------------------------------------------------------------------
```

first
$H_0 = \beta_{sig} = \beta_{fam} = 0$

$H_a = $ o.w.

F-test = Pvalue = 0.000

$\therefore$ Reject H_0, sutible

α,
then β_2 test

Step 1: H hypothesis

$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

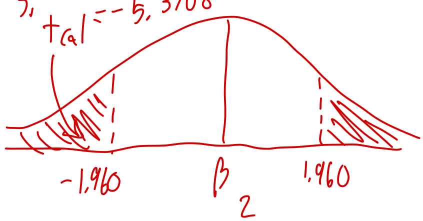
Step 2: T_{cal}, choose α

$$t_{cal} = \frac{\hat{\beta}_2 - 0}{\sigma_{\hat{\beta}_2}} = \frac{-0.5876985}{0.1090181} \approx -5.3908$$

$\alpha = 0.05$

$$t_{\frac{\alpha}{2}, n-k} = t_{0.025, 1191-3} = t_{0.025, 1188} = 1.960$$

Step 3: $t_{cal} = -5.3908$



Reject H_0, there is not enough evidence to suggest that $\beta_2 = 0$

$\therefore$ Smoking has an impact on birth weight.

Phutanate Techatanates

6204640814

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

```
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
      faminc |   .0538254   .0366502     Omitted for the purpose of this exam.
    fatheduc |   .4936695   .2832896
    motheduc |  -.4379234   .3197377
       _cons |   118.0741   3.500291
------------------------------------------------------------------------------
```

B. $\beta_2 = \hat{\beta}_2 \pm t_{\frac{\alpha}{2}, n-3} \frac{sd.\beta_2}{\sqrt{n}}$

$\beta_2 = -0.5876985 \pm 1.960 \left(\frac{0.1040181}{\sqrt{1191}}\right)$

$\beta_{2\ upper} = -0.5815$

$\beta_{2\ lower} = -0,5939$

F

C. (1) model 2,2 useful? → $\beta_1, \beta_2, \beta_3, \beta_4$ — T

F (2) model 2,1 useful? → β_1, β_2 — T

(3) Adjust R^2, which is better

Model 2.2 F test:

1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$

H_a: otherwise

2) Find F_{cal} & $F_{(\alpha, k-1, n-k)}$

Phutanate Techatanates
6204640814

$$F_{cal} = \frac{\frac{ESS}{TSS}/(K-1)}{\frac{RSS}{TSS}/(n-K)} \quad \text{or} \quad \frac{R^2/(K-1)}{1-R^2/(n-K)}$$

$$F_{cal} = \frac{\frac{15827.6593}{482746.692}/4}{\frac{466919.033}{482746.692}/1186} = 10.0508$$

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

```
regress  bwght  cigs   faminc fatheduc motheduc

      Source |       SS           df       MS      Number of obs   =     119
-------------+----------------------------------   F(  4,  1186)   =    10.0
       Model |  15827.6593         4  3956.91482   Prob > F        =   0.000
    Residual |  466919.033      1186   393.69227   R-squared       =   0.032
-------------+----------------------------------   Adj R-squared   =   0.029
       Total |  482746.692      1190  405.669489   Root MSE        =   19.84

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

$$F_{0.05,\,4,\,1186} = 2.21$$

Step 3:

$$F_{cal} > F_{4,1186}$$

Ho is rejected, not all slopes are 0 at the same time

∴ Suitable

Phutanate Techatanates
6204640814

Next: T-cal

β_1:

1) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

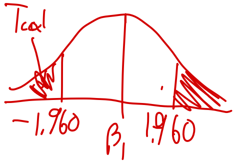
2) T_{cal} & $T_{\frac{\alpha}{2}, n-k}$

$$T_{cal} = \frac{\hat{\beta}_1 - \beta}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{-0.5894954}{0.1106172}$$

$$T_{cal} = -5.32914$$

$$T_{\frac{\alpha}{2}, n-k} = T_{0.025, 1186} = 1.960$$

3)



∴ Can Reject

C, ∴ My conclusion does not change, smoking smoking still impacts birthweight

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

```
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
      faminc |   .0538254   .0366502     Omitted for the purpose of this exam.
    fatheduc |   .4936695   .2832896
    motheduc |  -.4379234   .3197377
       _cons |   118.0741   3.500291
------------------------------------------------------------------------------
```

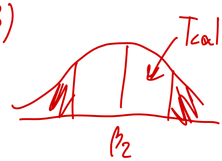
β_2: 1) $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

2) $T_{\frac{\alpha}{2}, n-k} = 1.960$

$$T_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{0.0538254}{0.0366502} = 1.4686$$

3)



∴ cannot be rejected.

Phutanate Techatanates

6204640814

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

```
regress bwght cigs faminc fatheduc motheduc
```

Source	SS	df	MS
Model	15827.6593	4	3956.91482
Residual	466919.033	1186	393.69227
Total	482746.692	1190	405.669489

Number of obs	= 1191
F(4, 1186)	= 10.05
Prob > F	= 0.0000
R-squared	= 0.0328
Adj R-squared	= 0.0295
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			

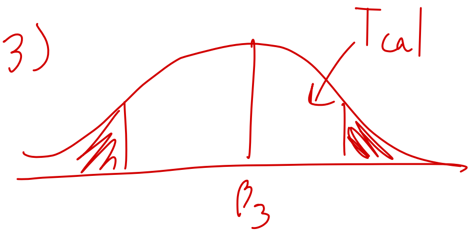
β_3 : 1) $H_0 : \beta_3 = 0$

$H_a : \beta_3 \neq 0$

2) $T_{\frac{\alpha}{2}, n-k} = 1.960$

$$T_{cal} = \frac{0.4936695}{0.2832896} = 1.72426$$

3)



$\therefore$ cannot reject H_0

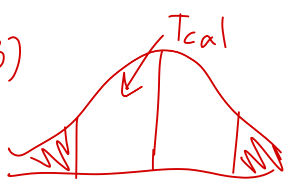
β_4 ; 1) $H_0 : \beta_4 = 0$

$H_a : \beta_4 \neq 0$

2) $T_{\frac{\alpha}{2}, n-k} = 1.96$

$$T_{cal} = \frac{\hat{\beta}_4 - 0}{\hat{\sigma}_{\hat{\beta}_4}} = \frac{-0.4379234}{0.3197377} = -1.3696$$

3)



$\therefore$ cannot reject

Model: 2,1

Phutanate Techatanates

620464081 4

Step 1: H hypothesis

$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

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

```
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
------------------------------------------------------------------------------
```

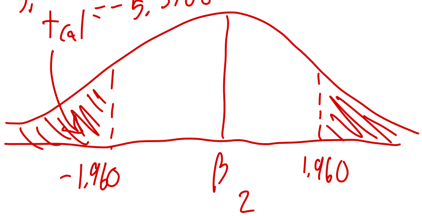
Step 2: t_{cal}, choose α

$$t_{cal} = \frac{\hat{\beta}_2 - 0}{\sigma_{\hat{\beta}_2}} = \frac{-0{,}5876985}{0{,}1090181} = -5{,}3908$$

$\alpha = 0{,}05$

$$t_{\frac{\alpha}{2}, n-k} = t_{0.025, 1191-3} = t_{0.025, 1188} = 1.960$$

Step 3: $t_{cal} = -5{,}3908$



Phutanate Techatanates

6204640814

β_2: $H_0: \beta_2 = 0$

$H_a: \beta_2 \neq 0$

$T_{\frac{\alpha}{2}, n-k} = 1.96$

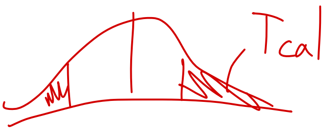
$|T_{cal}| = \dfrac{\hat{\beta}_2 - 0}{\sigma_{\hat{\beta}_2}} = \dfrac{0.0624684}{0.0324438} = 1.92543$

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

```
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   Omitted for the purpose of this exam.
      faminc |   .0624684   .0324438
       _cons |   118.5568   1.234278
------------------------------------------------------------------------------
```



∴ can reject

d. model 2.1 is better than model 2.2.

because the new variables cannot be used.

∴ model 2.2 isn't significant

Phutanate Techatanates
6204640814

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 $\alpha = 0.05$)

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

```
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]
-------------+----------------------------------------------------------------
β2     cigs |  -.5894954   .1106172
β3   faminc |   .0538254   .0366502      Omitted for the purpose of this exam.
β4 fatheduc |   .4936695   .2832896
β5 motheduc |  -.4379234   .3197377
β1    _cons |   118.0741   3.500291
------------------------------------------------------------------------------
```

Parent education = Mothereduc + fathereduc

Compare the two models

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

```
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
------------------------------------------------------------------------------
```

using marginal contribution

1) $H_0: \beta_4 = \beta_5 = 0$

H_1: otherwise

2) $$F_{cal} = \frac{ESS_{new} - ESS_{old} / \#\text{new regressors}}{RSS_{new}/(n-k_{new})}$$

$$= \frac{(15827.6593 - 14536.9538)/5-2}{466919.033/1186} = 1.6392$$

3) $\alpha = 0.05$, $F_{(\alpha, k-1, n-k)} = F_{(0.05, 4, 1186)} = 2.37$

$\therefore$ we cannot reject H_0, parents eduction is not significant.

Phutanate Techatanates
6204640814

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

$$\frac{6}{.446526442} = 13.9$$

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

$$\frac{ESS}{TSS} = MSE$$

Source	SS	df	MS
Model	35.3509 /	6 =	5.8874
Residual	187.9765 /	421	.446526442
Total	223.327441 /	427	0.5230

Number of obs	=	428
F(6 , 421)	=	13.19
Prob > F	=	0.0000
R-squared	=	0.1582
Adj R-squared	=	0.1464
Root MSE	=	.66823

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exper	.039819	.013393	2.97	0.003	.0134936	.0661444
expersq	-.0007812	.0004022	-1.94	0.053	-.0015718	9.37e-06
educ	.1078319	.0144021	7.49	0.000	.079523	.1361409
age	-.0014653	.0052925	-0.28	0.782	-.0118682	.0089377
kidslt6	-.0607106	.0887626	-0.68	0.494	-.2351836	.1137625
kidsge6	-.014591	.0278981	-0.52	0.601	-.069428	.0402459
_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?

Phutanate Techatanates
6204640814

a. $K-1 = 7-1 = 6$

$n-K = 428-7 = 421$

$n-1 = 428-1 = 427$

b. $MSR = \frac{RSS}{n-K}$

$0.4465 = \frac{RSS}{421}$

$RSS = 187.9765$

$ESS = 223.3274 - 187.9765 = 35.3509$

$MSE = 35.3509/6 = 5.8974$

$MST = 223.3274/427 = 0.5230$

c. Adjusted $R^2 = 1 - \frac{(n-1)}{(n-K)}(1-R^2)$

$= 1 - \frac{427}{421}(1-0.1582)$

$= 0.1462$

Phutanate Techatanates
6204640814

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 $F_{1,421} = 3.84$)

Perform F test and check R_2

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

MSE

Source	SS	df	MS
Model	37.24026 /	6 =	6,20671
Residual	188.434585 /	422	.446526442
Total	223.327441 /	427	0.5230150843

Number of obs	=	428
F(6 , 422)	=	13.19
Prob > F	=	0.0000
R-squared	=	0.1582
Adj R-squared	=	0.1464
Root MSE	=	.66823

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

If the excluded variable is excluded then model 3.2 is worse off. Since there's less variables R^2 of 3.2 would be lower than 3.1.

$$F_{cal} = \frac{R^2_{new} - R^2_{old} / \#new\ regressor}{1 - R^2_{new} / (n - k_{new})}$$

$$3{,}84 = \frac{0{,}1582 - R^2_{old} / 1}{1 - 0.1582 / (421)}$$

where $R^2 old$ is 3.2

$$3{,}84\left((1 - 0{,}1582)/421\right) = 0{,}1582 - R^2_{old}$$

$$0{,}0077 = 0{,}1582 - R^2_{old}$$

$$R^2_{old} = 1.505 \rightarrow \text{Max } R^2 \text{ of model } 3{,}2$$

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?

It makes economic sense because even though the older they are themore likely for them to get a raise and find high paying jobs the model does not take into account different professions where ones with higher wage tend to require more experience and eduction rather than age.