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_ใบ้. **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:

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

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)

dummy

$finance_i$ = 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.

```
    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 $\text{ROE}_i * utility_i$?

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

Source	SS	df (k)	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]
sig	lsales	.2571917	.0320348	8.03	0.000	.0194282	.3203553
sig	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
sig	utility	-.2830015	.0992337	-2.85	0.005	-.4786624	-.0873405
sig	_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 $\log(sales_i)$.

$$\log(salary_i) = 4.5881 + 0.2572 \log(sales_i) + 0.0112\ ROE_i + 0.1580\ Finance_i + 0.1809\ Consprod_i - 0.2830\ Utility_i$$

- The intercept is 4.5881 where whether we get sales, have ROE, etc. or not, we still get salary by $\log(Salary_i) = 4.5881$ or $\log^{-1}(4.5881)$.
- If $\log(sales_i)$ increase by 1 unit, $\log(salary_i)$ would increase by 0.2572 unit given other things 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.

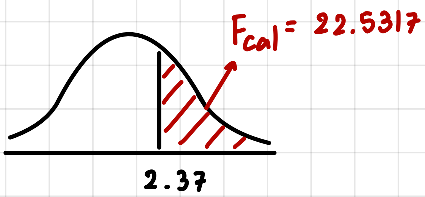
Overall significance test by using F-test :

$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_a: Otherwise

$$F_{cal} = \frac{R^2 / k}{(1-R^2)/n-k-1} = \frac{0.3569 / 5}{1-0.3569 / 209-5-1} = 22.5317$$

$$F_{cri} = F_{(k, n-k-1)\ \alpha=0.05} = F_{(5, 203)\ \alpha=0.05} = 2.37$$



$F_{cal} > F_{cri}$: can reject H_0 at the significant level of 0.05, heteroscedasticity

- significant (α) = 0.05 since CI = 95%
- Use t-test where set $\beta = 0$
- significant : All coefficients are significant

β_0 : cons : tcal = 4.5881 / 0.2950 = 15.5529

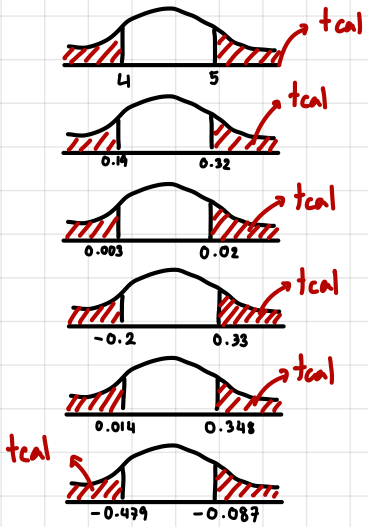
β_1 : sales : tcal = 0.2572 / 0.0320 = 8.0375

β_2 : ROE : tcal = 0.0112 / 0.3343 = 0.03350

β_3 : Finance : tcal = 0.1580 / 0.0890 = 1.7753

β_4 : Consprod : tcal = 0.1809 / 0.0848 = 2.1333

β_5 : Utility : tcal = -0.2830 / 0.0992 = -2.8528



T-testing of the model where $\alpha = 0.05$

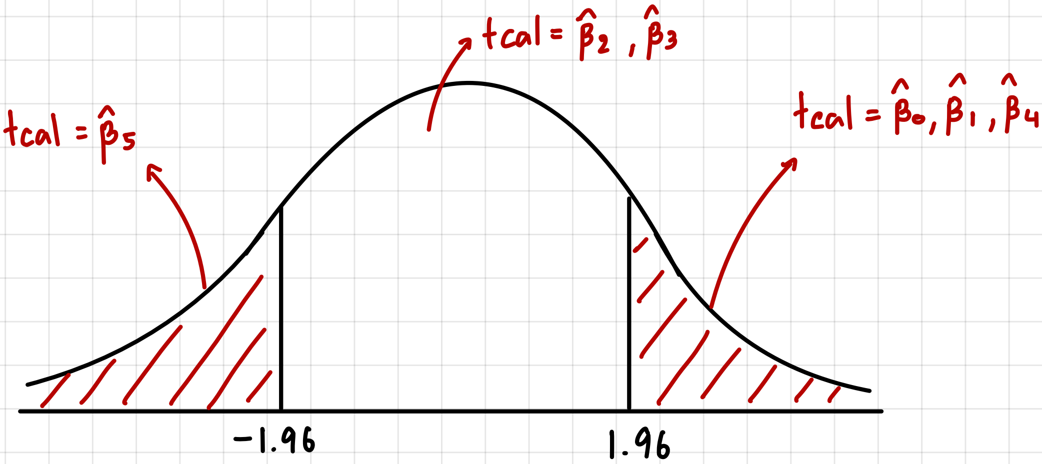
d.f. = n - k = 209 - 5 = 204

tlower : $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}}$ | tlower = -1.96

tupper : $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}}$ | tupper = 1.96

$\hat{\beta}_0 = 15.5529$ $\hat{\beta}_1 = 8.0375$ $\hat{\beta}_2 = 0.03350$

$\hat{\beta}_3 = 1.7753$ $\hat{\beta}_4 = 2.1333$ $\hat{\beta}_5 = -2.8528$



∴ We cannot reject Null hypothesis of ROE & Finance at the significant level of 0.05

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

Take a partial derivative of log (salary) to utility

$\longrightarrow \frac{\partial \log(\text{salary})}{\partial \text{Utility}} = \beta_5 \text{ Utility} = (0.2830) \text{ Utility}$

$\longrightarrow$ Transporation is the base line (intercept) $= \ln(4.5881)$ $\longrightarrow$ real number : $e^{4.5881} = 98.3075$

$\longrightarrow$ intercept of Utility $= \ln(4.5881 - 0.2830) = \ln(4.3051) \longrightarrow e^{4.3051} = 74.0766$

$e^{(4.5881)} = 98.3075$

$e^{(4.3051)} = 74.0766$

$\therefore$ Percentage different : $\frac{98.3075 - 74.0766}{98.3075}$

$= 0.2465$ or 24.6481% //

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?

- We don't put transporation$_i$ in the equation as it's already the base line to the salary of the CEO. If we put into equation and run stata anyway, the result is that we will have more errors. Further more, the result of dummy will not be 1 nor 0.

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 $\text{ROE}_i * utility_i$?

To make the interaction with dummies would effect the model that it will may or may not increase the R^2 of the relevant to the model. To test whether to make the interaction or not can be test by " Marginal Contribution". If the result in this test is to reject Null hypothesis test means that the interaction is benefit to the model

[H_0: Interaction (variables) has no marginal contribution to the model.

H_a: Otherwise]

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 (oz)
$cigs_i$ = 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 $\alpha = 0.05$) assume $\beta_1 = 0$

Model 2.1: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + u_i$ $H_0: \beta_1 = 0$ $H_a \cdot \beta_1 \neq 0$

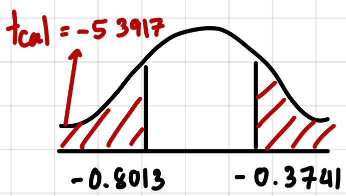
k=3

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

$bwght_i = 118\ 5568 - 0.5877 cigs_i + 0.0625 faminc_i$ $d.f. = n-k = 1191 - 3 = 1188$

$t_{cal} = \frac{-0.5877 - 0}{0.1090} = -5.3917$ significant $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = 1.96$

$\longrightarrow P[\hat{\beta}_1 - (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1}) \leqslant \hat{\beta}_1 \leqslant \hat{\beta}_1 + (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1})]$



Lower: $(-0.5877) - (1.96)(0.1090) = -0.8013$

Upper: $(-0.5877) + (1.96)(0.1090) = -0.3741$

∴ we can say that smoking has effect on birth weight 95 out of 100 times

b. Based on **Model 2.1**, construct a 99% confidence interval for β_2. assume $\beta_2 = 0$

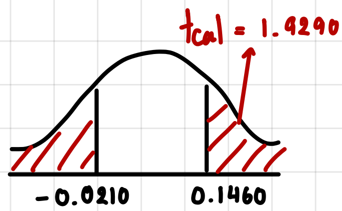
Model 2.1: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + u_i$ $H_0: \beta_2 = 0$ $H_a: \beta_2 \neq 0$

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

$bwght_i = 118\ 5568 - 0.5877 cigs_i + 0.0625 faminc_i$ $d.f. = n-k = 1191 - 3 = 1188$

$t_{cal} = \frac{0.0625 - 0}{0.0324} = 1.9290$ significant $t_{\frac{\alpha}{2}} = t_{\frac{0.01}{2}} = 2.576$

$\longrightarrow P[\hat{\beta}_2 - (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_2}) \leqslant \hat{\beta}_2 \leqslant \hat{\beta}_2 + (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_2})]$



Lower: $(0.0625) - (2.576)(0.0324) = -0.0210$

Upper: $(0.0625) + (2.576)(0.0324) = 0.1460$

∴ we can say that family income effect the birth weighted in significant level of 0.01.

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

$H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0$

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

$k = 5$

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

$\widehat{bwght}_i = 118\ 5568 - 0.5877\,cigs_i + 0.0625\,faminc_i + 0.4937\,fatherduc_i - 0.4379\,motherduc_i$

$t_{cal} = \dfrac{-0.5895 - 0}{0.1106} = -5.3300$ significant

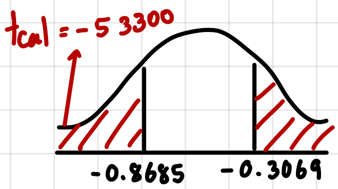
$d.f. = n-k = 1191 - 5 = 1{,}186$

$t_{\frac{\alpha}{2}} = t_{\frac{0.01}{2}} = 2.576$

$\longrightarrow P[\hat{\beta}_1 - (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1}) \leqslant \hat{\beta}_1 \leqslant \hat{\beta}_1 + (t_{\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1})]$

Lower: $(-0.5877) - (2.576)(0.1090) = -0.8685$

Upper: $(-0.5877) + (2.576)(0.1090) = -0.3069$



∴ We reject null hypothesis at the significant level of 0.05 where we can conclude that smoking effect the birth weighted.

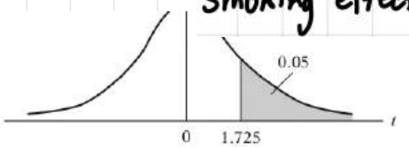
Percentage Points of the *t* Distribution

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

$Pr(t > 2.086) = 0.025$

$Pr(t > 1.725) = 0.05$ for df = 20

$Pr(|t| > 1.725) = 0.10$



Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

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. $\alpha = 0.05$

$k = 5$

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

Using F - test for test overall significance : $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_a : otherwise

$$F_{cal} = \frac{R^2 / k}{(1-R^2)/n-k-1} = \frac{0.0328/5}{(1-0.0328)/1{,}191-5-1} = \frac{0.0066}{0.0008}$$

$$F_{cal} = 8.0862$$

$F_{cal} > F_{cri}$: can reject H_0

$$F_{cri} = F_{(k, n-k-1)} = F_{(5, 1185)} = 2.37$$

$\alpha = 0.05$

heteroscedasticity

df for denominator N_2		k df for numerator N_1											
	Pr	1	2	3	4	5	6	7	8	9	10	11	12
22	.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
	.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
n-k-1 ∞	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

Using t-test for testing individually (Assume $\beta = 0$) $\alpha = 0.05$

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]
-------------+----------------------------------------------------------------
 β1     cigs |  -.5894954   .1106172
 β2   faminc |   .0538254   .0366502     Omitted for the purpose of this exam.
 β3 fatheduc |   .4936695   .2832896
 β4 motheduc |  -.4379234   .3197377
       _cons |   118.0741   3.500291
 β0
------------------------------------------------------------------------------
```

d.f. = n-k = 1191 − 5 = 1,186

$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = 1.96$

$H_0 : \beta_0 = 0$ $H_a : \beta_0 \neq 0$

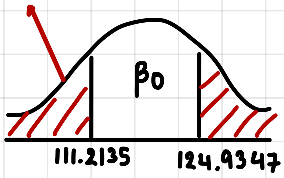
Test $\hat{\beta}_0$; $t_{cal} = 118.0741 / 3.5003 = 33.7326$ significant

$\longrightarrow P[\hat{\beta}_0 - (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_0}) \leq \hat{\beta}_0 \leq \hat{\beta}_0 + (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_0})]$

Lower: (118.0741) − (1.96)(3.5003) = 111.2135

Upper: (118.0741) + (1.96)(3.5003) = 124.9347

tcal = 33.7326



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

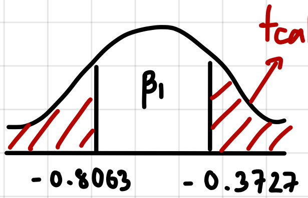
Test $\hat{\beta}_1$; $t_{cal} = -0.5895 / 0.1106 = -5.3300$ significant

$\longrightarrow P[\hat{\beta}_1 - (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_1}) \leq \hat{\beta}_1 \leq \hat{\beta}_1 + (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_1})]$

Lower: (−0.5895) − (1.96)(0.1106) = −0.8063

Upper: (−0.5895) + (1.96)(0.1106) = −0.3727

tcal = −5.3303



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

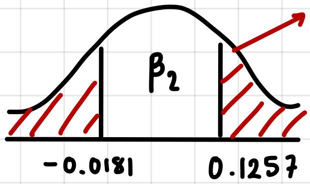
Test $\hat{\beta}_2$; $t_{cal} = 0.0538 / 0.0367 = 1.4659$ significant

$\longrightarrow P[\hat{\beta}_2 - (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_2}) \leq \hat{\beta}_2 \leq \hat{\beta}_2 + (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_2})]$

Lower: (0.0538) − (1.96)(0.0367) = −0.0181

Upper: (0.0538) + (1.96)(0.0367) = 0.1257

tcal = 1.4659



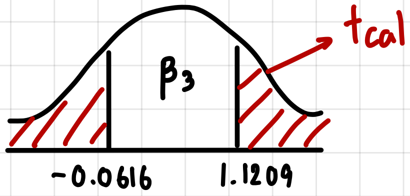
$H_0 : \beta = 0 \qquad H_a : \beta \neq 0$

Test $\hat{\beta}_3$; $t_{cal} = 0.4937/0.2833 = 1.7427$ significant

$\longrightarrow P[\hat{\beta}_3 - (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_3}) \leqslant \hat{\beta}_3 \leqslant \hat{\beta}_3 + (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_3})]$

Lower: $(0.4937) - (1.96)(0.2833) = -0.0616$

Upper: $(0.4937) + (1.96)(0.2833) = 1.1209$



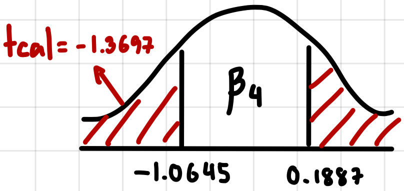
$H_0 : \beta = 0 \qquad H_a : \beta \neq 0$

Test $\hat{\beta}_4$; $t_{cal} = -0.4379/0.3197 = -1.3697$

$\longrightarrow P[\beta - (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}}) \leqslant \hat{\beta} \leqslant \hat{\beta} + (t_{\frac{\alpha}{2}} \times \sigma_{\hat{\beta}})]$

Lower: $(-0.4379) - (1.96)(0.3197) = -1.0645$

Upper: $(-0.4379) + (1.96)(0.3197) = 0.1887$

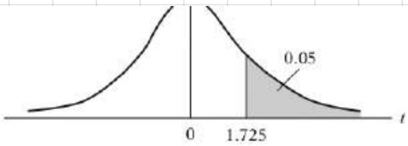


$\longrightarrow$ This is to find t-test and Confident Interval

Percentage Points of the *t* Distribution

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

$\Pr(t > 2.086) = 0.025$
$\Pr(t > 1.725) = 0.05$ for df = 20
$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

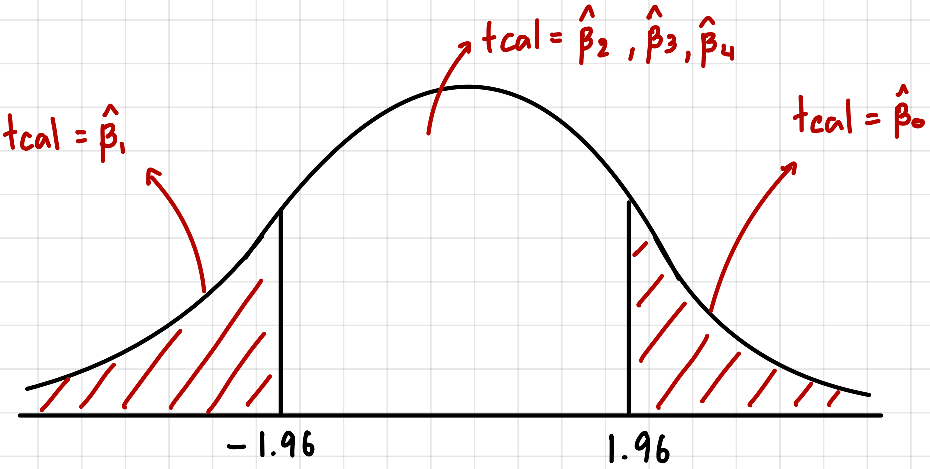
T-testing of the model where $\alpha = 0.05$

$d.f. = n-k = 209 - 5 = 204$

$t_{lower} : t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}}$	$t_{lower} = -1.96$
$t_{upper} : t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}}$	$t_{upper} = 1.96$

$\hat{\beta}_0 = 33.7326$ $\quad \hat{\beta}_1 = -5.3300$ $\quad \hat{\beta}_2 = 1.4659$

$\hat{\beta}_3 = 1.7427$ $\quad \hat{\beta}_4 = -1.3697$



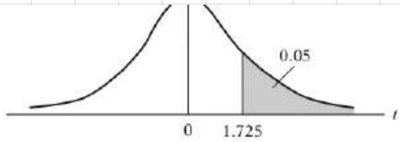
∴ We cannot reject Null hypothesis of family income & Parents' education at the significant level of 0.05

(3 estimators)

Percentage Points of the *t* Distribution

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.

$\Pr(t > 2.086) = 0.025$
$\Pr(t > 1.725) = 0.05$ for df = 20
$\Pr(|t| > 1.725) = 0.10$



Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

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 2.1: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + u_i$ $K = 3$

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

df: k; $n-k-1$; $n-1$. $F = \frac{MSR}{MSE}$

Model 2.2: $bwght_i = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + \beta_3 fathereduc_i + \beta_4 mothereduc_i + u_i$ $k = 5$

```
regress  bwght  cigs   faminc fatheduc motheduc

      Source |       SS       df       MS              Number of obs =    1191
-------------+------------------------------           F(  4,  1186) =   10.05
R      Model |  15827.6593     4  3956.91482           Prob > F      =  0.0000
E   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
------------------------------------------------------------------------------
```

df: k; $n-k-1$; $n-1$. $F = \frac{MSR}{MSE}$

We will test that the added estimators (Parents' education) is fit fit regression or not by using "Marginal Contribution"

H_0 : weighted has no marginal contribution to the model

H_a : Otherwise

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

$$= \frac{0.0328 - 0.0301 / 5 - 3}{1 - 0.0328 / (1191 - 5)}$$

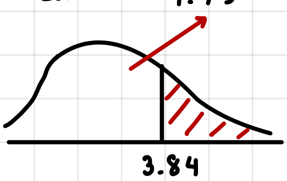
$$= \frac{0.0014}{0.0008}$$

$F_{cal} = 1.75$

$F_{cri}(1, n-3) \longrightarrow F_{cri}(1, 1188)$
$\alpha = 0.05$ $\alpha = 0.05$

$F_{cri} = 3.84$

$F_{cal} < F_{cri}$ cannot reject Null hypothesis

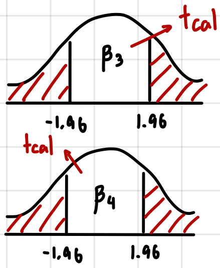


$\therefore$ We cannot reject Null hypothesis at the significant level of 0.05 where Parents' education has no marginal contribution to the model

Using t-test for testing individually (Assume $\beta = 0$) $\alpha = 0.05$

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



Test $\hat{\beta}_3$; $t_{cal} = 0.4937 / 0.2833 = 1.7427$

Test $\hat{\beta}_4$; $t_{cal} = -0.4379 / 0.3197 = -1.3697$

$t_{lower} = t_{\frac{0.05}{2}}$ | $t_{lower} = -1.96$

$t_{upper} = t_{\frac{0.05}{2}}$ | $t_{upper} = 1.96$

$\therefore$ we cannot reject Null hypothesis test or in other word we cannot say that Father and Mother education is signicifant at the level of 0.5

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	df	MS		
Model	________	___	________	Number of obs =	428
Residual	________	___	.446526442	F(___,_____) =	13.19
				Prob > F =	0.0000
				R-squared =	0.1582
				Adj R-squared =	_____
Total	223.327441	___	________	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?

$\frac{SS}{d.f.}$

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

model 3.1 included extented variables

Source	SS	df	MS
E Model	35.3304	6	5.8884
R Residual	187.9970	421	.446526442
T 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	=	0.1462
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

$$d.f.\ \text{model} = k-1 = 7-1 = 6$$

$$d.f.\ \text{Residual} = n-k = 428-7 = 421$$

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

$$R^2 = \frac{ESS}{TSS} \longrightarrow 0.1582 = \frac{ESS}{223.3274}$$

$$ESS = 35.3304 \,//$$

$$R^2 = 1 - \frac{RSS}{TSS} \longrightarrow 0.1582 = 1 - \frac{RSS}{223.3274}$$

$$RSS = 187.9970 \,//$$

$$MS = \frac{SS}{d.f.} \longrightarrow \frac{35.3304}{6} = 5.8884 \,// \quad (\text{Model})$$

$$\longrightarrow \frac{223.3274}{427} = 0.5230 \,// \quad (\text{Residual})$$

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

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

$$\bar{R}^2 = 0.1462 \,//$$

$k_{old} = k_{new} - 1 = 5$

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 $F_{1,421} = 3.84$)

→ Find R^2_{old}

We estimate to whether add more variables in the model or not by using "Marginal Contribution"

H_0 : weighted has no marginal contribution to the model

H_a : Otherwise

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

$$F_{cri}(1, n-3)_{\alpha = 0.05} = F_{cri}(1, 425)_{\alpha = 0.05} = 3.84$$

To maximize R^2 is $F_{cal} = F_{cri}$

$$3.84 = \frac{0.1582 - R^2_{old} / 6 - 5}{1 - 0.1582 / 428 - 6}$$

$$0.0077 = 0.1582 - R^2_{old}$$

$$R^2_{old} = 0.1505$$

∴ The maximum R^2 of model 3.2 is 0.1505

$\frac{SS}{d.f.}$

model 3.1 included extented variables

	Source	SS	df	MS
E	Model	35.3304	6	5.8884
R	Residual	187.9970	421	.446526442
T	Total	223.327441	427	0.5230

k_{new}

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

R_{new}

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

TABLE D.3 Upper Percentage Points of the *F* Distribution (*Continued*)

df for denom-inator N_2	Pr	df for numerator N_1: 1	2	3	4	5	6	7	8	9	10	11	12
22	.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

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?

In the economics, this making sense that the effect the hourly wage cannot say that age determined on that part since there are many ways to earn more money that the age doesn't really matter.