Question 1:

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 ($fathereduc_i$),
mother's year of education ($mothereduc_i$).

The following two regressions were estimated using data on n = 1,191 births:

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

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

Model 2.1

```
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
      _cons |   118.5568   1.234278
------------------------------------------------------------------------------
```

Omitted for the purpose of this exam.

Model 2.2

```
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
   fatheduc |   .4936695   .2832896
   motheduc |  -.4379234   .3197377
      _cons |   118.0741   3.500291
------------------------------------------------------------------------------
```

Omitted for the purpose of this exam.

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

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

(c) Based on **Model 2.1**, construct a 99% confidence interval for β_2.

(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.

(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 $\alpha = 0.05$)

Question 2: A model of wage equation is given by

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

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              Number of obs =     428
-------------+------------------------------           F(    ,    )  =   13.19
       Model |                                         Prob > F      =  0.0000
    Residual |                .446526442               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) 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?

(e) 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) 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 $F_{1,421} = 3.84$)