HW2

Omair Shafi Ahmed

7.1

College	5.46**	5.48**	5.44**
Female	-2.64**	-2.62**	-2.62**
Age		0.29**	0.29**
Ntheast			0.69**
Midwest			0.60**
South			-0.27**
Intercept	12.69**	4.40**	3.75**

7.2

A) The coefficient is statistically significant at the 5% level as the t-statistic is 26 (5.46/0.21). The 95% confidence interval is [5.46 - 1.96 x 0.21, 5.46 +1.96 x 0.21]

B) Again, the t-statistic is -13.2 (-2.64/0.20) which makes the coefficient statistically significant. The confidence interval is [-2.64 + 1.96 x 0.20, -2.64 - 1.96 x 0.20]

7.3

A) Age does determine earnings as shown by the t-statistic, which is 7.25. The confidence interval is [0.29 + 1.96 * 0.04, 0.29 - 1.96 * 0.04].

B) The difference in their earning is specified by: 5 X [0.29 + 1.96 x 0.04, 0.29 - 1.96 x 0.04] = [1.06 1.84]

7.4

A. The regional differences do seem statistically significant as the f-statistic is 6.10 and the 1% critical value is 3.78 and 6.10 > 3.78.
B. The expected difference in their earning can be specified by [-0.27 + 1.96 x 0.26, -0.27 - 1.96 x 0.26]
C. The difference between Juanita and Jennifer is expected to be between:

    [X5 (Juanita) - X5 (Jennifer)] x B5 + [X6 (Juanita) - X6 (Jennifer)] x B6 = -B5 + B6

    Including west and excluding midwest will make this computation easier and the coefficient would represent the difference between south and midwest. This will make it easier to find out the confidence interval as well.

7.5

To calculate statistical change in significance for the college coefficient, we have to calculate the t-statistic:

(5.48 - 5.29) / ((0.21^2 + 0.20^2)^1/2) = 0.6552 (Since the samples are independent and corr(X1(1998), X1(1992) = 0)

Since the t-statistic is under 1.98, it is not statistically significant.

7.6

```
. regress spendns_pc income_pc, tsscons vce(robust)

Linear regression                               Number of obs     =        342
                                                F(1, 340)         =      66.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1119
                                                Root MSE          =     347.17

------------------------------------------------------------------------------
             |               Robust
  spendns_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   income_pc |   .0141782   .0017335     8.18   0.000     .0107685     .017588
       _cons |   520.7574   54.13046     9.62   0.000     414.2846    627.2301
```

A) The coefficient of the per capita income, based on this analysis alone, most likely cannot be considered as an unbiased estimate of the causal effect of the per capita income as it is likely to have the omitted variable bias. Further analysis on the correlation between the variables could lead us to conclusively knowing if it could be considered an unbiased estimator.

B)

```
. regress spendns_pc unemploy income_pc, tsscons vce(robust)

Linear regression                               Number of obs     =        342
                                                F(2, 339)         =      40.00
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1670
                                                Root MSE          =     336.73

------------------------------------------------------------------------------
             |               Robust
  spendns_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    unemploy |   115.7982   49.08041     2.36   0.019     19.25774    212.3387
   income_pc |   .0220143   .0035741     6.16   0.000      .014984    .0290445
       _cons |   12.94953    207.389     0.06   0.950    -394.9818    420.8809
------------------------------------------------------------------------------
```

The coefficients appears to have changed. The implication of this is that in the single variate regression, the coefficient of income per capita absorbed the effect of unemployment. To confirm this we can run a correlation between unemployment and income_pc.

```
. correlate unemploy income_pc
(obs=342)

             | unemploy income~c
-------------+------------------
    unemploy |   1.0000
   income_pc |  -0.6189   1.0000
```

This confirms our hypothesis that income_pc is correlated with unemployment and the coefficient of income_pc in single-variate analysis absorbed the effect of unemployment.

C)

The coefficient for unemployment rate is: 115.798 with a standard error of 49 and a confidence interval [19.257, 212.3387]. This indicates that the municipal spending per capita goes up by $115.79 on average for every percentage increase in unemployment.

D)

```
. regress spendns_pc poverty unemploy income_pc, tsscons vce(robust)

Linear regression                               Number of obs     =        342
                                                F(3, 338)         =      45.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2371
                                                Root MSE          =     322.73

------------------------------------------------------------------------------
             |               Robust
  spendns_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     poverty |   28.68608   7.300084     3.93   0.000     14.32676     43.0454
    unemploy |   80.96017   55.10625     1.47   0.143    -27.43424    189.3546
   income_pc |   .0263618   .0033211     7.94   0.000     .0198291    .0328946
       _cons |  -185.8514   194.4673    -0.96   0.340      -568.37    196.6672
```

The coefficient on unemployment appears to have reduced from 115.79 to 80.96, whereas the coefficient of income_pc appears to have gone up from 0.22 to 0.26. This is a clear indication of the omitted variable bias where the coefficient for unemployment and income_pc appears to have had absorbed the effect of poverty. As it is likely that poverty is positively correlated with unemployment and negatively correlated with income per capita. To confirm this, we will verify the correlation for the two variables.

```
. correlate income_pc  unemploy poverty

             | income~c unemploy  poverty
-------------+---------------------------
   income_pc |   1.0000
    unemploy |  -0.6189   1.0000
     poverty |  -0.5001   0.4848   1.0000
```

With a correlation of 0.48, -0.50 and -0.61, our hypothesis has been confirmed.

E)

```
. regress spendns_pc population poverty unemploy income_pc, tsscons vce(robust)

Linear regression                               Number of obs     =        342
                                                F(4, 337)         =      51.00
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2505
                                                Root MSE          =     320.36

------------------------------------------------------------------------------
             |               Robust
  spendns_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  population |   .0012717   .0003958     3.21   0.001     .0004932    .0020502
     poverty |   22.30253   8.371596     2.66   0.008     5.835361    38.76969
    unemploy |   85.32992   55.04967     1.55   0.122    -22.95433    193.6142
   income_pc |   .0255616   .0033024     7.74   0.000     .0190656    .0320576
       _cons |  -161.0811   194.8913    -0.83   0.409    -544.4378    222.2756
------------------------------------------------------------------------------
```

The coefficient on per capita income cannot be considered as an unbiased estimate of the causal effect of per capita income on per capita non-school spending as per capita income is a function of the population and as a result, will be correlated.

F)

i) The proportion of the variance in the per capita non-school spending among these 342 municipalities that is explained by this model is it's R-sqaured: 0.25

ii) The typical magnitude of the error of the model in predicting a municipality's per capita non-school spending is it's RMSE: 320.36

G)

Municipality 1:

```
. display _b[_cons] + _b[ population]*5000 + _b[ unemploy]*2 +
_b[ income_pc]*40
> 000 + _b[ poverty]*2
1083.0077
```

Municipality 2:

```
. display _b[_cons] + _b[ population]*100000 + _b[ unemploy]*10 +
_b[ income_pc]
> *20000 + _b[ poverty]*15
1665.1594
```

H)

Municipality 1:

```
. display _b[_cons] + _b[ population]*5000 + _b[ unemploy]*2 +
_b[ income_pc]*40
> 000 + _b[ poverty]*2
1083.0077
```

Municipality 2:

```
. display _b[_cons] + _b[ population]*100000 + _b[ unemploy]*10 +
_b[ income_pc]
> *20000 + _b[ poverty]*15
1665.1594
```

The values are the same as the date is the same.

F) The unemployment vector should remain in the model because, although there is high negative correlation between unemployment and income, fundamentally, unemployment contains new information that is not present in the income vector.