Multiple Choice Questions [20 pts]

Please provide your answers to this section below:

1. 6.  
2. 7.  
3. 8.  
4. 9.  
5. 10.  

1) In the multiple regression model, the adjusted $R^2$, $\hat{R}^2$
   
   a. cannot be negative.  
   b. will never be greater than the regression $R^2$.  
   c. equals the square of the correlation coefficient $r$.  
   d. cannot decrease when an additional explanatory variable is added.  

   Answer: b

2) Under imperfect multicollinearity
   
   e. the OLS estimator cannot be computed.  
   f. two or more of the regressors are highly correlated.  
   g. the OLS estimator is biased even in samples of $n > 100$.  
   h. the error terms are highly, but not perfectly, correlated.  

   Answer: b

3) The following linear hypothesis can be tested using the $F$-test with the exception of
   
   a. $\beta_2 = 1$ and $\beta_3 = \beta_4 / \beta_5$.  
   b. $\beta_2 = 0$.  
   c. $\beta_1 + \beta_2 = 1$ and $\beta_3 = -2 \beta_4$.  
   d. $\beta_0 = \beta_1$ and $\beta_1 = 0$.  

   Answer: b
4) When there are omitted variables in the regression, which are determinants of the dependent variable, then
   a. you cannot measure the effect of the omitted variable, but the estimator of your included variable(s) is (are) unaffected.
   b. this has no effect on the estimator of your included variable because the other variable is not included.
   c. this will always bias the OLS estimator of the included variable.
   d. the OLS estimator is biased if the omitted variable is correlated with the included variable.

Answer: d

5) Imagine you regressed earnings of individuals on a constant, a binary variable (“Male”) which takes on the value 1 for males and is 0 otherwise, and another binary variable (“Female”) which takes on the value 1 for females and is 0 otherwise. Because females typically earn less than males, you would expect
   a. the coefficient for Male to have a positive sign, and for Female a negative sign.
   b. both coefficients to be the same distance from the constant, one above and the other below.
   c. none of the OLS estimators to exist because there is perfect multicollinearity.
   d. this to yield a difference in means statistic.

Answer: c

6) If you had a two regressor regression model, then omitting one variable which is relevant
   a. will have no effect on the coefficient of the included variable if the correlation between the excluded and the included variable is negative.
   b. will always bias the coefficient of the included variable upwards.
   c. can result in a negative value for the coefficient of the included variable, even though the coefficient will have a significant positive effect on \( Y \) if the omitted variable were included.
   d. makes the sum of the product between the included variable and the residuals different from 0.

Answer: c

7) To test joint linear hypotheses in the multiple regression model, you need to
   a. compare the sums of squared residuals from the restricted and unrestricted
model.
  b. use the heteroskedasticity-robust $F$-statistic.
  c. use several $t$-statistics and perform tests using the standard normal distribution.
  d. compare the adjusted $R^2$ for the model which imposes the restrictions, and the unrestricted model.

Answer: b

8) Under the least squares assumptions for the multiple regression problem (zero conditional mean for the error term, all $X_i$ and $Y_i$ being i.i.d., all $X_i$ and $u_i$ having finite fourth moments, no perfect multicollinearity), the OLS estimators for the slopes and intercept

a. have an exact normal distribution for $n > 25$.
b. are BLUE.
c. have a normal distribution in small samples as long as the errors are homoskedastic.
d. are unbiased and consistent.

Answer: d

9) If you reject a joint null hypothesis using the $F$-test in a multiple hypothesis setting, then

a. a series of $t$-tests may or may not give you the same conclusion.
b. the regression is always significant.
c. all of the hypotheses are always simultaneously rejected.
d. the $F$-statistic must be negative.

Answer: a

10) When your multiple regression function includes a single omitted variable regressor, then

a. use a two-sided alternative hypothesis to check the influence of all included variables.
b. the estimator for your included regressors will be biased if at least one of the included variables is correlated with the omitted variable.
c. the estimator for your included regressors will always be biased.
d. lower the critical value to 1.645 from 1.96 in a two-sided alternative hypothesis to test the significance of the coefficients of the included variables.

Answer: b
Problems [40 pts]

Instructions: The goal of the problem set is to understand what you are doing rather than just getting the correct result. Please show your work clearly and neatly. Please write your answers in the space provided.

1) The cost of attending your college has once again gone up. Although you have been told that education is investment in human capital, which carries a return of roughly 10% a year, you (and your parents) are not pleased. One of the administrators at your university/college does not make the situation better by telling you that you pay more because the reputation of your institution is better than that of others. To investigate this hypothesis, you collect data randomly for 100 national universities and liberal arts colleges from the 2000-2001 U.S. News and World Report annual rankings. Next you perform the following regression

\[
\text{Cost} = 7,311.17 + 3,985.20 \times \text{Reputation} - 0.20 \times \text{Size} \\
(2,058.63) \quad (664.58) \quad (0.13)
\]

\[
+ 8,406.79 \times \text{Dpriv} - 416.38 \times \text{Dlibart} - 2,376.51 \times \text{Dreligion} \\
(2,154.85) \quad (1,121.92) \quad (1,007.86)
\]

\[R^2 = 0.72, \text{SER} = 3,773.35\]

where Cost is Tuition, Fees, Room and Board in dollars, Reputation is the index used in U.S. News and World Report (based on a survey of university presidents and chief academic officers), which ranges from 1 (“marginal”) to 5 (“distinguished”), Size is the number of undergraduate students, and Dpriv, Dlibart, and Dreligion are binary variables indicating whether the institution is private, a liberal arts college, and has a religious affiliation. The numbers in parentheses are heteroskedasticity-robust standard errors.

(a) Interpret the results and indicate whether or not the coefficients are significantly different from zero. Do the coefficients have the expected sign?

Answer: An increase in reputation by one category, increases the cost by roughly $3,985. The larger the size of the college/university, the lower the cost. An increase of 10,000 students results in a $2,000 lower cost. Private schools charge roughly $8,406 more than public schools. A school with a religious affiliation is approximately $2,376 cheaper, presumably due to subsidies, and a liberal arts college also charges roughly $416 less. There are no observations close to the origin, so there is no direct interpretation of the intercept. Other than perhaps the coefficient on liberal arts colleges, all coefficients have the expected sign, although that coefficient is not significantly different from zero. All other coefficients are statistically significant at conventional levels, with the exception of the size coefficient, which carries a t-statistic of 1.54, and hence is not statistically significant at the 5% level (using a one-sided alternative hypothesis).

(b) What is the forecasted cost for a liberal arts college, which has no religious affiliation, a size of 1,500 students and a reputation level of 4.5? (All liberal arts colleges are private.)

Answer: $32,935.
(c) To save money, you are willing to switch from a private university to a public university, which has a ranking of 0.5 less and 10,000 more students. What is the effect on your cost? Is it substantial?

Answer: Roughly $12,400. Since over the four years of education, this implies approximately $50,000, it is a substantial amount of money for the average household.

(d) What is the p-value for the null hypothesis that the coefficient on Size is equal to zero? Based on this, should you eliminate the variable from the regression? Why or why not?

Answer: Using a one-sided alternative hypothesis, the p-value is 6.2 percent. Variables should not be eliminated simply on grounds of a statistical test. The sign of the coefficient is as expected, and its magnitude makes it important. It is best to leave the variable in the regression and let the reader decide whether or not this is convincing evidence that the size of the university matters.

(e) You want to test simultaneously the hypotheses that \( \beta_{\text{size}} = 0 \) and \( \beta_{\text{Dlibart}} = 0 \). Your regression package returns the \( F \)-statistic of 1.23. Can you reject the null hypothesis?

Answer: The critical value for \( F_{2,\infty} \) is 3.00 (5% level) and 4.61 (1% level). Hence you cannot reject the null hypothesis in this case.

(f) Eliminating the Size and Dlibart variables from your regression, the estimation regression becomes

\[
\hat{\text{Cost}} = 5,450.35 + 3,538.84 \times \text{Reputation} + 10,935.70 \times \text{Dpriv} - 2,783.31 \times \text{Dreligion};
\]

\[
(1,772.35) \quad (590.49) \quad (875.51) \quad (1,180.57)
\]

\( R^2 = 0.72, \ SER = 3,792.68 \)

Why do you think that the effect of attending a private institution has increased now?

Answer: Private institutions are smaller, on average, and some of these are liberal arts colleges. Both of these variables had negative coefficients.

(g) You give a final attempt to bring the effect of Size back into the equation by forcing the assumption of homoskedasticity onto your estimation. The results are as follows:

\[
\hat{\text{Cost}} = 7,311.17 + 3,985.20 \times \text{Reputation} - 0.20 \times \text{Size}
\]

\[
(1,985.17) \quad (593.65) \quad (0.07)
\]

\[
+ 8,406.79 \times \text{Dpriv} - 416.38 \times \text{Dlibart} - 2,376.51 \times \text{Dreligion}
\]

\[
(1,423.59) \quad (1,096.49) \quad (989.23)
\]

\( R^2 = 0.72, \ SER = 3,682.02 \)

Calculate the t-statistic on the Size coefficient and perform the hypothesis test that its coefficient is zero. Is this test reliable? Explain.

Answer: Although the coefficient would be statistically significant in this case, the test is unreliable and should not be used for statistical inference. There is no theoretical suggestion here that the errors might be homoskedastic. Since the standard errors are
quite different here, you should use the more reliable ones, \textit{i.e.} the heteroskedasticity-robust.

(h) What can you say about causation in the above relationship? Is it possible that \textit{Cost} affects \textit{Reputation} rather than the other way around?

\textbf{Answer:} It is very possible that the university president and chief academic officer are influenced by the cost variable in answering the \textit{U.S. News and World Report} survey. If this were the case, then the above equation suffers from simultaneous causality bias, a topic that will be covered in a later chapter. However, this poses a serious threat to the internal validity of the study.

2) You have collected data for 104 countries to address the difficult questions of the determinants for differences in the standard of living among the countries of the world. You recall from your macroeconomics lectures that the neoclassical growth model suggests that output per worker (per capita income) levels are determined by, among others, the saving rate and population growth rate. To test the predictions of this growth model, you run the following regression:

\begin{equation}
\text{RelPersInc} = 0.339 - 12.894 \times n + 1.397 \times s_k, R^2 = 0.621, SER = 0.177
\end{equation}

where \text{RelPersInc} is GDP per worker relative to the United States, \text{n} is the average population growth rate, 1980-1990, and \text{s_k} is the average investment share of GDP from 1960 to 1990 (remember investment equals saving). Numbers in parentheses are for heteroskedasticity-robust standard errors.

(a) Interpret the results. Do the signs correspond to what you expected them to be? Explain.

\textbf{Answer:} The Solow growth model predicts higher productivity with higher saving rates and lower population growth. The signs therefore correspond to prior expectations. A 10 percent point increase in the saving rate results in a roughly 14 percent increase in per capita income relative to the United States. Lowering the population growth rate by 1 percent results in a 13 percent higher per capita income relative to the United States. It is best not to interpret the intercept. The regression explains approximately 62 percent of the variation in per capita income among the 104 countries of the world.

(b) Calculate the \textit{t}-statistics and test whether or not each of the population parameters are significantly different from zero.

\textbf{Answer:} The \textit{t}-statistics for population growth and the saving rate are –4.06 and 6.10, making both coefficients significantly different from zero at conventional levels of significance.

(c) The overall \textit{F}-statistic for the regression is 79.11. What is the critical value at the 5\% and 1\% level? What is your decision on the null hypothesis?

\textbf{Answer:} The critical value is 3.00 and 4.61 respectively, allowing you to reject the null hypothesis that all slope coefficients are zero.

(d) You remember that human capital in addition to physical capital also plays a role in determining the standard of living of a country. You therefore collect additional data on the average educational attainment in years for 1985, and add this variable (\textit{Educ}) to the above regression. This results in the modified regression output:
\[ \text{RelPersInc} = 0.046 - 5.869 \times n + 0.738 \times s_K + 0.055 \times \text{Educ}, R^2=0.775, \text{SER} = 0.1377 \]

(0.079) (2.238) (0.294) (0.010)

How has the inclusion of Educ affected your previous results?

**Answer:** The coefficient on the population growth rate is roughly half of what it was originally, while the coefficient on the saving rate has approximately doubled. The regression \( R^2 \) has increased significantly.

(e) Upon checking the regression output, you realize that there are only 86 observations, since data for Educ is not available for all 104 countries in your sample. Do you have to modify some of your statements in (d)?

**Answer:** When comparing results, you should ensure that the sample is identical, since comparisons are not valid otherwise. In addition, there are now less than 100 observations, making inference based on the standard normal distribution problematic.

(f) Brazil has the following values in your sample: \( \text{RelPersInc} = 0.30, n = 0.021, s_K = 0.169, \text{Educ} = 3.5 \). Does your equation overpredict or underpredict the relative GDP per worker? What would happen to this result if Brazil managed to double the average educational attainment?

**Answer:** The predicted value for Brazil is 0.240. Hence the regression underpredicts Brazil’s per capita income. Increasing Educ to 7.0 would result in a predicted per capita income of 0.43, which is a substantial increase from both its current actual position and the previously predicted value.

3) (Requires Appendix Material) The rule-of-thumb \( F \)-statistic is given by the formula

\[
F = \frac{(\text{SSR}_{\text{restricted}} - \text{SSR}_{\text{unrestricted}})/q}{\text{SSR}_{\text{unrestricted}}/(n - k_{\text{unrestricted}} - 1)}
\]

where \( \text{SSR}_{\text{restricted}} \) is the sum of squared residuals from the restricted regression, \( \text{SSR}_{\text{unrestricted}} \) is the sum of squared residuals from the unrestricted regression, \( q \) is the number of restrictions under the null hypothesis, and \( k_{\text{unrestricted}} \) is the number of regressors in the unrestricted regression. Prove that this formula is the same as the following formula based on the regression \( R^2 \) of the restricted and unrestricted regression:

\[
F = \frac{(R^2_{\text{unrestricted}} - R^2_{\text{restricted}})/q}{(1 - R^2_{\text{restricted}})/(n - k_{\text{unrestricted}} - 1)}
\]

**Answer:** Since \( R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} \), \( \text{SSR} = \text{TSS}(1 - R^2) \). Substitution into the first equation then results in the second equation, once the “1” in the numerator is canceled, and the TSS is factored out in the numerator.
EViews Exercise [40 pts]

Part 1

![Scatter plots showing TESTSCR vs. EL_PCT, TESTSCR vs. MEAL_PCT, and TESTSCR vs. CALW_PCT.]

Part 2

Dependent Variable: TESTSCR
Method: Least Squares
Date: 05/14/03   Time: 15:21
Sample: 1 420
Included observations: 420
White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>698.9330</td>
<td>10.36436</td>
<td>67.43619</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-2.279808</td>
<td>0.519489</td>
<td>-4.38557</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.051240  Mean dependent var 654.1565
Adjusted R-squared 0.048970  S.D. dependent var 19.05335
S.E. of regression 18.58097  Akaike info criterion 8.686903
Sum squared resid 144315.5  Schwarz criterion 8.706143
Log likelihood -1822.250  F-statistic 22.57511
Durbin-Watson stat 0.129062  Prob(F-statistic) 0.000003
### Regression Results

**Dependent Variable:** TESTSCR  
**Method:** Least Squares  
**Date:** 05/14/03  **Time:** 15:23  
**Sample:** 1 420  
**Included observations:** 420  
**White Heteroskedasticity-Consistent Standard Errors & Covariance**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>686.0322</td>
<td>8.728224</td>
<td>78.59930</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-1.101296</td>
<td>0.432847</td>
<td>-2.544307</td>
<td>0.0113</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.649777</td>
<td>0.031032</td>
<td>-20.93909</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| R-squared | 0.426431   | Mean dependent var | 654.1565 |
| Adjusted R-squared | 0.423680 | S.D. dependent var | 19.05335 |
| S.E. of regression | 14.46448 | Akaike info criterion | 8.188387 |
| Sum squared resid | 87245.29 | Schwarz criterion | 8.217246 |
| Log likelihood | -1716.561 | F-statistic | 155.0136 |
| Durbin-Watson stat | 0.685575 | Prob(F-statistic) | 0.000000 |

---

**Dependent Variable:** TESTSCR  
**Method:** Least Squares  
**Date:** 05/14/03  **Time:** 15:24  
**Sample:** 1 420  
**Included observations:** 420  
**White Heteroskedasticity-Consistent Standard Errors & Covariance**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>700.1500</td>
<td>5.568450</td>
<td>125.7352</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-0.998309</td>
<td>0.270080</td>
<td>-3.696348</td>
<td>0.0002</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.121573</td>
<td>0.032832</td>
<td>-3.702926</td>
<td>0.0002</td>
</tr>
<tr>
<td>MEAL_PCT</td>
<td>-0.547346</td>
<td>0.024107</td>
<td>-22.70460</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| R-squared | 0.774516   | Mean dependent var | 654.1565 |
| Adjusted R-squared | 0.772890 | S.D. dependent var | 19.05335 |
| S.E. of regression | 9.080079 | Akaike info criterion | 7.259521 |
| Sum squared resid | 34298.30 | Schwarz criterion | 7.298000 |
| Log likelihood | -1520.499 | F-statistic | 476.3064 |
| Durbin-Watson stat | 1.437595 | Prob(F-statistic) | 0.000000 |

---

**Dependent Variable:** TESTSCR  
**Method:** Least Squares  
**Date:** 05/14/03  **Time:** 15:27  
**Sample:** 1 420  
**Included observations:** 420  
**White Heteroskedasticity-Consistent Standard Errors & Covariance**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>697.9987</td>
<td>6.920369</td>
<td>100.8615</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-1.307984</td>
<td>0.339076</td>
<td>-3.857494</td>
<td>0.0001</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.487620</td>
<td>0.029582</td>
<td>-16.48352</td>
<td>0.0000</td>
</tr>
<tr>
<td>CALW_PCT</td>
<td>-0.789965</td>
<td>0.067660</td>
<td>-11.67557</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| R-squared | 0.628543   | Mean dependent var | 654.1565 |
| Adjusted R-squared | 0.625864 | S.D. dependent var | 19.05335 |
| S.E. of regression | 11.65429 | Akaike info criterion | 7.758704 |
| Sum squared resid | 56502.17 | Schwarz criterion | 7.797183 |
| Log likelihood | -1625.328 | F-statistic | 234.6381 |
Dependent Variable: TESTSCR
Method: Least Squares
Date: 05/14/03   Time: 15:26
Sample: 420
Included observations: 420

White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>700.3918</td>
<td>5.537418</td>
<td>126.4835</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-1.014353</td>
<td>0.268861</td>
<td>-3.772775</td>
<td>0.0002</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.129822</td>
<td>0.036258</td>
<td>-3.580509</td>
<td>0.0004</td>
</tr>
<tr>
<td>MEAL_PCT</td>
<td>-0.528619</td>
<td>0.038117</td>
<td>-13.86844</td>
<td>0.0000</td>
</tr>
<tr>
<td>CALW_PCT</td>
<td>-0.047854</td>
<td>0.058654</td>
<td>-0.815863</td>
<td>0.4150</td>
</tr>
</tbody>
</table>

R-squared 0.774850     Mean dependent var 654.1565
Adjusted R-squared 0.772680     S.D. dependent var 19.05335
S.E. of regression 9.084273     Akaike info criterion 7.262800
Sum squared resid 34247.46     Schwarz criterion 7.310898
Log likelihood -1520.188     F-statistic 357.0541
Durbin-Watson stat 1.094470 Prob(F-statistic) 0.000000

Part 3

Dependent Variable: TESTSCR_STD
Method: Least Squares
Date: 05/14/03   Time: 15:47
Sample: 420
Included observations: 420

White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.350057</td>
<td>0.543965</td>
<td>4.320233</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-0.119654</td>
<td>0.027265</td>
<td>-4.388557</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.051240     Mean dependent var 2.52E-06
Adjusted R-squared 0.048970     S.D. dependent var 1.000000
S.E. of regression 0.975207     Akaike info criterion 2.792417
Sum squared resid 397.5303     Schwarz criterion 2.811657
Log likelihood -584.4076     F-statistic 22.57511
Durbin-Watson stat 1.429595 Prob(F-statistic) 0.000000

Dependent Variable: TESTSCR_STD
Method: Least Squares
Date: 05/14/03   Time: 15:47
Sample: 420
Included observations: 420

White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.672973</td>
<td>0.458094</td>
<td>3.652032</td>
<td>0.0003</td>
</tr>
<tr>
<td>STR</td>
<td>-0.057801</td>
<td>0.022718</td>
<td>-2.544307</td>
<td>0.0113</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.034103</td>
<td>0.001629</td>
<td>-20.93909</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.426431     Mean dependent var 2.52E-06
Adjusted R-squared 0.423680  S.D. dependent var 1.000000
S.E. of regression 0.759157  Akaike info criterion 2.293901
Sum squared resid 240.3252  Schwarz criterion 2.322760
Log likelihood -478.7192  F-statistic 155.0136
Durbin-Watson stat 0.685575  Prob(F-statistic) 0.000000

Dependent Variable: TESTSCR_STD
Method: Least Squares
Date: 05/14/03   Time: 15:46
Sample: 1 420
Included observations: 420
White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.413931</td>
<td>0.292256</td>
<td>8.259654</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-0.052395</td>
<td>0.014175</td>
<td>-3.696348</td>
<td>0.0002</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.006381</td>
<td>0.001723</td>
<td>-3.702926</td>
<td>0.0002</td>
</tr>
<tr>
<td>MEAL_PCT</td>
<td>-0.028727</td>
<td>0.001265</td>
<td>-22.70460</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.774516  Mean dependent var 2.52E-06
Adjusted R-squared 0.772890  S.D. dependent var 1.000000
S.E. of regression 0.476561  Akaike info criterion 1.365035
Sum squared resid 94.47783  Schwarz criterion 1.403514
Log likelihood -282.6574  F-statistic 476.3064
Durbin-Watson stat 1.437595  Prob(F-statistic) 0.000000

Dependent Variable: TESTSCR_STD
Method: Least Squares
Date: 05/14/03   Time: 15:46
Sample: 1 420
Included observations: 420
White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.301023</td>
<td>0.363210</td>
<td>6.335239</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-0.068649</td>
<td>0.017796</td>
<td>-3.857494</td>
<td>0.0001</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.025592</td>
<td>0.001553</td>
<td>-16.48352</td>
<td>0.0000</td>
</tr>
<tr>
<td>CALW_PCT</td>
<td>-0.041461</td>
<td>0.003551</td>
<td>-11.67557</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.628543  Mean dependent var 2.52E-06
Adjusted R-squared 0.625864  S.D. dependent var 1.000000
S.E. of regression 0.611666  Akaike info criterion 1.864219
Sum squared resid 155.6404  Schwarz criterion 1.902697
Log likelihood -387.4859  F-statistic 234.6381
Durbin-Watson stat 1.094470  Prob(F-statistic) 0.000000

Dependent Variable: TESTSCR_STD
Method: Least Squares
Date: 05/14/03   Time: 15:44
Sample: 1 420
Included observations: 420
White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.301023</td>
<td>0.363210</td>
<td>6.335239</td>
<td>0.0000</td>
</tr>
<tr>
<td>STR</td>
<td>-0.068649</td>
<td>0.017796</td>
<td>-3.857494</td>
<td>0.0001</td>
</tr>
<tr>
<td>EL_PCT</td>
<td>-0.025592</td>
<td>0.001553</td>
<td>-16.48352</td>
<td>0.0000</td>
</tr>
<tr>
<td>CALW_PCT</td>
<td>-0.041461</td>
<td>0.003551</td>
<td>-11.67557</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.628543  Mean dependent var 2.52E-06
Adjusted R-squared 0.625864  S.D. dependent var 1.000000
S.E. of regression 0.611666  Akaike info criterion 1.864219
Sum squared resid 155.6404  Schwarz criterion 1.902697
Log likelihood -387.4859  F-statistic 234.6381
Durbin-Watson stat 1.094470  Prob(F-statistic) 0.000000
The coefficients from these regressions show, for a one unit increase in the regressor, how many standard deviations away from the mean will the test score vary.

**Part 4**

(a) Specification 3. The adjusted R-square increases when we add percent of English learners and percent eligible for subsidized lunch as independent variables. Addition of percent of public income assistance variable does not increase the explanatory power of the regression. Also, the coefficient on this variable is insignificant. These results are the same originally obtained since the transformation of the dependent variable is such that it does not affect the fit of the model.

(b) Here is the Eviews output

Wald Test:
Equation: Untitled

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Value</th>
<th>df</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-statistic</td>
<td>0.388461</td>
<td>(2, 415)</td>
<td>0.6783</td>
</tr>
<tr>
<td>Chi-square</td>
<td>0.776922</td>
<td>2</td>
<td>0.6781</td>
</tr>
</tbody>
</table>

Null Hypothesis Summary:

<table>
<thead>
<tr>
<th>Normalized Restriction (= 0)</th>
<th>Value</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(2) - 2*C(4)</td>
<td>0.002251</td>
<td>0.014142</td>
</tr>
<tr>
<td>C(5)</td>
<td>-0.002512</td>
<td>0.003078</td>
</tr>
</tbody>
</table>

The p-value for F-statistics is 0.6783, hence we fail to reject the null hypothesis.

(c) The book offers a good way to check this answer against your interpretation of the results.