University of Canterbury
ECON 213

Professor Bob Reed
Second Semester 2009

WEEK #5 ASSIGNMENT

***DUE*: Monday, August 10th**

Do the following exercises:

1A. Read Section 5-2 on page 94-95 of the textbook.

NOTE: The purpose of including interaction terms is that they allow the effect of one explanatory variable, say age, to vary depending on the size of another variable, say education.

1B. (Estimate an Income Equation with an AGE-EDUCATION Interaction Variable.) Create a folder on your P: drive called "WEEK5."

Download the Excel spreadsheet file "Income" from the course website: "*http://www.econ.canterbury.ac.nz/personal_pages/bob_reed/ Courses/ECON_213/Week5/index.html*" and save it into your WEEK5 folder.

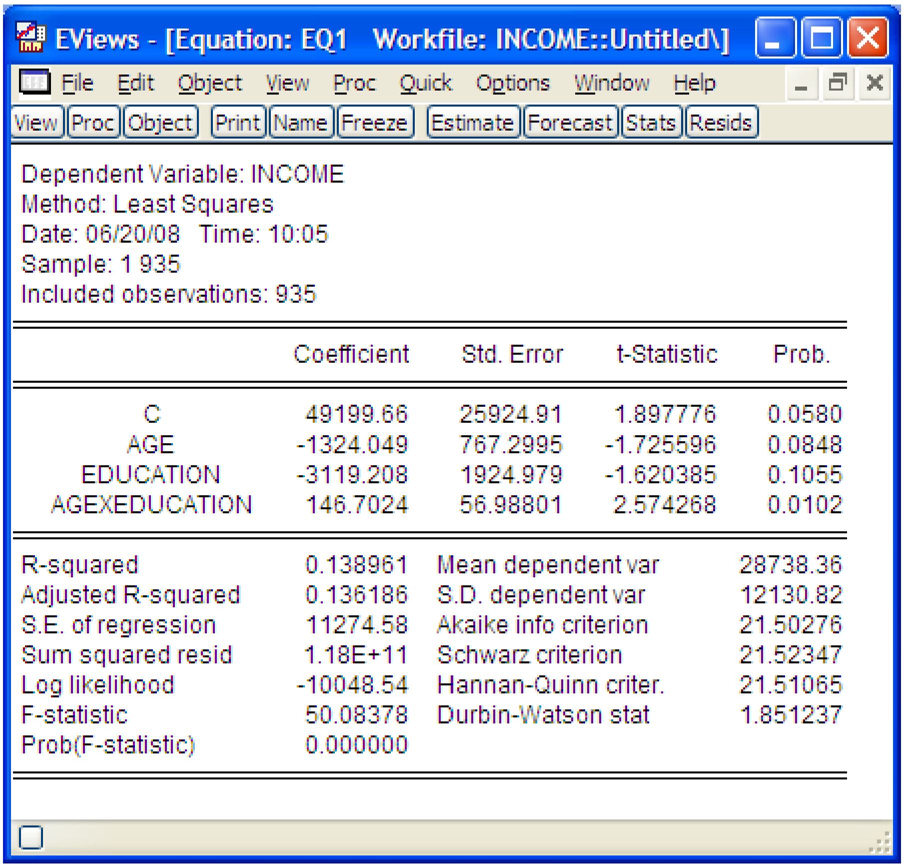
The data set "Income" consists of 935 observations. After you download the EXCEL spreadsheet, inspect it to confirm that it contains data on workers' income, age, and education.

Create a new EViews workfile, and import the "Income" data into this workfile. Name the new workfile "Income" and save it into your WEEK5 folder.

Create an "interaction variable" that is the product of age times education. Name it AGExEDUCATION. (HINT: You will use the same procedure that you used to create a "slope dummy" in last week's assignment. Accordingly, you may find it helpful to review Exercise #7B from the Week #4 Assignment.)

Estimate the regression model of Equation (5-14) on page 94.

Your regression results should be the same as those reported below. (NOTE: You will not obtain the same regression results as TABLE 5-F in the text because the data are different.)



EViews - [Equation: EQ1 Workfile: INCOME::Untitled\]

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: INCOME
Method: Least Squares
Date: 06/20/08 Time: 10:05
Sample: 1 935
Included observations: 935

	Coefficient	Std. Error	t-Statistic	Prob.
C	49199.66	25924.91	1.897776	0.0580
AGE	-1324.049	767.2995	-1.725596	0.0848
EDUCATION	-3119.208	1924.979	-1.620385	0.1055
AGEXEDUCATION	146.7024	56.98801	2.574268	0.0102

R-squared	0.138961	Mean dependent var	28738.36
Adjusted R-squared	0.136186	S.D. dependent var	12130.82
S.E. of regression	11274.58	Akaike info criterion	21.50276
Sum squared resid	1.18E+11	Schwarz criterion	21.52347
Log likelihood	-10048.54	Hannan-Quinn criter.	21.51065
F-statistic	50.08378	Durbin-Watson stat	1.851237
Prob(F-statistic)	0.000000		

1C. Answer Questions #1-12.

2A. Review hypothesis testing from the Week #3 Assignment.

2B. Answer Question #13.

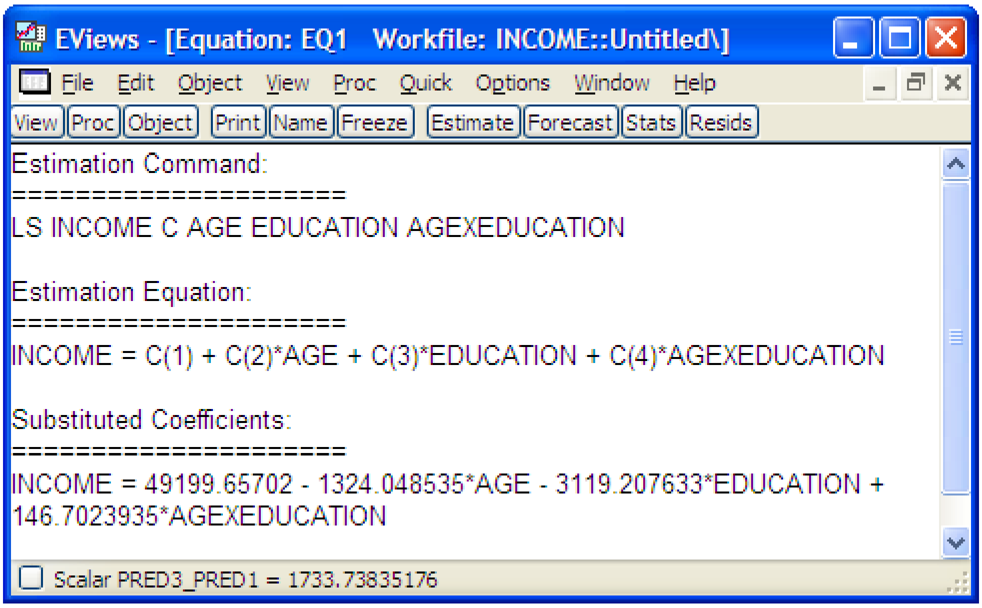
3A. (Hypothesis Testing of More Than One Coefficient.) Consider again the regression model:

$$INCOME = B_0 + B_1\ AGE + B_2\ EDUCATION + B_3\ AGExEDUCATION + e\ .$$

The problem with testing whether age affects income in this model is that the variable AGE appears twice: It appears once as itself in the variable AGE, and again in the interaction variable AGExEDUCATION. A test of whether age affects income must take into account both appearances of age. Specifically, we need to test whether <u>both</u> B_{AGE} and $B_{AGExEDUCATION}$ equal zero. When more than one coefficient is being tested, we need to use **the F-test**.

The F-test works just like a two-sided t-test except -- you guessed it – we use an F-statistic rather than a t-statistic. The textbook provides a discussion of how this statistic is constructed (you can read pages 96-100 if you want), but I will tell you all you have to know right here. I will show you how to do the test for whether EDUCATION affects income, and leave the test for AGE for you to do in the next question.

In this next exercise, you will test whether EDUCATION affects income. Note that EDUCATION, like AGE, appears twice in the regression model. Therefore, you need to test whether both $B_{EDUCATION}$ and $B_{AGExEDUCATION}$ equal zero. Consider the "Representations View" of your regression results from Exercise #1B.

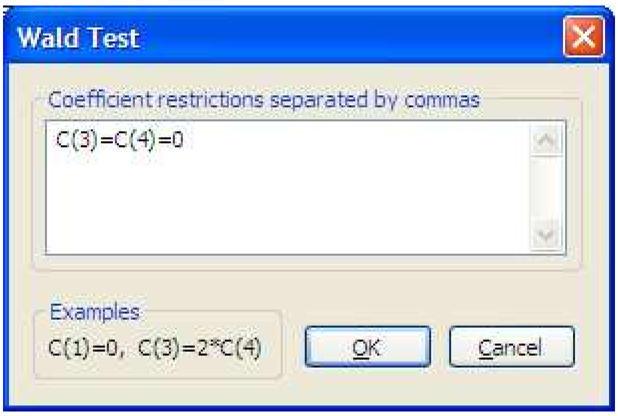


Note that a test of whether $B_{EDUCATION} = 0$ and $B_{AGExEDUCATION} = 0$ is represented under "Estimation Equation:" by C(3) = 0 and C(4) = 0. To have EViews test this for you, click on the "View" button in the Equation window, and then select **Coefficient Tests / Wald - Coefficient Restrictions…** .

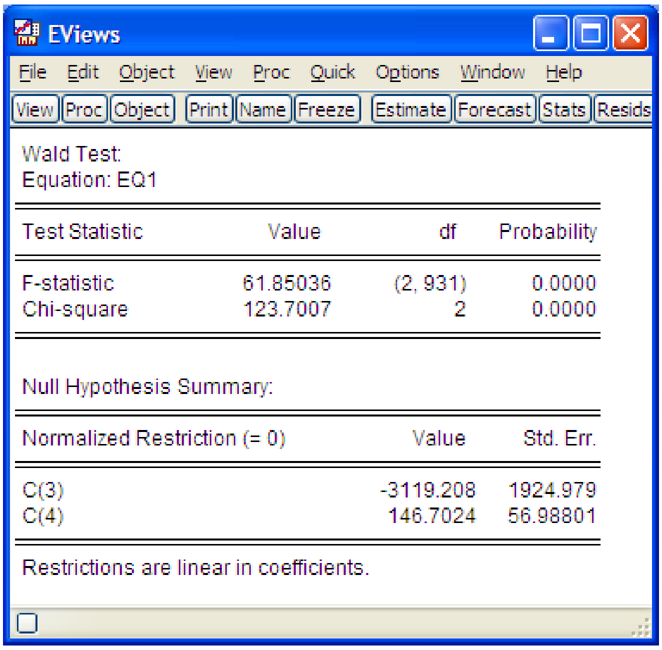
A "Wald Test" dialog box should open up. Type the following in the work area:

```
C(3)=0, C(4)=0
```

Your "Wald Test" dialog box should now look as follows:



Click “OK”. A dialog box with the following test results should appear:



Wald Test:
Equation: EQ1

Test Statistic	Value	df	Probability
F-statistic	61.85036	(2, 931)	0.0000
Chi-square	123.7007	2	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	-3119.208	1924.979
C(4)	146.7024	56.98801

Restrictions are linear in coefficients.

In this test, the F-statistic is 61.85036 and the corresponding p-value is 0.0000 (it’s not exactly equal to zero, but in this case the actual p-value is rounded off to zeros through the fourth decimal place).

Just as in the case of a two-sided hypothesis test, you compare the p-value with the level of statistical significance. If the p-value is less than the significance level, you “Reject.” If the p-value is greater than the significance level, you “Cannot Reject.”

You should confirm that the completed “Hypothesis Test Table” for a test of whether a person’s education affects their income is the following:

1)	H_0: $B_{EDUCATION} = B_{AGExEDUCATION} = 0$	(A person’s education does not affect their income)
	H_A : At least one of $B_{EDUCATION}$ or $B_{AGExEDUCATION}$ is not equal to zero	(A person’s education does affect their income)
2)	LEVEL OF STATISTICAL SIGNIFICANCE = 0.05	
3)	STATISTICS: F-statistic = 61.850; p-value = 0.00	
4)	DECISION = Reject	

We conclude that education affects a person’s income.

<u>NOTE</u>: F-tests can be used to test any "two-sided" hypotheses involving more than one coefficient. That is, they can be used to test that any subset of the coefficients are all equal to zero (such as, $B_{AGE} = B_{AGExEDUCATION} = 0$). However, they cannot be used to test "one-sided" hypotheses involving more than one coefficient. What should you do in this case? That is a topic for advanced econometrics classes!

3B. Answer Question #14.

4A. Read Section 5-4 on page 105 of the text.

<u>NOTE</u>: Polynomial models are like interaction terms in that they allow the effect of one variable, say X, to vary depending on the value of another variable. Only in this case, the "other" variable is X! Consider the "Average Cost" example of Figure 5-3 on page 107. When the number of cars being produced is relatively small, an increase in car production lowers Average Costs. However, when the number of cars being produced is relatively large, further increases in car production cause Average Costs to rise.

4B. (How to Estimate a Model with Non-linear Effects Using Polynomial Terms.) Download the Excel spreadsheet file "ProductionCosts" from the course website: "*http://www.econ.canterbury.ac.nz/personal_pages/bob_reed/Courses/ECON_213/Week5/index.html*" and save it into your WEEK5 folder.

The data set "ProductionCosts" consists of 20 observations. After you download the EXCEL spreadsheet, open it up and check it against TABLE 5-M on page 106 of the text.

Create a new EViews workfile, and import the "ProductionCosts" data into this workfile. Name the new workfile "ProductionCosts" and save it into your WEEK5 folder.

We will estimate the model:

$\text{AVECOST} = B_0 + B_1\,\text{CARS} + B_2\,\text{CARS2} + e$,

where $\text{CARS2} = \text{CARS}^2$. Your data set does not include the variable CARS2 so you will have to create it. Since CARS2 = CARS x CARS, you could create it using the same EViews commands you previously employed to generate interaction terms. Alternatively, you can use the following commands:

```
genr CARS2 = CARS^2
```

The "^" operator (usually located above the number 6 on your keyboard) raises the respective variable to a given power. For example, if we wanted to generate a variable that was the cube of CARS, we would use `CARS^3`.

Create the variable CARS2 and use the "Show" button to make sure that you have generated it correctly. Then estimate the model and compare your results with those in TABLE 5-N on page 107.

4C. Answer Questions #15-21.

5A. Read Section 5-5 on page 105-110 of the textbook.

5B. (Choosing Amongst Alternative Specifications of Variables.) You have now been exposed to a number of ways to enter variables into your model specifications. In estimating the effect of an explanatory variable X on a dependent variable Y, you can do the following: (i) regress Y on X; (ii) regress Y on X and an interaction variable of X with something else; (iii) regress Y on X and X^2 (and maybe even higher-order X variables); (iv) regress Y on ln(X); (v) regress ln(Y) on X; and (vi) regress ln(Y) on ln(X).

What a bewildering number of choices!!! Which one should you choose? The following table offers some helpful advice:

Regress:	***When…***
Y on X	X has a constant effect on Y (that is, a unit increase in X is expected to always have the same numerical effect on Y)
Y on X and an interaction variable of X with something else	The effect of X on Y is expected to depend on another variable (for example, the effect of education on income is expected to be greater over time as workers get older)
Y on X and X^2 (and possibly higher-order X variables)	The effect of X on Y is expected to change signs, so that sometimes the effect is positive and sometimes negative (for example, the production cost example of Figure 5-3 on page 107)
Y on ln(X)	The effect of X on Y is expected to diminish as X gets bigger (for example, X is hours studying, and Y is your grade)
ln(Y) on X	The effect of X on Y is expected to get larger as X gets bigger (for example, most researchers believe that an extra year of education has a larger effect for a person who already has a lot of education, compared to a person with just a little education – this is why most wage/income equations log their dependent variable)
ln(Y) on ln(X)	The effect of X on Y is expected to have "constant elasticity;" i.e., a one percent increase in X is expected to result in a constant percent change in Y (these models are commonly used when the research interest consists of estimating a "price" or "income" elasticity

5C. Read Exercise #13 on page 113 of the textbook. Then download the Excel spreadsheet file "Consumption" from the course website: "*http://www.econ.canterbury.ac.nz/personal_pages/bob_reed/Courses/ECON_213/Week5/index.html*" and save it into your WEEK5 folder.

The data set "Consumption" consists of 30 observations. After you download the EXCEL spreadsheet, open it up and check it against the data on page 114 of the text.

Create a new EViews workfile, and import the "Consumption" data into this workfile. Name the new workfile "Consumption" and save it into your WEEK5 folder.

We are going to estimate 4 different models and compare them:

1) $\text{CONSUMPTION} = B_0 + B_1 \text{ INCOME} + B_2 \text{ WEALTH} + e$

2) $\text{CONSUMPTION} = B_0 + B_1 \ln(\text{INCOME}) + B_2 \ln(\text{WEALTH}) + e$

3) $\ln(\text{CONSUMPTION}) = B_0 + B_1 \text{ INCOME} + B_2 \text{ WEALTH} + e$

4) $\ln(\text{CONSUMPTION}) = B_0 + B_1 \ln(\text{INCOME}) + B_2 \ln(\text{WEALTH}) + e$

Use the following command to create ln(CONSUMPTION):

```
genr lnconsumption = log(consumption)
```

Use similar commands to create the variables ln(INCOME) and ln(WEALTH).

Estimate each of the four models above, naming them "EQ1" through "EQ4", respectively. The following table will help you to interpret the coefficients you estimate in each of the models.

Form of Dependent Variable	***Form of Explanatory Variable***	***If $B_1 = 0.5$, then…***
CONSUMPTION	INCOME	a one unit (1 dollar) increase in INCOME causes consumption to increase by 0.5 units (50 cents)
CONSUMPTION	ln(INCOME)	a 1% increase in INCOME causes consumption to increase by 0.5/100 = 0.005 units (half a penny)
ln(CONSUMPTION)	INCOME	a one unit (1 dollar) increase in INCOME causes consumption to increase by 50%
ln(CONSUMPTION)	ln(INCOME)	a 1% increase in INCOME causes consumption to increase by 0.5% percent

5D. Answer Questions #22-25.

5E. (Making Predictions When the Explanatory Variable is in Log Form.) Use the regression results from Model 2) to predict the value of CONSUMPTION when INCOME = 20,000 and WEALTH = 50,000. Typing the following line in the "Command Window" work area and hitting "enter" will produce the predicted value, "pred1".

```
scalar pred1 = C(1) + C(2)*LOG(20000) + C(3)*LOG(50000)
```

WARNING!! Be sure you re-estimate Model 2) immediately before calculating your prediction. This will insure that Eviews uses the correct c(1) through c(3) coefficients. If you do the problem successfully, you should obtain a predicted value of $43,906.09.

5F. Answer Question #26.

<u>NOTE</u>: Unfortunately, it is difficult to use your estimated results to make predictions when the dependent variable is in log form. Therefore, we won't cover this in our class. But if you are interested, you can check out Jeffrey Wooldridge's textbook, <u>Introductory Econometrics</u>, Fourth Edition, Thomson South-Western, 2009, pages 212-214.

5G. (Selecting the Best Model When the Dependent Variables Are the Same.) You estimated 4 models in Exercise #5C. Which one is the best? It turns out that it is relatively easy to compare estimated regression models that have the same form of the dependent variable; but difficult to compare models with different forms of the dependent variable.

For our purposes, we will use the "AIC" -- Eviews calls this the "Akaike info criterion" -- to select the best model. (You will find it helpful to read the note below Exercise #8B in the Week #3 Assignment.)

Consider Models 1) and 2) from Exercise #5C above. Both have the same form of the dependent variable (CONSUMPTION). To determine which of the models fits best, compare the value of the "Akaike info criterion" for each of the two models. Choose the model with the lowest value.

The "Akaike info criterion" value for Model 1) is 21.82499; and 22.96205 for Model 2). Therefore, Model 1) is the better model.

5H. Answer Question #27.

<u>NOTE</u>: (Selecting the Best Model When the Dependent Variables Are Different.) By now you should have established that Model 1) is the best model when "CONSUMPTION" is the dependent variable, and Model 3) is best when "ln(CONSUMPTION)" is the dependent variable. But of these two, which is best? There is no easy statistical answer to this question. But we can use some common sense! Given the different coefficient interpretations (refer to the box in Exercise #5C) and what you know about people's consumption behaviour, which model makes the most sense to you?

Be prepared to answer the following questions in class:

1. (Provide a short answer in the space below.)
Your regression results from Exercise #1B should show an estimated coefficient for AGE equal to -1324.049. In plain English, provide a brief interpretation of this number.

2. (Provide a short answer in the space below.)
Your regression results from Exercise #1B should show an estimated coefficient for EDUCATION equal to -3119.208. In plain English, provide a brief interpretation of this number.

3. (Provide a short answer in the space below.)
Does anything seem strange to you about the estimated coefficients for AGE and EDUCATION? What? Write your thoughts in plain English in the space below.

4. (Supply the correct numerical answer on the line below.)
What is the average (mean) value of the variable AGE? Write your answer below:

____

5. (Supply the correct numerical answer on the line below.)
What is the average (mean) value of the variable EDUCATION? Write your answer below:

____

6. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #1B to predict the income of a person with an average age and an average amount of education. Write your answer below:
(<u>WARNING</u>: Let $\overline{X}_1$ be the average value of AGE, and let $\overline{X}_2$ be the average value of EDUCATION. To get the right value of AGExEDUCATION for a person with average age and an average amount of education, use $\overline{X}_1 \times \overline{X}_2$. Thus, if average age is 33.08021, and average education is 13.46845, use 33.08021×13.46845 (=445.53915).

____

7. (Supply the correct numerical answer on the line below.)
Repeat the analysis of the previous problem, only do it for a person who is one year younger.
(<u>HINT</u>: Every place you used average age to get your prediction in the previous case, now use average age minus 1. Thus, if average age is 33.08021, use 32.08021 instead. And for AGExEDUCATION, use $32.08021 \times 13.46845 = 432.0707$).

____

8. (Supply the correct numerical answer on the line below.)
Use your answers from the two previous questions to answer the following: For a person with an average age and an average amount of education, what is the effect on income of an extra year of age? Write your answer below:

____

9. (Provide a short answer in the space below.)
Compare your answer to Question #8 with that of Question #1. Why are they different? Provide an explanation in plain English.
(<u>HINT</u>: Does the fact that the coefficient on AGExEDUCATION is positive have something to do with it?)

10. (Multiple choice: Select the single best answer.)

According to the regression results of Exercise #1B, is the effect of an extra year of age on income larger (more positive/less negative) or smaller for people with a lot of education compared to those with only a little bit of education?

A) The effect of an extra year of age is more positive/less negative for well-educated people.

B) The effect of an extra year of age is less positive/more negative for well-educated people.

C) Education has no effect on the amount that age contributes to a person's income.

D) None of the above.

11. (Supply the correct numerical answer on the line below.)

Use a similar procedure to that above to answer the following question: For a person with an average age and an average amount of education, what is the effect on income of an extra year of education? Write your answer below:

____

12. (Multiple choice: Select the single best answer.)

According to the regression results of Exercise #1B, is the effect of an extra year of education on income larger (more positive/less negative) or smaller for older people compared to younger people?

(HINT: Does the fact that the coefficient on AGExEDUCATION is positive help you with this question?)

A) The effect of an extra year of education is more positive/less negative for older people.

B) The effect of an extra year of education is more positive/less negative for younger people.

C) Age has no effect on the amount that education contributes to a person's income.

D) None of the above.

13. (Fill in the required information in the "Hypothesis Test Table" below.) ***Consider the regression results obtained in Exercise #1B above. Use the standard hypothesis testing procedure to test whether the true value of the coefficient on AGE is zero (i.e., $B_{AGE} = 0$). Complete the "Hypothesis Test Table."***

1)	H_0: $B_{AGE} = 0$	(The true value of the coefficient on AGE is zero)
	H_A: $B_{AGE} \neq 0$	(The true value of the coefficient on AGE is not zero)
2)	LEVEL OF STATISTICAL SIGNIFICANCE =	
3)	STATISTICS:	
4)	DECISION =	

14. (Fill in the required information in the "Hypothesis Test Table" below.) ***Consider the regression results obtained in Exercise #1B above. A researcher is studying the following statement: A person's age has no effect on how much income they earn. Complete the corresponding "Hypothesis Test Table."***

1)	H_0:
	H_A:
2)	LEVEL OF STATISTICAL SIGNIFICANCE =
3)	STATISTICS:
4)	DECISION =

15. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #4B to predict average cost per car when 100 cars are produced per week. Write your answer below:

____

16. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #4B to predict average cost per car when 99 cars are produced per week. Write your answer below:

____

17. (Multiple choice: Select the single best answer.)
Use your answers from the two previous questions to answer the following: What is predicted effect on average cost per car when the car company increases production from 99 to 100 cars per week?

A) The average cost per car goes up.
B) The average cost per car goes down.
C) The average cost per car stays the same.
D) None of the above.

18. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #4B to predict average cost per car when 900 cars are produced per week. Write your answer below:

____

19. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #4B to predict average cost per car when 899 cars are produced per week. Write your answer below:

____

20. (Multiple choice: Select the single best answer.)
Use your answers from the two previous questions to answer the following: What is predicted effect on average cost per car when the car company increases production from 899 to 900 cars per week?

A) The average cost per car goes up.
B) The average cost per car goes down.
C) The average cost per car stays the same.
D) None of the above.

21. (Multiple choice: Select the single best answer.)
Compare your answers to Questions #20 and #17. Are they the same?

A) Yes.
B) No.

22. (Provide a short answer in the space below.)
Your regression results from Exercise #5C, Model 1) should show an estimated coefficient for INCOME equal to 0.641166. In plain English, provide a brief interpretation of this number.

23. (Provide a short answer in the space below.)
Your regression results from Exercise #5C, Model 2) should show an estimated coefficient for ln(INCOME) equal to 23433.86. In plain English, provide a brief interpretation of this number.

24. (Provide a short answer in the space below.)
Your regression results from Exercise #5C, Model 3) should show an estimated coefficient for INCOME equal to 1.44E-05 (=0.0000144). In plain English, provide a brief interpretation of this number.

25. (Provide a short answer in the space below.)
Your regression results from Exercise #5C, Model 4) should show an estimated coefficient for ln(INCOME) equal to 0.352034. In plain English, provide a brief interpretation of this number.

26. (Supply the correct numerical answer on the line below.)
Use the regression results from Exercise #5C, Model 2) to predict consumption for a person with an average amount of income and wealth. Write your answer below:

____

27. (Multiple choice: Select the single best answer.)
Compare the regression results for Models 3) and 4) from Exercise #5C. Which is the better model?

A) Model 3)
B) Model 4)
C) The two models are equally good.