MSCI:9110 *Advanced Analytics*
Kurt Anstreicher
Spring, 2017

SAMPLE FINAL EXAM

Name: SOLUTION

This is a two and one-half hour, open book and notes exam. There are a total of 85 points. Be sure to show your work to receive partial credit. Good Luck!

Question 1. (40 points) A regression analysis is being used to compare a production index for the G7 countries and the worldwide consumption of copper, in thousands of tons, over a 40-year period. The regression output and residual plots are shown below.

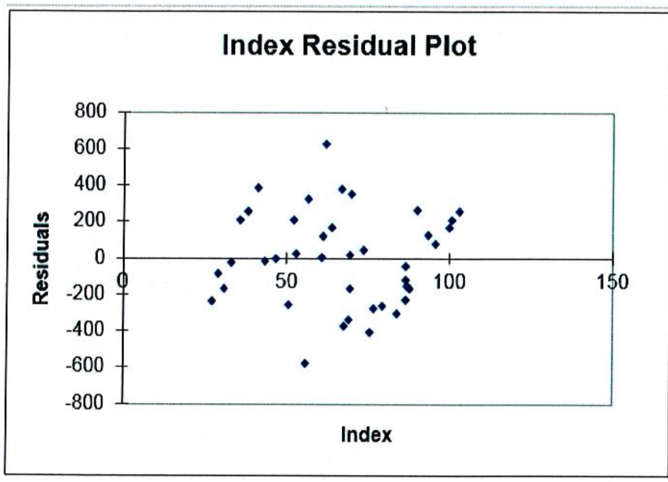
Regression of Copper on Index

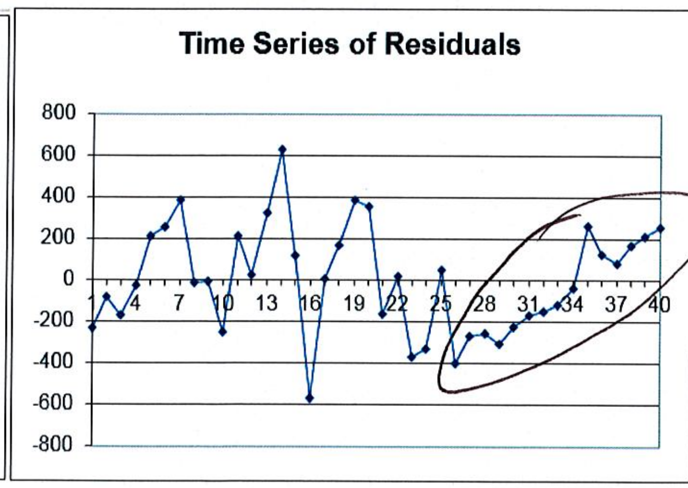
Regression Statistics	
Multiple R	0.969
R Square	0.939
Adjusted R Square	0.938
Standard Error	263.126
Observations	40

ANOVA

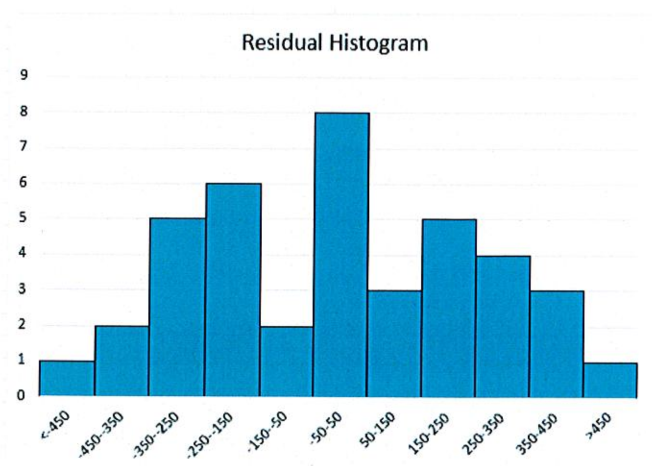
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	40840277.821	40840277.821	589.876	0.000
Residual	38	2630943.315	69235.350		
Total	39	43471221.136			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2081.001	136.411	15.255	0.000	1804.852	2357.150
Index	47.331	1.949	24.287	0.000	43.386	51.276





Autocorrelation= .439



A. (5 points) Describe problems, if any, that you see with the residuals from this regression.

THE RESIDUAL AUTOCORRELATION IS HIGHER THAN DESIRABLE. THERE IS ALSO A CLEAR INCREASING TREND IN THE RESIDUALS FOR THE LAST 15 YEARS OF DATA.

B. (10 points) Suppose that this regression is going to be used to generate a forecast for copper consumption based on a forecast of 105 for next year's G7 production index. Using the above regression, what is the forecast value for copper consumption? Approximately how large is a 95% prediction interval for this forecast? Do you have any reason to believe that this forecast might tend to be high, or low?

FORECAST IS

$$\hat{y} = b_0 + b_1 X = 2081 + 47.331(105)$$
$$= \underline{7050.755}$$

THE ERROR FOR A 95% PREDICTION INTERVAL IS APPROXIMATELY $\pm 2 S_e = 2(263.126) = \underline{\pm 526.25}$

ONE MIGHT SUSPECT THAT THIS FORECAST IS **LOW** BASED ON THE TIME SERIES OF RESIDUALS, SINCE POSITIVE RESIDUALS CORRESPOND TO PREDICTED VALUES BEING **LESS THAN** ACTUAL VALUES.

An alternative regression uses the variables Diff_Copper and Diff_Index, where "Diff" for a time series X_t is equal to $X_t - X_{t-1}$
The regression of Diff_Copper on Diff_Index, and residual plots, is given below.

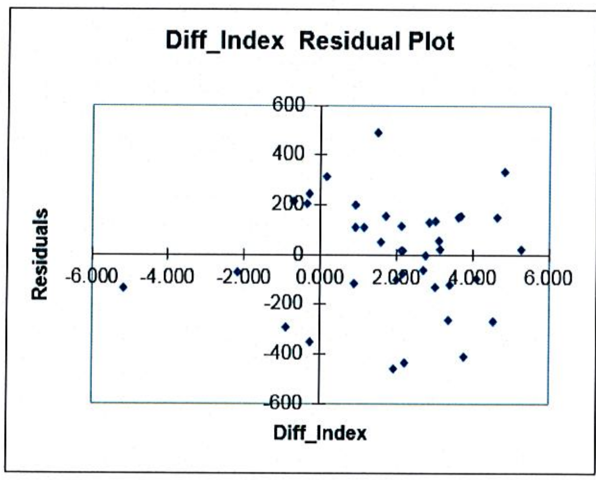
Regression of Diff_Copper on Diff_Index

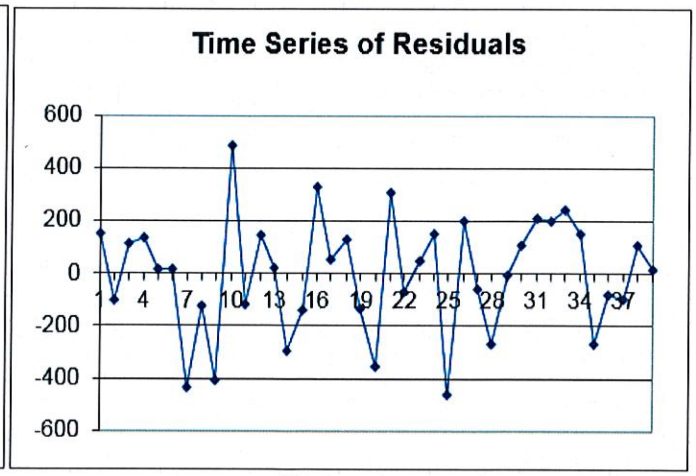
Regression Statistics	
Multiple R	0.770
R Square	0.593
Adjusted R Square	0.582
Standard Error	222.279
Observations	39

ANOVA

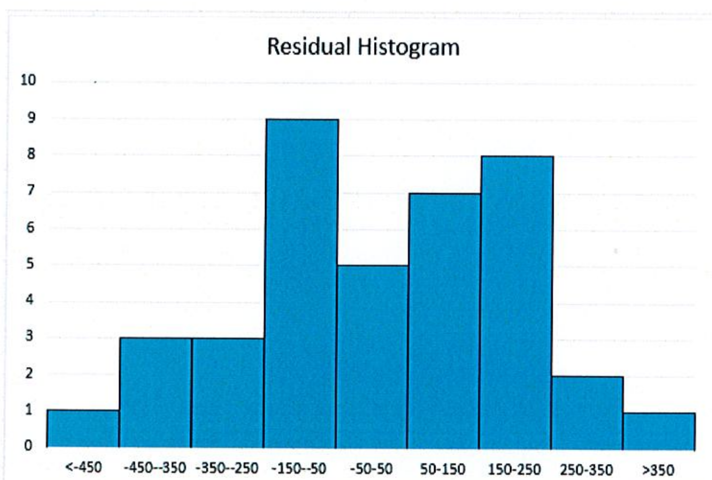
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2665282.605	2665282.605	53.944	0.000
Residual	37	1828098.182	49408.059		
Total	38	4493380.787			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-143.423	49.060	-2.923	0.006	-242.829	-44.018
Diff_Index	127.534	17.364	7.345	0.000	92.351	162.717





Autocorrelation = -.162



C. (5 points) Why does the second regression have only 39 observations, when there are 40 years of data?

THE FIRST OBSERVATION IS LOST DUE TO THE USE OF THE "DIFF" OPERATION (THERE IS NO X_0 OR Y_0)

D. (5 points) What advantages, if any, do you see in the residuals from the second regression compared to the first regression?

THE AUTOCORRELATION IS SUBSTANTIALLY SMALLER AND THE TREND AT THE END OF THE RESIDUAL TIME SERIES IS GONE.

E. (5 points) Do you agree or disagree with the statement "The first regression fits the data better because the R-Square value is much higher?" Explain your answer.

DISAGREE. THE STD. ERROR IS LOWER IN THE SECOND REGRESSION THAN IN THE FIRST.

THE SYSTEMATIC ERRORS AT THE END OF THE TIME SERIES IN THE FIRST REGRESSION ARE ALSO ELIMINATED IN THE SECOND.

F. (10 points) Suppose that the final value for the G7 index in the data is 103, and the final value for copper consumption is 7,212. Use the second regression to give a forecast for copper consumption based on a forecast of 105 for next year's G7 production index.

NOTE THAT IF $X_{41} = 105$ AND $X_{40} = 103$, THEN $DIFF\_X_{41} = 2$. FROM THE REGRESSION,

$$\hat{y} = a + bX = -143.423 + 2(127.534)$$
$$= 111.645.$$

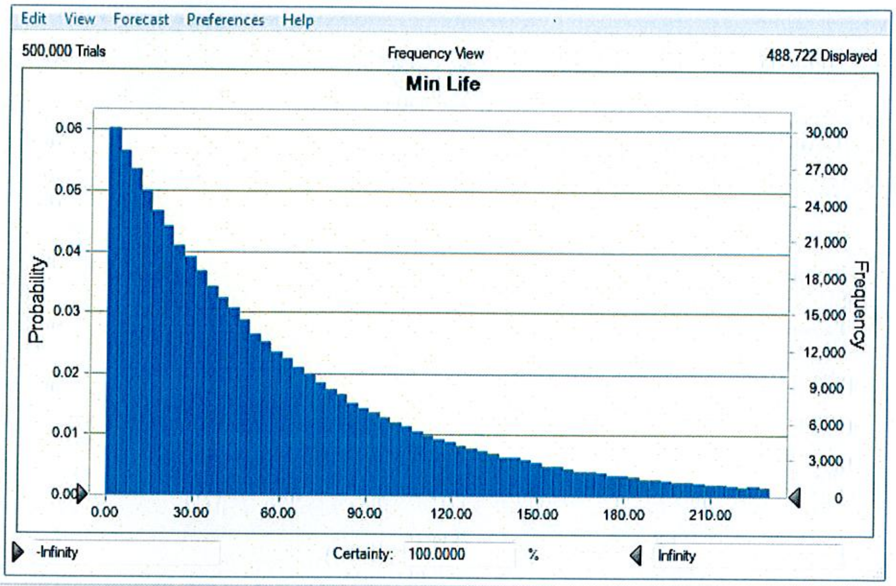
SINCE THE DEPENDENT VARIABLE IS DIFF_COPPER, THE FORECAST FOR COPPER IS

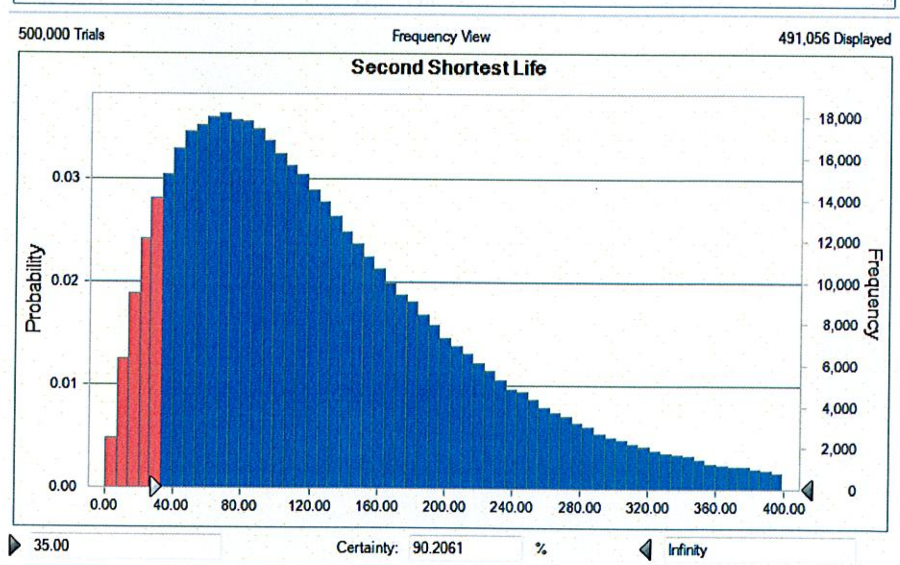
$$7212 + 111.645 = \underline{7323.65}$$

(NOTE THAT THIS IS HIGHER THAN THE FORECAST FROM THE FIRST REGRESSION)

Question 2. (15 points) An industrial sterilizer uses 6 high-power UV bulbs. The sterilizer requires that at least 5 bulbs be working to perform adequately. The operating life of an individual UV bulb has an exponential distribution with mean 365 days. A Crystal Ball simulation has been written to determine how often the sterilizer will need to be shut down to replace bulbs. The lives of the individual bulbs are CB Assumption cells, which are then sorted in the spreadsheet. The two CB Forecast cells are the minimum life, and the second shortest life. The distributions for these two forecast cells based on a run with 500,000 trials are shown below. The mean values for the two forecast cells are also shown in the spreadsheet.

Lamp #	Life	Sort	Mean
1	365	297	60.8
2	400	350	133.7
3	405	365	
4	297	400	
5	350	405	
6	500	500	





A. (5 points) Based on the simulation output, what is the distribution of the minimum of 6 random variables, each of which has an exponential distribution with mean 365 days?

THIS APPEARS TO BE AN EXPONENTIAL DISTRIBUTION WITH MEAN $= \frac{365}{6} = \underline{\underline{60.83}}$ DAYS

B. (5 points) Based on the simulation output, if the sterilizer is started up with 6 new bulbs, what is the probability that it will need to be shut down to have bulbs changed after 35 days or less of operation?

$\underline{\underline{9.8\%}}$, OR A BIT LESS THAN 10%.

C. (5 points) Suppose that 2 bulbs have burned out and the sterilizer is shut down to have bulbs changed. If the individual bulbs actually have lives with an exponential distribution, do the 4 bulbs that are still working need to be replaced? Explain.

<u>NO</u>. FOR AN EXPONENTIAL DISTRIBUTION, THE REMAINING LIFE AFTER ANY GIVEN TIME IS <u>EQUAL</u> TO THE ORIGINAL LIFE DISTRIBUTION. THIS MEANS THAT THE BULBS THAT ARE STILL WORKING ARE "AS GOOD AS NEW."

Question 3. (30 points) A producer of agricultural machinery must determine the production schedule for a particular tractor for the 2017 calendar year. Currently there are 100 completed tractors in inventory. Based on historical data, and market research, there are projected demands for tractors in each month of 2017. There are also production capacities for each month. These capacities are 150 units/month for the months of January-April, October, and November, 100 units/month for May-September, and 50 units for December. (Capacity is lower in the summer months because of required production of other items, and is lower in December due to holidays and scheduled maintenance.) A linear program is formulated to minimize the sum of tractors in inventory at the start of each month, from January 2017 to January 2017, subject to meeting the projected demands, not exceeding monthly production capacities, and maintaining non-negativity of inventories and production levels. A spreadsheet with the problem data, and solution, is shown below. Numbers in boldface are data, corresponding to the initial inventory, monthly demands, and monthly production capacities. The complete Excel Solver Sensitivity Report is on the following page.

	A	B	C	D	E
1		Inventory	Demand	Production	Capacity
2	January	**100**	**0**	0	**150**
3	February	100	**0**	0	**150**
4	March	$\boxed{100}$	$\boxed{\mathbf{20}}$	120	**150**
5	April	$\boxed{200}$	**50**	150	**150**
6	May	300	**100**	100	**100**
7	June	300	**200**	100	**100**
8	July	200	**200**	100	**100**
9	August	100	**200**	100	**100**
10	September	0	**100**	100	**100**
11	October	0	**20**	20	**150**
12	November	0	**10**	10	**150**
13	December	0	**0**	0	**50**
14	January	0			
15	Total	1400			

A. (10 points) What formula is contained in cell D4? (Hint: Note that in the LP solved by Solver the variables, or adjustable cells, are the starting inventories for February 2017-January 2017)

FOR EACH MONTH WE HAVE

$I_{i+1} = I_i - D_i + P_i$, WHICH MEANS THAT

$P_i = D_i + I_{i+1} - I_i$.

THE FORMULA IN CELL D4 IS THEN

$= \underline{C4 + B5 - B4}$.

(TO CHECK, WE SHOULD THEN HAVE

120 = 20 + 200 − 100 ✓)

Microsoft Excel 8.0a Sensitivity Report

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B3	February Inventory	100	0	1	1E+30	2
B4	March Inventory	100	0	1	1E+30	1
B5	April Inventory	200	0	1	1E+30	1
B6	May Inventory	300	0	1	1E+30	2
B7	June Inventory	300	0	1	1E+30	3
B8	July Inventory	200	0	1	1E+30	4
B9	August Inventory	100	0	1	1E+30	5
B10	September Inventory	0	6	1	1E+30	6
B11	October Inventory	0	1	1	1E+30	1
B12	November Inventory	0	1	1	1E+30	1
B13	December Inventory	0	2	1	1E+30	2
B14	January Inventory	0	1	1	1E+30	1

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
D2	January Production	0	0	150	1E+30	150
D3	February Production	0	0	150	1E+30	150
D4	March Production	120	0	150	1E+30	30
D5	April Production	150	-1	150	120	30
D6	May Production	100	-2	100	120	30
D7	June Production	100	-3	100	120	30
D8	July Production	100	-4	100	120	30
D9	August Production	100	-5	100	100	30
D10	September Production	100	0	100	1E+30	0
D11	October Production	20	0	150	1E+30	130
D12	November Production	10	0	150	1E+30	140
D13	December Production	0	0	50	1E+30	50
D2	January Production	0	2	0	120	30
D3	February Production	0	1	0	120	30
D4	March Production	120	0	0	120	1E+30
D5	April Production	150	0	0	150	1E+30
D6	May Production	100	0	0	100	1E+30
D7	June Production	100	0	0	100	1E+30
D8	July Production	100	0	0	100	1E+30
D9	August Production	100	0	0	100	1E+30
D10	September Production	100	0	0	100	1E+30
D11	October Production	20	0	0	20	1E+30
D12	November Production	10	0	0	10	1E+30
D13	December Production	0	1	0	50	0

B. (15 points) Suppose that it is possible to increase production capacity in the months of June, July, and August by 30 units/month, by hiring summer workers. The total cost of hiring the summer workers would be $60,000. The cost of carrying one tractor in inventory for one month (including the cost of capital, insurance, storage cost, etc.) is $250. Is the option of hiring the summer workers attractive? Explain. (Note: Assume that each tractor in inventory at the start of each month incurs the $250 carrying charge for that month.)

THIS IS A CHANGE OF MULTIPLE RHS VALUES, SO WE NEED TO CHECK THE 100% RULE. HERE

$$\sum_{\text{CONSTRAINTS}} \frac{\text{CHANGE}}{\text{RANGE}} = \frac{30}{120} + \frac{30}{120} + \frac{30}{100} = .8 < 1 \checkmark$$

THE SHADOW PRICES THEREFORE APPLY, SO THE CHANGE IN THE OBJECTIVE WILL BE

$$-3(30) - 4(30) - 5(30) = -12(30) = \underline{\underline{-360}}$$

THE ASSOCIATED INVENTORY COST SAVINGS WILL THEN BE

$$360(250) = \underline{\underline{\$90,000}}$$

SINCE THE COST SAVINGS IS SUBSTANTIALLY HIGHER THAN THE COST OF HIRING THE WORKERS, THIS OPTION IS ATTRACTIVE.

C. (5 points) Give brief explanations of why the reduced costs for September and December inventories are 6 and 2, respectively. (Recall that the reduced cost is the shadow price for the non-negativity constraint.)

SEPTEMBER

TO HAVE A UNIT IN INVENTORY AT THE START OF SEPTEMBER, IT WILL HAVE TO BE PRODUCED IN MARCH. (NOTE THAT PRODUCTION IS AT CAPACITY FOR APRIL – AUGUST.) THIS WILL ADD ONE UNIT TO THE STARTING INVENTORY FOR APRIL, MAY, JUNE, JULY, AUGUST AND SEPTEMBER – A TOTAL OF 6 MONTHS.

DECEMBER

THERE IS NO DEMAND IN DECEMBER, SO A UNIT IN INVENTORY AT THE START OF DECEMBER WILL ALSO BE IN INVENTORY AT THE START OF JANUARY, FOR A TOTAL OF 2 MONTHS.