INTERDISCIPLINARY LIVELY APPLICATIONS PROJECT

Lake Pollution

6 PARTS

Interdisciplinary Lively Applications Project

I

Title: Lake Pollution

Authors: William Fox
William Goetz
Hermann Kolev
Ronald Miller
Jeffery Strickland
Wendell King

Managing Editor: David C. Arney
Production Editor: Pauline Wright
Editors: Donald Small
Wade Ellis
Kathleen Snook

Department of Mathematical Sciences, Department of Geography and Environmental Engineering, and Department of Chemistry, United States Military Academy, West Point, NY

Date: January 1995

Carry-Through Subjects: Discrete Mathematics, Differential Equations, Probability, Statistics

Mathematics Classifications: Difference Equations, Differential Equations, Discrete Mathematics, Probability, Statistics, Linear Regression, Linear Algebra

Disciplinary Classifications: Environmental Engineering

Prerequisite Skills:

1 .Modeling with difference equations
2. Solving first- and second-order difference equations
3. Solving linear systems of equations
4. Solving linear systems of difference equations
5. Modeling with differential equations
6. Solving first-order differential equations using numerical and analytic methods
7. Modeling with stochastic processes (Markov chains)
8. Using Descriptive statistics, classical probability, Bayes' theorem, and probability distributions
9. Solving linear regression problems

Materials Available:

A. Problem Statement (6 Parts) (Student)
B. Sample Solution (Instructor)
C. Background Material (Instructor)

Interdisciplinary Lively Applications Project is funded by the National Science Foundation, Directorate of Education and Human Resources Division of Undergraduate Education, NSF Grant #9455980.

NSF Initiative: Mathematics Sciences and their Applications Throughout the Curriculum (CCD-MATH)

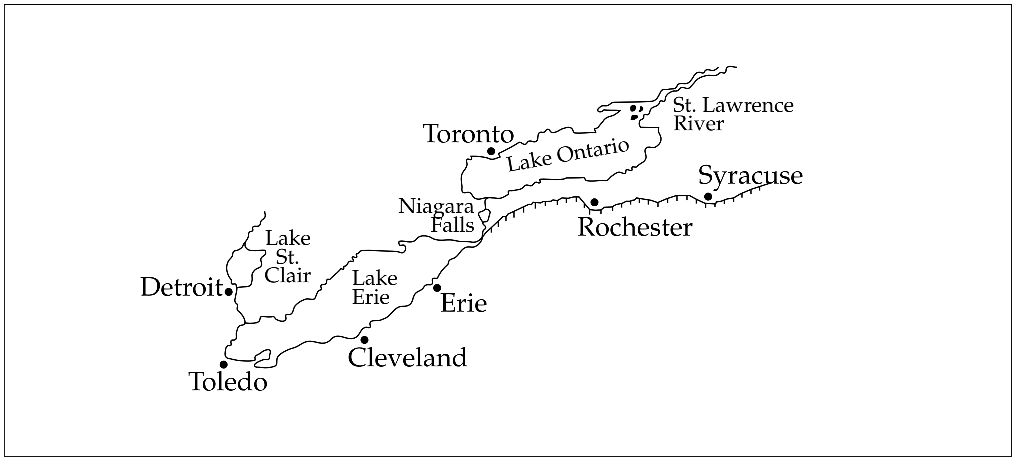
PART 1

PREREQUISITE SKILLS:

1. MODELING WITH DIFFERENCE EQUATIONS
2. SOLVING FIRST- AND SECOND-ORDER DIFFERENCE EQUATIONS

Most of the water flowing into Lake Ontario is from Lake Erie. Assume that both lakes contain some amount of a certain pollutant. Suppose that this kind of pollution of the lakes has ceased, except for pollution from a factory on Lake Ontario. How long would it take for the pollution level in each lake to be reduced to 10 percent of its present level?

First, to simplify matters, let's assume that 100 percent of the water in Lake Ontario comes from Lake Erie. Let $a(n)$ and $b(n)$ be the total amount of pollution in Lake Erie and Lake Ontario, respectively, after n years. Since pollution has stopped, the concentration of pollution in the water coming into Lake Erie is 0. It has also been determined that, each year, the percentage of water replaced in Lakes Erie and Ontario is approximately 38 and 13 percent, respectively. Additionally, suppose that the one remaining factory on Lake Ontario directly dumps 25 units of this pollutant into the lake each year. Initially, data indicates that there were 2500 units of this pollutant in Lake Ontario last year, and there are 3150 units of this pollutant in the lake this year.



REQUIREMENT 1.

Write a system of difference equations which models this pollution process, and convert the first-order system to a second order equation that will model the amount of this pollutant in Lake Ontario.

REQUIREMENT 2.

Find the general solution to this equation.

REQUIREMENT 3.

Does this equation have an equilibrium value? Justify your answer. Interpret the meaning of an equilibrium value in this situation.

REQUIREMENT 4.

Find the particular solution and determine how long it would take for the pollution level in Lake Ontario to be reduced to 10 percent of its present level.

REQUIREMENT 5.

Describe the long term behavior of this model.

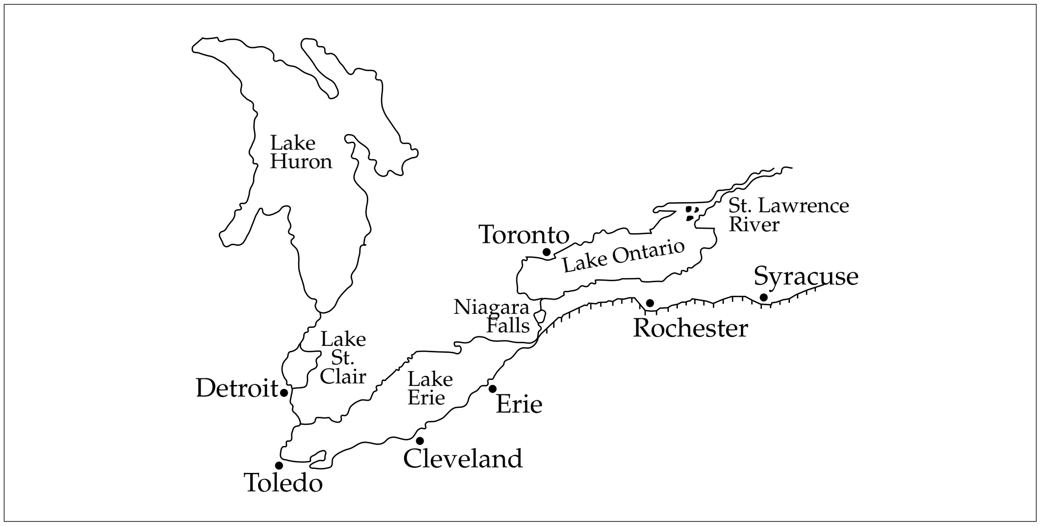
PART 2

Prerequisite Skills:

3. Solving linear systems of equations
4. Solving linear systems of difference equations

Most of the water flowing into Lake Erie is from Lake Huron, and most of the water flowing into Lake Ontario is from Lake Erie. Assume that the three lakes contain some amount of a certain pollutant. Suppose that this kind of pollution of the lakes has ceased, except for pollution introduced by two factories, one each on Lake Huron and Lake Ontario. How long would it take for the pollution level in each lake to be reduced to 10 percent of its present level?

To simplify matters, let's assume that 100 percent of the water in Lake Erie comes from Lake Huron and 100 percent of the water in Lake Ontario comes from Lake Erie. Let $a(n)$, $b(n)$ and $c(n)$ be the total amount of pollution in Lake Huron, Lake Erie and Lake Ontario, respectively, after n years. It has also been determined that, each year, the percentage of water replaced in Lakes Huron, Erie and Ontario is approximately 11, 36 and 12 percent, respectively. Additionally, suppose that the two factories on Lake Huron and Lake Ontario directly dump 30 units of this pollutant into each lake each year. Initially, there are 3500, 1800 and 2400 units of this pollutant in Lakes Huron, Erie and Ontario, respectively.



REQUIREMENT 1.

Write a system of difference equations which models this process.

REQUIREMENT 2.

Find the general solution to this system.

REQUIREMENT 3.

Does this system have an equilibrium vector? Justify your answer.

REQUIREMENT 4.

Find the particular solution and determine how long it would take for the pollution level in each lake to be reduced to 10 percent of its present level.

REQUIREMENT 5.

Describe the long term behavior of this model.

PART 3

PREREQUISITE SKILLS:

5. MODELING WITH DIFFERENTIAL EQUATIONS
6. SOLVING FIRST-ORDER DIFFERENTIAL EQUATIONS USING NUMERICAL AND ANALYTIC METHODS

Modeling pollution in a river-bay system can be thought of as a mixing problem where the rate of a quantity Q of a pollutant is equal to the difference between the rate of input(s) and the rate of output(s):

$$\frac{dQ}{dt} = R_{inputs} - R_{outputs}.$$

Our particular problem concerns pollution in a small stream and bay connected to the Great Lakes. The primary source of pollution is located upstream on the river. Additionally, there is an aluminum factory located directly on the bay that contributes to the problem. Environmentalists are very concerned about the level of pollutants in the bay and have lobbied for laws to protect the bay. The present law forces the aluminum factory to temporarily shut down anytime the average concentration of pollutants reaches 1.6 milligrams per liter of bay water.

To model the change over time of the concentration of pollutants in the bay, we will consider it to be of a constant volume V containing at time t an amount of pollutant $Q(t)$. Assume the pollutant is evenly distributed throughout the bay with a concentration $C(t)$ where $C(t) = \frac{Q(t)}{V}$. Also, assume that water from the river containing a constant concentration k of pollutant enters the bay at a rate r and that the polluted water in the bay flows out at the same rate. From this information, we determine that $R_{outputs}=rk$ and $R_{outputs}=rC$.

Requirement 1.

Write a differential equation that models the concentration of pollutants in the bay. Include a boundary condition if the bay begins with N milligrams per liter of pollutants. **Explain** your model.

Problem Continuation:

We will assume that the approximate volume of the bay is four million liters. We further assume that the water flows into the bay at a rate of 40,000 liters per day, and that the water flows out of the bay at the same rate. The amount of pollutants $Q(t)$ will be measured in milligrams and the concentration $C(t)$ will therefore have units of milligrams per liter. Time will be measured in days. The bay begins with 0.8 milligrams of pollutants per liter of water. The stream normally has a pollution concentration of $k = 0.5$ milligrams per liter.

Requirement 2.

A. Using a numerical method, find the concentration of the pollutants in the bay for the first thirty days.

B. Solve the differential equation by the method of Separation of Variables to obtain the concentration of pollutants in the bay for the same data as in part (**A.**) of this requirement.

C. Compare the answers to the two methods. Which method is more accurate? Why?

D. Will the pollution concentration achieve a steady state value? If so, what is it?

Requirement 3.

Four million milligrams of pollution are instantaneously spilled into the stream. Assume that all of the pollution in the spill reaches the bay, that it arrives as one big input (or impulse), and that it is instantly mixed throughout the lake upon arrival. [Reminder: The bay has a pollution concentration of 0.8 mg/1 already.] For how long will the factory have to shut down? Assume that after the spill reaches the bay the amount of pollution in the stream returns to normal.

Requirement 4.

List and discuss at least four factors not taken into account by this model that would have an impact on the pollution level in the bay.

PART 4

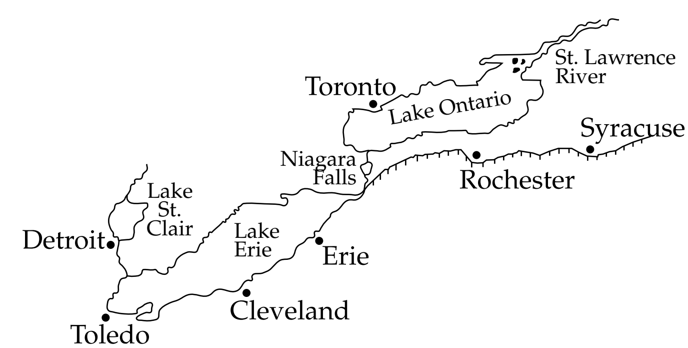
PREREQUISITE SKILLS:

7. MODELING WITH STOCHASTIC PROCESSES (MARKOV CHAINS)

The pollution of the Great Lakes has, in the past few decades, been a significant issue on the agenda of both local and regional environmentalist groups. There have been a number of clean-up initiatives that have been started in the past 20 years which have greatly increased the outlook for the future of this natural waterway. The primary polluters have been some of the heavy industries that use the lakes as a means of transportation to get their products to both domestic and international markets. Some of the most significant violators are the steel industries around the Pittsburgh area on Lake Erie and around the Gary, Indiana area of Lake Michigan, and the mining industries around the Superior/Duluth areas of Lake Superior. Both of these industries have made significant strides in recent years to clean up their operations and their efforts are starting to slowly pay off.

SITUATION

You have been hired as an analyst by a regional aluminum processing plant that is located on the shores of Lake Erie. They are in the process of analyzing various clean-up techniques that they have been using over the past few years and they want to know which technique is the most cost effective from both an economic and environmental point of view. You are assigned to do a study of a two lake system which includes Lake Erie and Lake Ontario. Weekly test samples have been taken over the past few years and the amount of aluminum related pollutants in each sample has been recorded. After taking a look at all of the data, you have been able to come up with some preliminary results. You have determined that if there are traces of aluminum pollutants in this week's test of Lake Erie, then 35% of the time there will be traces of aluminum pollutants in next week's test of Lake Erie and 65% of the time there will be traces of aluminum pollutants in next week's test of Lake Ontario. If there are traces of aluminum pollutants in this week's test of Lake Ontario, then 90% of the time there will be traces of aluminum pollutants in next week's test of Lake Ontario and 10% of the time there will be traces of aluminum pollutants in next week's test of Lake Erie. You are making the assumption that the water flows in both directions even though you know that the primary direction of flow goes from Lake Erie to Lake Ontario. You are also assuming that the traces of aluminum pollutants are confined to Lake Ontario and Lake Erie. You are ultimately interested in the long term probabilities of finding aluminum pollutants in Lake Ontario and Lake Erie.



Requirement 1.

Draw a classical probability tree diagram which correctly portrays the probability of finding traces of aluminum in next week's tests of Lake Erie and Lake Ontario based on the probability of finding traces of aluminum in this week's tests of the two lakes. Ensure that the branches of your tree diagram are properly labeled. Also ensure that you include any probabilities that you may know at this time. If there are probabilities that you do not currently know, put "NA" on the branch where you would normally put the numerical value of the probability.

REQUIREMENT 2.

We have made some assumptions in our analysis. List two (2) other plausible assumptions that you made in order to construct the tree diagram in requirement **1** above.

REQUIREMENT 3.

You now have to calculate the unknown probabilities on your tree diagram. Construct a model for the probability of finding traces of aluminum pollutants in Lake Ontario and Lake Erie from one week to the next. Clearly define any variables that you use in your model. Use the model to solve for the unknown probabilities on your tree diagram. HINT: Solve for the following probabilities: P(aluminum pollutants in Lake Ontario) and P(aluminum pollutants in Lake Erie) during this week's test.

REQUIREMENT 4.

You should now be able to fill in the unknown probabilities on your tree diagram. Use the tree diagram to answer the following questions.

A. What is the probability that traces of aluminum pollutants will be found in either Lake Erie or Lake Ontario in two consecutive weekly tests?

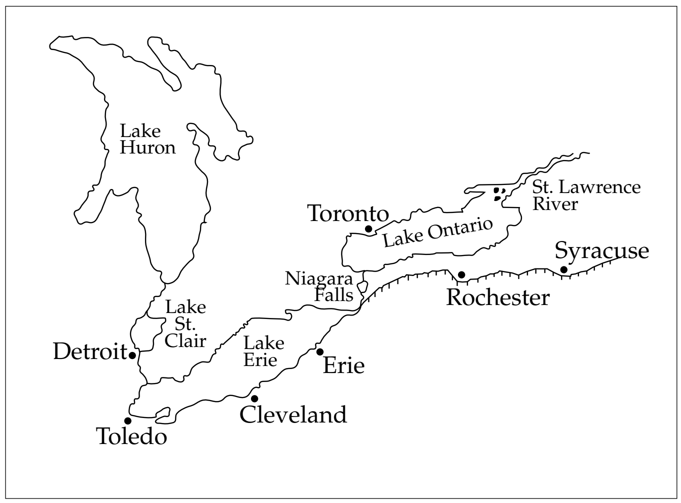
B. What is the probability of finding traces of aluminum pollutants in last week's test of Lake Erie given that there were traces of aluminum pollutants found in this week's test of Lake Erie?

C. In the long term, which lake has the higher probability of having traces of aluminum pollutants in it? What is the numerical value of this long run probability?

D. Based upon your answers above and your knowledge of the primary direction of flow of the Great Lakes, does your analysis make sense? Explain.

Situation(continued):

Your boss is impressed with your work so far and is interested in seeing you do some more in-depth analysis of the pollution situation. He thinks that constraining your analysis to the two lake system of Lake Erie and Lake Ontario is a bit too simplistic. He would like you to add Lake Huron into your analysis and see if that has any effect on your results. The company has been making weekly tests of Lake Huron for the past several years so it should not be that difficult to incorporate Lake Huron's data into your analysis. You take a look at all of the data and you are able to come up with the following table of probabilities:



	Probability that traces of aluminum will be found in next week's test of Lake_______		
Traces of aluminum were found in this week's test of Lake _________	**Huron**	**Erie**	**Ontario**
Huron	0.25	0.75	_____
Erie	0.10	0.25	0.65
Huron	____	0.15	0.85

REQUIREMENT 5.

You notice that there are two blank spaces in your probability table. What values should be put in those spaces? Explain in words why you assigned the probabilities that you did.

REQUIREMENT 6.

Draw and properly label a new classical probability tree diagram which includes all three lakes. Construct the equation that models this new situation. Clearly define any variables that you use in your model. Use your model to calculate any unknown probabilities that are on your tree diagram.

REQUIREMENT 7.

Use your completed tree diagram to answer the following questions:

A. What is the probability that traces of aluminum will be found in either Lake Huron or Lake Erie or Lake Ontario in two consecutive weekly tests? How does this compare to your answer in requirement **4(A.)**?

B. What is the probability that traces of aluminum were found in last week's test of Lake Huron given that there were traces of aluminum found in this week's test of Lake Erie?

C. What is the probability that there are traces of aluminum found in this week's test of Lake Ontario and next week's test of Lake Erie?

PART 5

PREREQUISITE SKILLS:

8. USING DESCRIPTIVE STATISTICS, CLASSICAL PROBABILITY, BAYES' THEOREM, AND PROBABILITY DISTRIBUTIONS

TWO years have passed since you completed your analysis of the Great Lakes pollution problem. The major polluters along the chain of lakes put together a fairly comprehensive clean-up plan which temporarily calmed the fears of the concerned environmentalist groups. Recently, a study was published in an environmentally friendly journal which implied that the clean-up plans that had been implemented by the major polluters were not effective. The journal article basically claimed that the clean-up plans were mere eyewash to quiet the environmental watchdog groups. This new revelation has energized all of the key players involved in the Great Lakes pollution issue and each side is currently assessing their ability to defend their position in a court of law.

SITUATION:

You are still working as an analyst for the regional aluminum processing plant that is located on Lake Erie. You have heard rumors that the study of the Great Lakes pollution problem has been reopened and that the CEO of your firm is relying on your "institutional knowledge" to head up a team of analysts which will finally put this issue to rest. Your boss walks into your office and confirms that the rumors are true. He provides you with two years worth of data and says that he wants some quality statistical analysis which will withstand any scrutiny by the environmentalists. The data consist of weekly test samples that were taken from Lake Erie and analyzed for various attributes.

Sample Data

ROW	alum_ppm	oth_ppm	puretime	epacode	temp
1	36.2846	59.7390	12.105	0	59.7710
2	35.1909	43.6978	37.864	0	64.9707
3	38.5753	45.4508	67.275	0	63.1579
4	29.8268	55.6974	45.286	0	61.6976
5	32.7359	43.6578	81.583	1	63.7523
6	33.4072	56.6556	6.297	1	71.4984
7	31.5494	49.9503	32.174	0	65.7153
8	29.4830	59.2501	15.345	0	68.2051
9	28.3125	59.2778	103.952	0	67.0155
10	33.3508	30.8085	44.684	0	63.7843
11	35.4065	50.9860	16.497	0	60.9638
12	40.8595	59.0104	28.704	0	60.3454
13	38.0167	57.1471	67.667	0	67.6119
14	48.5467	48.4696	36.098	0	69.8011
15	38.2178	47.4461	17.372	0	69.9018
16	35.7263	37.0215	33.567	0	58.6964
17	44.3164	52.0141	23.307	0	62.4671
18	33.7658	53.8839	109.182	0	59.0033
19	38.6447	35.8155	1.579	1	64.0707
20	44.4616	43.8406	2.717	0	65.7709
21	37.8688	58.4949	72.884	1	64.6611
22	33.2331	56.7300	131.370	1	61.7126
23	39.3477	45.6032	57.226	0	63.9489
24	34.2471	37.7449	58.106	1	71.5458
25	38.9567	45.7960	16.483	0	72.5465
26	37.2145	44.0806	11.400	1	80.2774
27	39.3465	55.0693	63.580	1	60.4900
28	34.6710	31.5343	8.304	0	67.0364
29	26.6301	57.3854	18.576	1	66.2645
30	30.8516	30.2706	91.128	0	71.2584
31	30.8723	42.1799	5.037	0	63.9580
32	29.7076	44.9115	17.931	0	66.5546
33	42.7976	40.9325	38.233	1	63.0884
34	44.4128	57.4171	52.001	0	70.0097
35	28.4938	30.2881	12.810	1	70.9777
36	32.8682	46.1233	30.378	0	66.3875
37	35.1048	39.4139	17.783	0	67.7514
38	40.9571	46.7199	0.608	1	63.1022
39	28.7226	36.3994	27.310	0	66.6784
40	32.7272	42.1318	65.823	0	66.2453
41	30.2132	54.0219	18.391	1	70.7118
42	29.0897	34.4747	56.934	0	67.7272
43	34.0419	49.2455	168.731	0	68.2565
44	28.1496	43.0440	29.275	0	64.8418
45	35.2347	58.2225	5.884	0	67.0687
46	37.2364	57.4492	4.351	1	66.6186
47	29.1620	53.2136	42.727	1	76.5815
48	40.4460	30.9330	32.152	1	68.9791
49	35.7201	48.8363	11.040	0	66.2452
50	29.2285	32.6732	137.548	0	71.2823
51	42.9233	53.6791	40.148	0	73.7126
52	34.9017	31.7362	69.898	1	74.0190

Sample Data

ROW	alum_ppm	oth_ppm	puretime	epacode	temp
53	26.6193	35.7099	21.848	1	70.0692
54	32.8131	39.6966	19.566	0	68.8021
55	30.6090	48.3676	25.052	1	71.9412
56	28.5616	42.1876	1.559	1	62.3037
57	30.4919	53.2427	19.179	0	64.9527
58	39.5930	47.2459	44.036	1	66.1605
59	32.4821	31.4635	46.615	1	63.9796
60	35.1442	43.3965	22.226	1	63.8191
61	33.9530	44.8305	18.446	1	61.8358
62	29.4919	46.8731	150.477	0	67.2914
63	28.4642	39.5781	15.591	1	68.2996
64	26.1369	37.0135	14.602	1	72.8905
65	35.1675	37.8624	4.265	1	70.5572
66	37.4473	37.1184	40.328	1	80.3929
67	37.6916	36.5494	33.648	1	70.1867
68	24.6438	41.2847	2.789	0	68.3591
69	30.0035	48.9843	91.200	1	75.7433
70	26.0824	50.6743	146.261	1	66.4860
71	31.5738	48.5370	33.855	0	70.8557
72	33.6830	33.1112	43.153	1	76.2052
73	30.7659	54.3590	26.966	1	70.2122
74	28.5338	56.7471	192.736	1	66.7332
75	29.9502	34.9079	58.938	1	72.0411
76	39.0303	43.4007	7.306	0	66.5119
77	40.7136	34.6556	29.328	1	71.3661
78	47.5382	42.0995	42.509	1	69.1802
79	27.8366	55.4403	15.455	1	71.9098
80	34.5073	42.9701	13.940	0	67.6962
81	33.2366	56.2188	18.492	1	60.7372
82	38.1850	39.4222	15.752	0	64.3912
83	31.3074	33.5046	6.902	0	63.3019
84	32.5488	38.6973	22.425	0	62.7968
85	30.3595	59.6348	25.319	1	74.7648
86	36.6709	38.3582	75.050	1	75.9615
87	37.7908	46.5821	3.351	0	62.3881
88	33.5284	36.8290	9.954	0	66.9608
89	33.9880	35.3832	28.603	0	67.8019
90	30.6810	56.9938	41.422	1	73.3128
91	33.2926	33.0390	9.068	0	62.4025
92	34.0935	37.8167	6.904	0	65.2770
93	37.5900	51.4286	13.276	1	64.4110
94	35.6707	37.3356	0.195	0	62.9142
95	35.0608	30.3771	23.171	1	67.5920
96	30.6652	37.7568	40.732	0	62.4702
97	33.3158	43.9597	78.473	1	67.3556
98	32.9712	50.1189	2.845	1	70.5495
99	43.8148	47.6849	28.864	1	63.2839
100	37.1492	53.0739	53.365	0	72.5938
101	34.0851	34.6360	13.298	1	69.5670
102	37.8289	36.0122	106.067	0	62.9715
103	40.5501	31.8986	228.922	1	75.5686
104	41.6061	49.6269	57.271	0	68.4558

Select a sample of 104 rows of the data to use (each row represents a weekly test sample so that the 104 total rows represent the two year's worth of data). You should examine your worksheet to confirm that it has 104 rows in each of the following columns:

Column C1: 'alum_ppm'	The amount of aluminum pollutants in the weekly test sample in parts per million (ppm)
Column C2: 'oth_ppm'	The amount of other pollutants in the weekly test sample in parts per million (ppm)
Column C3: 'puretime'	The amount of time that it takes to purify a weekly test sample in seconds
Column C4: 'epacode'	This is a coded variable that takes on the value: 0 if the sample is not "polluted" per EPA guidelines 1 if the sample is "polluted" per EPA guidelines
Column C5 'temp'	The temperature of the sample in degrees Fahrenheit

Requirement 1.

Your first mission is to "see" what the data looks like. You decide to use some visual descriptive statistics.

A. Construct a histogram, stem and leaf plot, and dotplot of each of the random variables in columns C1-C3.

B. For each variable, comment on the shape and symmetry of the various plots.

C. Conjecture a known distribution that might model each one of the variables.

Requirement 2.

You have a good handle on what each variable looks like. Now you want to summarize the data numerically. You decide to use some numerical descriptive statistics.

A. Produce the numerical descriptive statistics for each of the random variables in columns C1-C3.

B. With this information, estimate the value(s) of the parameter(s) for each of the distributions that you identified in requirement **1** above.

Requirement 3.

You now want to do some analysis of the variable 'epacode'.

A. What type of random variable is 'epacode'?

B. Find the percent of samples that were "polluted" per EPA guidelines.

C. What is the probability that at least 4 of the next 16 weekly test samples will be "polluted" per the EPA guidelines?

D. What assumptions are you making to calculate the probability in part (**B.**)?

REQUIREMENT 4.

The next variable you want to analyze is ‘puretime’

A. What type of random variable does ‘puretime’ appear to be?

B. What is the probability that a randomly selected weekly test sample will take at most 35 seconds to purify?

C. The analysis of all of this data has taken you about five weeks so far. During this time five new weekly test samples have been taken. You define a “success” as a sample that will take at most 35 seconds to purify. What is the probability that the five new test samples will all be considered successes?

REQUIREMENT 5.

Your boss says that the CEO wants a quick update on your work so far. You have time to analyze one more variable. You decide to investigate ‘oth_ppm’.

A. What type of random variable is ‘oth_ppm’?

B. What is the probability that the amount of other pollutants in a randomly selected sample is between 40 ppm and 55 ppm?

PART 6

PREREQUISITE SKILLS:

9. SOLVING LINEAR REGRESSION PROBLEMS

You just completed your update briefing with the CEO. He was very impressed with the work that you and your team have done so far. The CEO told you during the briefing that the environmentalist groups are pursuing pollution problems that are being generated by a number of different industries. The environmentalists have limited resources so they plan to focus in on one industry and see if they can make that industry fix its pollution problem. The CEO's primary objective is to make sure that the aluminum processing industry is not the industry that the environmentalists select to be their target. He wants you to take a real hard look at the aluminum pollutants data and see what kind of information you can gather which might help the aluminum processing industry's cause.

REQUIREMENT 1.

Since the CEO is interested in aluminum pollutants, you decide to analyze the variable 'alum_ppm' from your data set.

A. What type of random variable does the variable 'alum_ppm' appear to be?

B. What is the probability that the amount of aluminum in a randomly selected weekly test sample is at least 36 parts per million (ppm)?

C. You now calculate the average amount of aluminum that is in a sample by averaging all 104 weekly test samples. What is the probability that the average amount of aluminum in a sample is at least 36 ppm?

D. Give the name of the theorem that you used to calculate the answer in part (**C.**) and state why the theorem applies in this case.

REQUIREMENT 2.

The CEO wants a "solid" estimate of the amount of aluminum pollutants that are in the Great Lakes.

A. Provide a point estimate of the population mean of the amount of aluminum pollutants that are in a weekly test sample.

B. Provide an interval estimate of the population mean of the amount of aluminum pollutants that are in a weekly test sample. Use a confidence level of 90%.

C. In words, give an interpretation of what the interval estimate that you found in part (**B.**) is telling you.

D. Provide an interval estimate of the population standard deviation of aluminum pollutants that are in a weekly test sample. Use a confidence level of 95%.

REQUIREMENT 3.

You know that another technique you can use to draw conclusions about your aluminum pollutants data is hypothesis testing.

A. You want to show that the true population mean of the amount of aluminum pollutants that are in a weekly test sample is less than 36 ppm. Test this claim at a significance level of 0.05. Use a sketch to support your conclusion.

B. Historical data over the years shows that the true mean of the amount of aluminum pollutants that are in a weekly test sample is 35 ppm. Perform a Type II error analysis on your test assuming that the historical data is accurate. Use a sketch to support your conclusions.

C. What is the power of your test in part (**B.**)? What can you do to increase the power of your test without increasing the probability of other errors? What is the possible drawback with using this method?

REQUIREMENT 4.

You are concerned that there is a significant relationship between the amount of aluminum pollutants in a weekly test sample and the temperature (degrees Fahrenheit) of the test sample. You decide to explore this issue in more detail.

A. Construct a plot of the amount of aluminum pollutants in a weekly test sample versus the sample temperature. Comment on the shape of the plot.

B. Determine the correlation between the amount of aluminum pollutants in a weekly test sample and the temperature of the sample. What does this value tell you?

C. Give the equation of the line that best fits the data that you have plotted in part (**A.**). What is the amount of aluminum pollutants that you would expect in a weekly sample that was measured at 65 degrees Fahrenheit?

REQUIREMENT 5.

You came up with an equation of a line in requirement **4** to summarize your data. You want to make sure that the CEO knows the statistical principles that you used to obtain that line.

A. State the simple linear regression model that you used and list any assumptions that you made when you decided to use this model.

B. Give the value of the coefficient of determination for your regression analysis. What can you say about your model based on this value?

C. Test the hypothesis that your regression line is statistically significant. Use a significance level of 0.05. Does this reinforce your conclusions from part (**B.**)?

Lake Pollution

SAMPLE SOLUTIONS: 6 PARTS

INTERDISCIPLINARY LIVELY APPLICATIONS PROJECT

TITLE: LAKE POLLUTION

AUTHORS: WILLIAM FOX
WILLIAM GOETZ
HERMANN KOLEV
RONALD MILLER
JEFFERY STRICKLAND
WENDELL KING

MANAGING EDITOR: DAVID C. ARNEY

PRODUCTION EDITOR: PAULINE WRIGHT

EDITORS: DONALD SMALL
WADE ELLIS
KATHLEEN SNOOK

DEPARTMENT OF MATHEMATICAL SCIENCES, DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL ENGINEERING, AND DEPARTMENT OF CHEMISTRY, UNITED STATES MILITARY ACADEMY, WEST POINT, NY

DATE: JANUARY 1995

CARRY-THROUGH SUBJECTS: DISCRETE MATHEMATICS, DIFFERENTIAL EQUATIONS, PROBABILITY, STATISTICS

MATHEMATICS CLASSIFICATIONS: DIFFERENCE EQUATIONS, DIFFERENTIAL EQUATIONS, DISCRETE MATHEMATICS, PROBABILITY, STATISTICS, LINEAR REGRESSION, LINEAR ALGEBRA

DISCIPLINARY CLASSIFICATIONS: ENVIRONMENTAL ENGINEERING

PREREQUISITE SKILLS:

1 .MODELING WITH DIFFERENCE EQUATIONS
2. SOLVING FIRST- AND SECOND-ORDER DIFFERENCE EQUATIONS
3. SOLVING LINEAR SYSTEMS OF EQUATIONS
4. SOLVING LINEAR SYSTEMS OF DIFFERENCE EQUATIONS
5. MODELING WITH DIFFERENTIAL EQUATIONS
6. SOLVING FIRST-ORDER DIFFERENTIAL EQUATIONS USING NUMERICAL AND ANALYTIC METHODS
7. MODELING WITH STOCHASTIC PROCESSES (MARKOV CHAINS)
8. USING DESCRIPTIVE STATISTICS, CLASSICAL PROBABILITY, BAYES' THEOREM, AND PROBABILITY DISTRIBUTIONS
9. SOLVING LINEAR REGRESSION PROBLEMS

MATERIALS AVAILABLE:

A. PROBLEM STATEMENT (6 PARTS) (STUDENT)
B. SAMPLE SOLUTION (INSTRUCTOR)
C. BACKGROUND MATERIAL (INSTRUCTOR)

INTERDISCIPLINARY LIVELY APPLICATIONS PROJECT IS FUNDED BY THE NATIONAL SCIENCE FOUNDATION, DIRECTORATE OF EDUCATION AND HUMAN RESOURCES DIVISION OF UNDERGRADUATE EDUCATION, NSF GRANT #9455980.

NSF INITIATIVE:
MATHEMATICS SCIENCES AND THEIR APPLICATIONS THROUGHOUT THE CURRICULUM (CCD-MATH)

PART 1

REQUIREMENT 1.

The system that models the amount of pollutants in Lakes Erie and Ontario is

$$a(n+1) = 0.62\, a(n)$$

$$b(n+1) = 0.87\, b(n) + 0.38\, a(n) + 25$$

where $a(n)$ = the amount of pollution in Lake Erie after n years,

and $b(n)$= the amount of pollution in Lake Ontario after n years.

The second order system that models the amount of pollutants in Lake Ontario is

$$b(n+2) = 1.49\, b(n+1) - 0.5394\, b(n) + 9.5$$

REQUIREMENT 2.

The general solution to this equation is

$$b(k)= c_1(.87)^k + c_2\,(.62)^k + 192.3077$$

REQUIREMENT 3.

The equilibrium value is 192.3077. This means that if we begin with an initial value of 192.3077, then each iteration will yield 192.3077.

REQUIREMENT 4.

Using the boundary conditions $b(0) = 2500$ and $b(1) = 3150$, the particular solution to the discrete dynamical system is

$$b(k) = 6107.6923\,(.87)^k - 3799.9999\,(.62)^k + 192.3077$$

It will take 34 years for the pollution level in Lake Ontario to be reduced to 10 percent of its present level.

REQUIREMENT 5.

As k increases, the system approaches the equilibrium value. In other words, after a long period of time, the pollution level in Lake Ontario reaches 192.3077 units of pollution. This is less than eight percent of the initial level.

PART 2

REQUIREMENT 1.

The system that models the amount of pollutants in Lakes Huron, Erie and Ontario is

$$A(n+1) = \begin{bmatrix} 0.89 & 0.0 & 0.0 \\ 0.11 & 0.64 & 0.0 \\ 0.0 & 0.36 & 0.88 \end{bmatrix} \begin{bmatrix} a(n) \\ b(n) \\ c(n) \end{bmatrix} + \begin{bmatrix} 30 \\ 0 \\ 30 \end{bmatrix}$$

where $a(n)$ = the amount of pollution in Lake Huron after n years,

$b(n)$ = the amount of pollution in Lake Erie after n years,

and $c(n)$ = the amount of pollution in Lake Ontario after n years.

REQUIREMENT 2.

The general solution to the system is

$$A(k) = c_1 \left(\frac{16}{25}\right)^k \begin{bmatrix} 0 \\ 1 \\ -\frac{3}{2} \end{bmatrix} + c_2 \left(\frac{22}{25}\right)^k \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + c_3 \left(\frac{89}{100}\right)^k \begin{bmatrix} 1 \\ \frac{11}{25} \\ \frac{396}{25} \end{bmatrix} + \begin{bmatrix} 272.7273 \\ 83.3333 \\ 500 \end{bmatrix}$$

REQUIREMENT 3.

The equilibrium vector is $\begin{bmatrix} 272.7273 \\ 83.3333 \\ 500 \end{bmatrix}$. If we begin with an initial vector $\begin{bmatrix} 272.7273 \\ 83.3333 \\ 500 \end{bmatrix}$,

then each iteration will yield $\begin{bmatrix} 272.7273 \\ 83.3333 \\ 500 \end{bmatrix}$

REQUIREMENT 4.

The particular solution to the discrete dynamical system is

$$A(k) = 296.666\left(\frac{16}{25}\right)^k \begin{bmatrix} 0 \\ 1 \\ -\frac{3}{2} \end{bmatrix} - 48775\left(\frac{22}{25}\right)^k \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 3227.27\left(\frac{89}{100}\right)^k \begin{bmatrix} 1 \\ \frac{11}{25} \\ \frac{396}{25} \end{bmatrix} + \begin{bmatrix} 272.7273 \\ 83.3333 \\ 500 \end{bmatrix}$$

It will take 33 years and 24 years, respectively, for the pollution levels in lakes Huron and Erie to be reduced to 10 percent of their present levels. The pollution level in Lake Ontario will never be reduced to the 10 percent level.

REQUIREMENT 5.

As k increases, the system approaches the equilibrium vector. In other words, after a long period of time, the pollution level in Lake Huron becomes 272.7273, the pollution level in Lake Erie becomes 83.3333, and the pollution level in Lake Ontario reaches 500 units of pollution. This is less than ten percent of the initial level for Lake Huron and Lake Erie, but is 20.83 percent of the initial level for Lake Ontario.

PART 3

REQUIREMENT 1.

Given that $C(t)$ is the concentration, we want to write a differential equation in terms of $C(t)$. We use the relationship, $C(t) = \frac{Q(t)}{V}$, where V is a constant volume, and differentiate both sides of the equation:

$$\frac{dC}{dt} = \frac{1}{V}\frac{dQ}{dt} .$$

We know that $\frac{dQ}{dt} = R_{inputs} - R_{outputs}$. Using this, we substitute into the above differential equation, obtaining

$$\frac{dC}{dt} = \frac{1}{V}(R_{inputs} - R_{outputs})$$

$$\frac{dC}{dt} = \frac{R_{inputs} - R_{outputs}}{V} .$$

Since we know that $R_{inputs} = rk$ and $R_{outputs} = rC$, we substitute these directly into the differential equation, obtaining

$$\frac{dC}{dt} = \frac{rk}{V} - \frac{rC}{V} .$$

We also know that the bay begins with N milligrams per liter of pollutants, so our boundary condition is $C(0) = N$. Therefore, our boundary value problem is

$$\frac{dC}{dt} = \frac{rk}{V} - \frac{rC}{V} , \qquad C(0) = N .$$

In the boundary value problem above, $\frac{dC}{dt}$ is the time rate of change of pollutants in the lake, rk is the rate at which a constant concentration k pollutants enter the bay, and rC is the rate at which a concentration of C pollutants flow out of the bay. A constant volume of bay water is represented by V. Therefore, the time rate of change of pollutants in the bay is equal to the difference between rate of pollutant inputs and the rate of pollutant outputs, divided by the volume of water in the bay.

REQUIREMENT 2.

A. We find the concentration of the pollutants in the bay using the *First Order Euler Method* with a step size of 0.5 day for the first thirty days. Substituting our data, r=50,000, k=0.5, V=4,000,000, and N=0.8, into our differential equation found in Requirement 1, we obtain

$$\frac{dC}{dt} = 0.00625 - 0.0125C$$

By applying Euler's Method to this problem, we can estimate the concentration of pollution in the bay. Using an initial value of $C(0)$ = 0.8, and a step size of 0.5, our first iteration of Euler's formula yields

$$C_1 = C_0 + 0.5(0.00625 - 0.0125\,(C_0)) = 0.8 + 0.5\,(0.00625 - 0.0125\,(0.8)) = 0.798125.$$

We continue this iterating process with a spreadsheet, where the n+1 iteration of Euler's formula is

$$C_{n+1} = C_n + 0.5(0.00625 - 0.0125(C_n)).$$

Our spreadsheet results (below) show us that our estimate for the concentration of pollutants in the bay after 30 days is 0.705944 milligrams of pollutants per liter of bay water.

# days	concentration	# days	concentration	# days	concentration
0	0.8	10.5	0.762991	21	0.730548
0.5	0.798125	11	0.762991	21.5	0.730548
1	0.796262	11.5	0.759714	22	0.727675
1.5	0.79441	12	0.758091	22.5	0.726252
2	0.79257	12.5	0.756478	23	0.724838
2.5	0.790741	13	0.754875	23.5	0.723433
3	0.788924	13.5	0.753282	24	0.722036
3.5	0.787119	14	0.751699	24.5	0.720649
4	0.785324	14.5	0.750126	25	0.71927
4.5	0.783541	15	0.748562	25.5	0.717899
5	0.781769	15.5	0.748562	26	0.716537
5.5	0.780008	16	0.745465	26.5	0.715184
6	0.778258	16.5	0.743931	27	0.713839
6.5	0.776518	17	0.742406	27.5	0.712503
7	0.77479	17.5	0.740891	28	0.711174
7.5	0.773073	18	0.739386	28.5	0.709855
8	0.771366	18.5	0.73789	29	0.708543
8.5	0.766967	19	0.736403	29.5	0.70724
9	0.767985	19.5	0.734925	30	0.705944
9.5	0.76631	20	0.733457		
10	0.764645	20.5	0.731988		

B. Again, our differential equation is $\frac{dC}{dt} = \frac{50{,}000(0.5)}{4{,}000{,}000} - \frac{50{,}000C}{4{,}000{,}000}$, $C(0) = 0.8$

or

$$\frac{dC}{dt} = \frac{1}{160} - \frac{C}{80}, \quad C(0) = 0.8.$$

We can use the technique of Separation of Variables on a differential equation of the form $\frac{dy}{dx} = \frac{g(x)}{h(y)}$.

Applying the technique here, we have

$$\frac{dC}{dt} = \frac{1}{160} - \frac{C}{80} \qquad (1)$$

$$\frac{dC}{\frac{1}{160} - \frac{C}{80}} = dt \quad \text{or} \quad \frac{-160dC}{2C-1} = dt \qquad (2)$$

We obtain the second equation by dividing equation (1) by the term on the right hand side, and multiplying equation(1) by dt. Next, we integrate both sides of equation (2).

$$-160\int \frac{dC}{2C-1} = \int dt \quad .$$

We use u-substitution to integrate the left-hand side, letting $u = 2C - 1$, obtaining $du = 2\, dC$. Applying this, we have

$$-80\int \frac{2dC}{2C-1} = \int dt$$

$$-80\int \frac{du}{u} = \int dt$$

$$-80 \ln u = t + b_1 \quad \text{or} \quad -80 \ln (2C-1) = t + b_1 \qquad (3)$$

Now, if we exponentiate equation (3), we have

$$e^{ln(2C-1)^{-80}} = e^{t+b_1} \qquad (4)$$

Simplifying equation (4), we obtain

$$(2C-1)^{-80} = be^{t} \quad \text{or} \quad \left(\frac{1}{2C-1}\right)^{80} = be^{t}$$

where $b = e^{b_1}$.

We can simplify this expression to

$$\frac{1}{2C-1} = b^{0.0125}\, e^{0.0125t} \ . \qquad \textbf{(5)}$$

Now we use the initial condition $C(0) = 0.8$ to solve for b.

$$\frac{1}{2(.8)-1} = b^{\,0.0125}\, e^{0.0125(0)} \ .$$

$$b = 5.59628 \times 10^{17} \ .$$

Substituting $b = 5.59628 \times 10^{17}$ into equation (5), we obtain

$$\frac{1}{2C-1} = (5.59628 \times 10^{17})^{(0.0125)}\, e^{0.0125t} \ .$$

$$\frac{1}{2C-1} = 1.66666e^{\,0.0125t} \quad \text{or} \quad C = 0.5+0.3e^{-0.0125t}.$$

Therefore, our particular solution to the differential equation is

$$C(t) = 0.5+0.3e^{-0.0125t}$$

This equation gives us the concentration of pollutants in the bay at any time t. At t=30, we have $C(30)$=0.706187 milligrams of pollutants per liter of water.

C. Euler's Method gives us $C(30)$=0.705944, while the solution to the differential equation yields $C(30)$=0.706187. The second method is much more accurate. When we use an approximating or estimating method, like Euler's, we have both round-off error and theoretical error. In the computation using our particular solution, we only have round-off error.

D. The pollution concentration will achieve a steady state value. We determine this value by computing the limit of $C(t)$ as t goes to infinity.

$$\lim_{t \to \infty} C(t) = \lim_{t \to \infty} (0.5+0.3e^{-.0125t}) = 0.5 \ .$$

Therefore, the steady state value is 0.5 milligrams of pollutants per liter of water.

REQUIREMENT 3.

We know that 4,000,000 milligrams are mixed with 4,000,000 liters of bay water, giving us a concentration of 1 milligram of pollutant per liter of water. This is added to the 0.8 milligram already present, giving us an initial concentration of 1.8 milligrams of pollutant per liter of water. We solve our general solution to obtain a new particular solution corresponding to $C(0) = 1.8$.

$$C(0) = 1.8 = 0.5 + be^{-.0125(0)}$$

$$b = 1.3 .$$

So our particular solution is $C(t) = 0.5 + 1.3e^{-.0125t}$.

Now we want to know when the concentration of pollutant will be back down to a safe level of 1.6 milligrams of pollutant per liter of water.

So we solve $1.6 = 0.5 + 1.3e^{-.0125t}$ for t.

$$1.6 = 0.5 + 1.3e^{-.0125t} \qquad \text{or} \qquad 1.1 = 1.3e^{-.0125t}.$$

Taking the natural log of both sides of the equation, we have

$$\ln(1.1) = \ln(1.3) + \ln(e^{-.0125t})$$

$$\ln(1.1) = \ln(1.3) - 0.0125t$$

$$\ln(1.3) - \ln(1.1) = 0.0125t$$

$$t = 13.36433$$

This means that the factory will have to shut down for approximately 13.36 years.

REQUIREMENT 4.

This model assumes that the bay has a constant volume of water, that the only sources of pollution come from the stream and the factory, and that the rates of pollutant input and output are the same.

Factors not accounted for in this model include, (1) other sources of pollution, (2) variation of water level, (3) seasonal differences in input, and factors which affect input and output, such as (4) collection of pollutants in bottom sediment and plants, as well as (5) bacterial conversion of pollutants. More specifically:

1) Other pollutants might enter the bay through run-off from farms and residences along or near the shore. These include soap and petroleum products, and other solid and liquid wastes.

2) Variation in the water level due to drought or heavy rains will affect the concentration of pollutants. The same factors will affect the amount of pollutants entering due to run-off or drainage.

3) Seasonal changes may affect the concentration of pollutants. In addition to water level changes, we should consider the effect of storms, wind currents, etc.

4) Some of the pollutants will settle into the bottom sediment and in areas where underwater plant growth is dense.

5) Some pollutants may be converted to a harmless form by certain bacteria.

PART 4

REQUIREMENT 1.

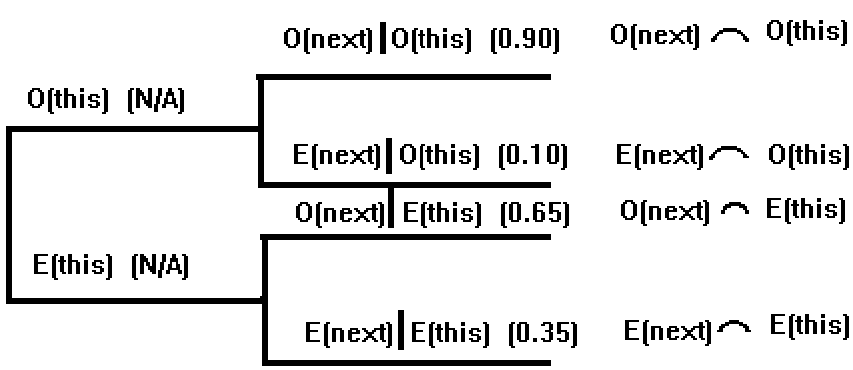
Define:

O(this) as the event that traces of aluminum pollutants are found in this week's test of Lake Ontario

O(next) as the event that traces of aluminum pollutants are found in next week's test of Lake Ontario

E(this) as the event that traces of aluminum pollutants are found in this week's test of Lake Erie

E(next) as the event that traces of aluminum pollutants are found in next week's test of Lake Erie



REQUIREMENT 2.

Assumptions (there are others):

a. You are assuming that the only source of aluminum pollutants in the system is the aluminum plant that you work for.

b. You are assuming that the results of next week's test are only dependent on this week's test and no other earlier tests.

REQUIREMENT 3.

Define:

O(*n*) = the probability that traces of aluminum were found in Lake Ontario's test from week *n*.

E(*n*) = the probability that traces of aluminum were found in Lake Erie's test from week *n*.

System of Equations:

$$O(n+1) = (0.90)\,O(n) + (0.65)\,E(n)$$

$$E(n+1) = (0.10)\,O(n) + (0.35)\,E(n)$$

or in matrix form:

$$\mathbf{A}(n+1) = \begin{bmatrix} O(n+1) \\ E(n+1) \end{bmatrix} = \begin{bmatrix} 0.90 & 0.65 \\ 0.10 & 0.35 \end{bmatrix} \begin{bmatrix} O(n) \\ E(n) \end{bmatrix}$$

Eigenvalues: 0.25, 1.0

Eigenvectors: $\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0.8667 \\ 0.1333 \end{bmatrix}$

General Solution:

$$\mathbf{A}(k) = C_1\,(0.25)^k \begin{bmatrix} 1 \\ -1 \end{bmatrix} + C_2\,(1)^k \begin{bmatrix} 0.8667 \\ 0.1333 \end{bmatrix}$$

$$\lim_{k \to \infty} \mathbf{A}(k) = C_2 \begin{bmatrix} 0.8667 \\ 0.1333 \end{bmatrix}$$

∴ P(aluminum in this week's test of Lake Ontario) = 0.8667

P(aluminum in this week's test of Lake Erie) = 0.1333

REQUIREMENT 4.

A. $P(O(next) \cap O(this)) + P(E(next) \cap E(this)) =$

$$(0.90)(0.8667) + (0.35)(0.1333) = 0.8267$$

B. $P(E(this) | E(next)) = \dfrac{P(E(this) \cap E(next))}{P(E(next))} = \dfrac{0.0467}{0.0467 + 0.0867} = 0.3501$

C. Lake Ontario has the higher probability of having traces of aluminum in it. The numerical value of the probability is 0.8667.

D. The answer in **C.** above makes some sense since the aluminum plant is on Lake Erie and the primary direction of flow is from Lake Erie to Lake Ontario. In the long term, you would expect to see a better chance of pollutants eventually show up in Lake Ontario.

REQUIREMENT 5.

The value of zero should be in the two blank spaces because Lake Huron and Lake Ontario are not directly connected to one another.

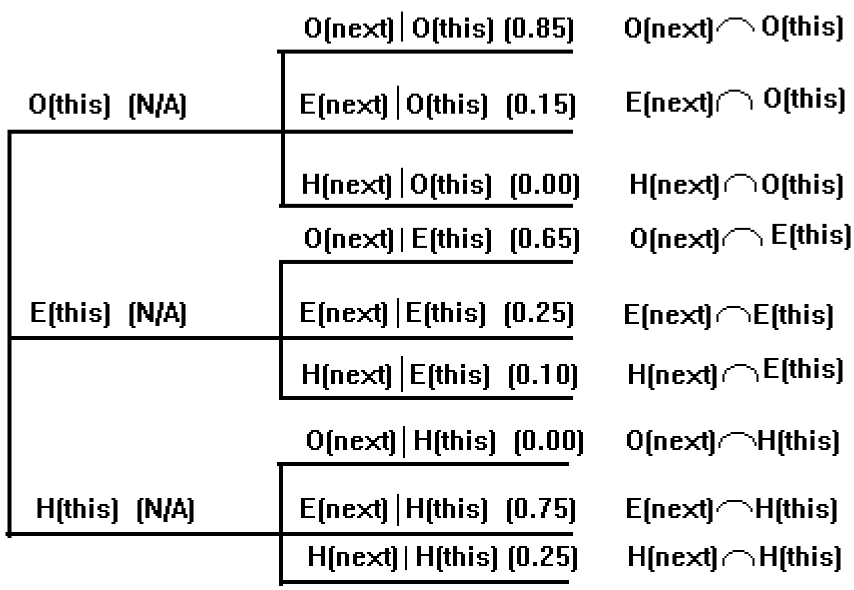
	Probability that traces of aluminum will be found in next week's test of Lake_______		
	Huron	**Erie**	**Ontario**
Traces of aluminum were found in this week's test of Lake _________			
Huron	0.25	0.75	**0.00**
Erie	0.10	0.25	0.65
Ontario	**0.00**	0.15	0.85

REQUIREMENT 6.

Define:

H(this) as the event that traces of aluminum pollutants are found in this week's test of Lake Huron

H(next) as the event that traces of aluminum pollutants are found in next week's test of Lake Huron



Define:

H(*n*) = the probability that traces of aluminum were found in Lake Huron's test from week *n*.

System of Equations:

$$O(n+1) = (0.85)\ O(n) + (0.65)\ E(n)$$

$$E(n+1) = (0.15)\ O(n) + (0.25)\ E(n) + (0.75)\ H(n)$$

$$H(n+1) = (0.10)\ E(n) + (0.25)\ H(n)$$

or in matrix form:

$$\mathbf{A}(n+1) = \begin{bmatrix} O(n+1) \\ E(n+1) \\ H(n+1) \end{bmatrix} = \begin{bmatrix} 0.85 & 0.65 & 0.00 \\ 0.15 & 0.25 & 0.75 \\ 0.00 & 0.10 & 0.25 \end{bmatrix} \begin{bmatrix} O(n) \\ E(n) \\ H(n) \end{bmatrix}$$

Eigenvalues (from Derive): 1.0, 0.4312, - 0.0812

Eigenvector corresponding to the eigenvalue of 1.0:

$$\begin{bmatrix} 0.7927 \\ 0.1829 \\ 0.0244 \end{bmatrix}$$

Note: We do not have to find the other two eigenvectors since they will be multiplied by eigenvalues which will approach zero as *k* approaches infinity.

General Solution:

$$\mathbf{A}(k) = C_1\ (1.0)^k \begin{bmatrix} 0.7927 \\ 0.1829 \\ 0.0244 \end{bmatrix} + C_2\ (0.4312)^k \begin{bmatrix} xx \\ xx \\ xx \end{bmatrix} + C_3\ (-0.0812)^k \begin{bmatrix} xx \\ xx \\ xx \end{bmatrix}$$

$$\lim_{k \to \infty} \mathbf{A}(k) = C_1 \begin{bmatrix} 0.7927 \\ 0.1829 \\ 0.0244 \end{bmatrix}$$

$\therefore$ P(aluminum in this week's test of Lake Ontario) = 0.7927

P(aluminum in this week's test of Lake Erie) = 0.1829

P(aluminum in this week's test of Lake Huron) = 0.0244

REQUIREMENT 7.

A. $P(O(this) \cap O(next)) + P(E(this) \cap E(next)) + P(H(this) \cap H(next))$ =

(0.7927)(0.85) + (0.1829)(0.25) + (0.0244)(0.25) = 0.7256

This value is slightly lower than the value we found in requirement 4 (A.) You cannot really make any other comparison since the two requirements used data matrices that were not related.

B. $P(H(this)|E(next)) = \dfrac{P(H(this) \cap E(next))}{P(E(next))}$ =

$$\frac{(0.0244)(0.75)}{(0.0244)(0.75) + (0.1829)(0.25) + (0.7927)(0.15)} = 0.1000$$

C. $P(O(this) \cap E(next)) = (0.7927)(0.15) = 0.1189$

PART 5

SAMPLE DATA

ROW	alum_ppm	oth_ppm	puretime	epacode	temp
1	36.2846	59.7390	12.105	1	70.0692
2	35.1909	43.6978	37.864	0	68.8021
3	38.5753	45.4508	67.275	1	71.9412
4	29.8268	55.6974	45.286	1	62.3037
5	32.7359	43.6578	81.583	0	64.9527
6	33.4072	56.6556	6.297	1	66.1605
7	31.5494	49.9503	32.174	1	63.9796
8	29.4830	59.2501	15.345	1	63.8191
9	28.3125	59.2778	103.952	1	61.8358
10	33.3508	30.8085	44.684	0	67.2914
11	35.4065	50.9860	16.497	1	68.2996
12	40.8595	59.0104	28.704	1	72.8905
13	38.0167	57.1471	67.667	1	70.5572
14	48.5467	48.4696	36.098	1	80.3929
15	38.2178	47.4461	17.372	1	70.1867
16	35.7263	37.0215	33.567	0	68.3591
17	44.3164	52.0141	23.307	1	75.7433
18	33.7658	53.8839	109.182	1	66.4860
19	38.6447	35.8155	1.579	0	70.8557
20	44.4616	43.8406	2.717	1	76.2052
21	37.8688	58.4949	72.884	1	70.2122
22	33.2331	56.7300	131.370	1	66.7332
23	39.3477	45.6032	57.226	1	72.0411
24	34.2471	37.7449	58.106	0	66.5119
25	38.9567	45.7960	16.483	1	71.3661
26	37.2145	44.0806	11.400	1	69.1802
27	39.3465	55.0693	63.580	1	71.9098
28	34.6710	31.5343	8.304	0	67.6962
29	26.6301	57.3854	18.576	1	60.7372
30	30.8516	30.2706	91.128	0	64.3912
31	30.8723	42.1799	5.037	0	63.3019
32	29.7076	44.9115	17.931	0	62.7968
33	42.7976	40.9325	38.233	1	74.7648
34	44.4128	57.4171	52.001	1	75.9615
35	28.4938	30.2881	12.810	0	62.3881
36	32.8682	46.1233	30.378	0	66.9608
37	35.1048	39.4139	17.783	0	67.8019
38	40.9571	46.7199	0.608	1	73.3128
39	28.7226	36.3994	27.310	0	62.4025
40	32.7272	42.1318	65.823	0	65.2770
41	30.2132	54.0219	18.391	1	64.4110
42	29.0897	34.4747	56.934	0	62.9142
43	34.0419	49.2455	168.731	1	67.5920
44	28.1496	43.0440	29.275	0	62.4702
45	35.2347	58.2225	5.884	1	67.3556
46	37.2364	57.4492	4.351	1	70.5495
47	29.1620	53.2136	42.727	1	63.2839
48	40.4460	30.9330	32.152	0	72.5938
49	35.7201	48.8363	11.040	1	69.5670
50	29.2285	32.6732	137.548	0	62.9715
51	42.9233	53.6791	40.148	1	75.5686
52	34.9017	31.7362	69.898	0	68.4558

SAMPLE DATA

ROW	alum_ppm	oth_ppm	puretime	epacode	temp
53	26.6193	35.7099	21.848	0	59.7710
54	32.8131	39.6966	19.566	0	64.9707
55	30.6090	48.3676	25.052	0	63.1579
56	28.5616	42.1876	1.559	0	61.6976
57	30.4919	53.2427	19.179	1	63.7523
58	39.5930	47.2459	44.036	1	71.4984
59	32.4821	31.4635	46.615	0	65.7153
60	35.1442	43.3965	22.226	0	68.2051
61	33.9530	44.8305	18.446	0	67.0155
62	29.4919	46.8731	150.477	0	63.7843
63	28.4642	39.5781	15.591	0	60.9638
64	26.1369	37.0135	14.602	0	60.3454
65	35.1675	37.8624	4.265	0	67.6119
66	37.4473	37.1184	40.328	0	69.8011
67	37.6916	36.5494	33.648	0	69.9018
68	24.6438	41.2847	2.789	0	58.6964
69	30.0035	48.9843	91.200	0	62.4671
70	26.0824	50.6743	146.261	0	59.0033
71	31.5738	48.5370	33.855	1	64.0707
72	33.6830	33.1112	43.153	0	65.7709
73	30.7659	54.3590	26.966	1	64.6611
74	28.5338	56.7471	192.736	1	61.7126
75	29.9502	34.9079	58.938	0	63.9489
76	39.0303	43.4007	7.306	1	71.5458
77	40.7136	34.6556	29.328	0	72.5465
78	47.5382	42.0995	42.509	1	80.2774
79	27.8366	55.4403	15.455	1	60.4900
80	34.5073	42.9701	13.940	0	67.0364
81	33.2366	56.2188	18.492	1	66.2645
82	38.1850	39.4222	15.752	0	71.2584
83	31.3074	33.5046	6.902	0	63.9580
84	32.5488	38.6973	22.425	0	66.5546
85	30.3595	59.6348	25.319	1	63.0884
86	36.6709	38.3582	75.050	0	70.0097
87	37.7908	46.5821	3.351	1	70.9777
88	33.5284	36.8290	9.954	0	66.3875
89	33.9880	35.3832	28.603	0	67.7514
90	30.6810	56.9938	41.422	1	63.1022
91	33.2926	33.0390	9.068	0	66.6784
92	34.0935	37.8167	6.904	0	66.2453
93	37.5900	51.4286	13.276	1	70.7118
94	35.6707	37.3356	0.195	0	67.7272
95	35.0608	30.3771	23.171	0	68.2565
96	30.6652	37.7568	40.732	0	64.8418
97	33.3158	43.9597	78.473	0	67.0687
98	32.9712	50.1189	2.845	1	66.6186
99	43.8148	47.6849	28.864	1	76.5815
100	37.1492	53.0739	53.365	1	68.9791
101	34.0851	34.6360	13.298	0	66.2452
102	37.8289	36.0122	106.067	0	71.2823
103	40.5501	31.8986	228.922	0	73.7126
104	41.6061	49.6269	57.271	1	74.0190

REQUIREMENT 1.

A.

dotplot alum_ppm

```
                 . .
```

Histogram of alum_ppm *N* = 104

Midpoint	Count	
24	1	*
26	4	****
28	8	********
30	18	******************
32	10	**********
34	18	******************
36	12	************
38	15	***************
40	9	*********
42	3	***
44	4	****
46	0	
48	2	**

Stem-and-leaf of alum_ppm *N* = 104

Leaf Unit = 1.0

1	2	4
6	2	66667
21	2	888888899999999
34	3	0000000000111
52	3	222222233333333333
52	3	44444445555555555
35	3	66777777777
24	3	8888889999
14	4	000001
8	4	223
5	4	444
2	4	7
1	4	8

```
             :   .   .:: :    :

             :..::.. :::: ::  ..:  . .       .

        . ::  :::::::: ::::::: ::::::: ::. : .:     . .

   -----+---------+---------+---------+---------+---------+------alum_ppm

      25.0      30.0      35.0      40.0      45.0      50.0      55.0
```

Histogram of oth_ppm   *N* = 104

Midpoint	Count	
32	13	*************
36	19	*******************
40	8	********
44	18	******************
48	15	***************
52	10	**********
56	14	**************
60	7	*******

Stem-and-leaf of oth_ppm   *N* = 104

Leaf Unit = 1.0

9	3	000001111
13	3	2333
20	3	4444555
32	3	666677777777
38	3	889999
40	4	01
52	4	222223333333
52	4	444555
46	4	6666777
39	4	88888999
31	5	0001
27	5	233333
21	5	44555
16	5	666667777
7	5	8899999

dotplot oth_ppm

```
                  .
     .   . ..::  :   : ::  . .. :        .     :. .
 ::::.:..:.::::: : ..: :: ::.:: ::::.....::.:.::::.:.
—-+—————-+—————-+—————-+—————-+—————-+————oth_ppm
30.0     36.0     42.0     48.0     54.0     60.0     66.0
```

Histogram of puretime   *N* = 104

```
Midpoint  Count
     0      19  *******************
    20      37  *************************************
    40      20  ********************
    60      12  ************
    80       4  ****
   100       5  *****
   120       0
   140       3  ***
   160       2  **
   180       0
   200       1  *
   220       1  *
```

Stem-and-leaf of puretime   *N* = 104

```
Leaf Unit = 10

  42    0 00000000000000000001111111+
 (23)   0 22222222222222333333333
  39    0 444444444445555555
  21    0 66666777
  13    0 899
  10    1 000
   7    1 33
   5    1 45
   3    1 6
   2    1 9
   1    2
   1    2 2
```

dotplot puretime

```
    :
 :..:. .
 ::::::: .:
:::::::: ::  . .
:::::::::::.:::::..:  :  .:     .. ..  .     .       .
+---------+---------+---------+---------+---------+---------+puretime
0        40        80       120       160       200       240
```

B. alum_ppm: The plots look fairly symmetric and have a shape similar to the normal distribution

oth_ppm: The plots look fairly symmetric and have a shape similar to the uniform distribution

puretime: The plots are not symmetric and have a shape similar to the exponential distribution

C. alum_ppm: normal distribution

oth_ppm: uniform distribution

puretime: exponential distribution

REQUIREMENT 2.

A.

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN	MIN	MAX	Q1	Q3
alum_ppm	104	34.496	34.015	34.342	4.925	0.483	24.644	48.547	30.623	37.819
oth_ppm	104	44.724	44.020	44.699	8.769	0.860	30.271	59.739	37.046	52.809
puretime	104	41.24	28.65	36.12	42.62	4.18	0.20	228.92	14.11	56.04

B. alum_ppm: mean= 34.496, variance= 24.256

oth_ppm: *a*= 30.271, *b*= 59.739

puretime: lambda = 1/41.24 = 0.02425

REQUIREMENT 3.

A. The random variable epacode appears to be Bernoulli or binomial

B. Per EPA guidelines, 47.12% of the samples are “polluted”

epacode:

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN	MIN	MAX	Q1	Q3
epacode	104	0.4712	0.0000	0.4681	0.5016	0.0492	0.0000	1.0000	0.0000	1.0000

C. Let X = number of weekly test samples that are polluted in 16 trials.

$$X \sim BIN(n=16, p=0.4712)$$

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.0189 = 0.9811$$

D. The assumptions required for a binomial experiment

1. The experiment consists of a sequence of 16 trials, where 16 is fixed in advance of the experiment.
2. The trials are identical and each trial can result in one of the same two possible outcomes, which we denote as “polluted” and “not polluted”.
3. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.
4. The probability of a “polluted” sample (0.4712) is constant from trial to trial.

REQUIREMENT 4.

A. The random variable puretime appears to be exponential.

B. Let X = Time it takes for a weekly sample to purify

X ~ EXP(Lambda = 0.02425)

P(X < 35) = F(35) = 1 - Exp(-35/41.24) = 1 - 0.4279 = 0.5721

C. Let Y = Number of weekly samples that take at most 35 seconds to purify in 5 trials

Y ~ BIN(*n*=5, *p*=0.5721)

P(Y = 5) = 0.0613

REQUIREMENT 5.

A. The random variable oth_ppm appears to be uniform.

B. Let X = amount of other pollutants in a sample

X ~ UNIF(*a*= 30.271, *b*= 59.739)

P(40 < X < 55) = F(55) - F(40)

$$= \left(\frac{55-30.271}{59.739-30.271}\right)-\left(\frac{40-30.271}{59.739-30.271}\right) = 0.8392 - 0.3302 = 0.5090$$

PART 6

REQUIREMENT 1.

A. The random variable alum_ppm appears to be normally distributed.

B. Let X = the amount of aluminum in a randomly selected weekly sample

$$X \sim \text{Norm}(\mu = 34.496, \sigma = 4.925)$$

$$P(X \geq 36) = 0.3800$$

C. Let $\bar{X}$= the average amount of aluminum in the 104 weekly samples

$$\bar{X} \sim \text{Norm}(\mu = 34.496, = \sigma \frac{4.925}{\sqrt{104}})$$

$$P(\bar{X} \geq 36) = 0.0009$$

D. The Central Limit Theorem or the Reproductive Theorem apply in this case. The average in this requirement is a linear combination of normal random variables, which is also normal by the Reproductive Theorem. Because the number of samples is greater than 30, the Central Limit Theorem also applies. We are making the assumption that the original X's are independent.

REQUIREMENT 2.

A. The sample mean = 34.496 ppm provides a point estimate of the population mean.

B. An interval estimate of the population mean using a 90% confidence interval is given by

$$\bar{x} \pm z(\text{critical}) \cdot \frac{s}{\sqrt{n}}$$

$$34.496 \pm (1.645) \cdot \frac{4.925}{\sqrt{104}} \Rightarrow [35.2904, 33.7016]$$

C. The theoretical confidence interval before we sample says that the probability of capturing the true population mean in the interval is 0.90. The specific interval that we found in part (B.) after we sampled has a probability of 0 or 1 of capturing the true population mean. If we were to sample a large number of times, we would expect to see 90% of the specific intervals that we obtained to contain the true population mean.

D. Limits: $\sqrt{\dfrac{(n-1)\bullet s^2}{Chi-square(critical)}}$

Upper Limit: Chi-square(critical: ν= 103)≅ 76.803

$$\sqrt{\frac{(104-1)\bullet(4.925)^2}{76.803}} = 5.7034$$

Lower Limit: Chi-square(critical: ν=103)≅ 132.978

$$\sqrt{\frac{(104-1)\bullet(4.925)^2}{132.978}} = 4.3345$$

Interval for σ: [5.7034,4.3345]

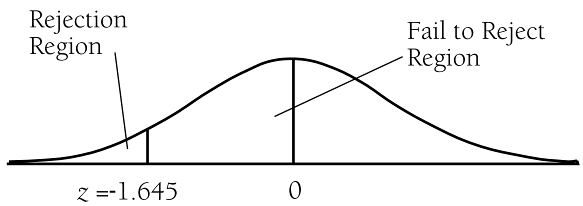
REQUIREMENT 3.

A. Ho: μ = 36

Ha: μ < 36

α = 0.05

$$z = \frac{\bar{x}-36}{\frac{s}{\sqrt{n}}}$$

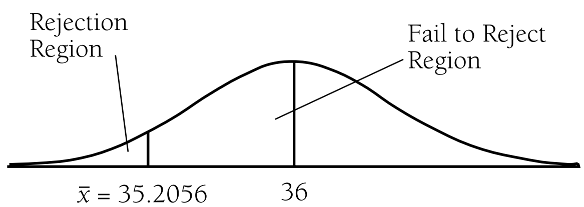


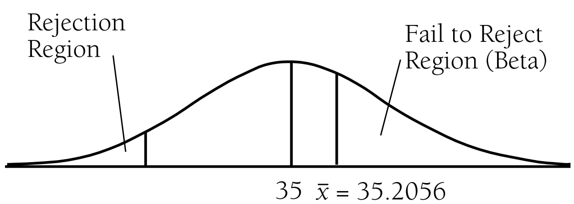
$$z = \frac{34.496-36}{\frac{4.925}{\sqrt{104}}} = -3.11$$

Since -3.11 < -1.645 , Reject Ho and conclude that the amount of aluminum pollutants in a weekly test sample is less than 36 ppm.

B. $$\frac{\bar{x}-36}{\frac{4.925}{\sqrt{104}}}=-1.645$$

$$\therefore \bar{x}=35.2056$$





$\beta = P(\bar{X} > 35.2056)=0.3351$

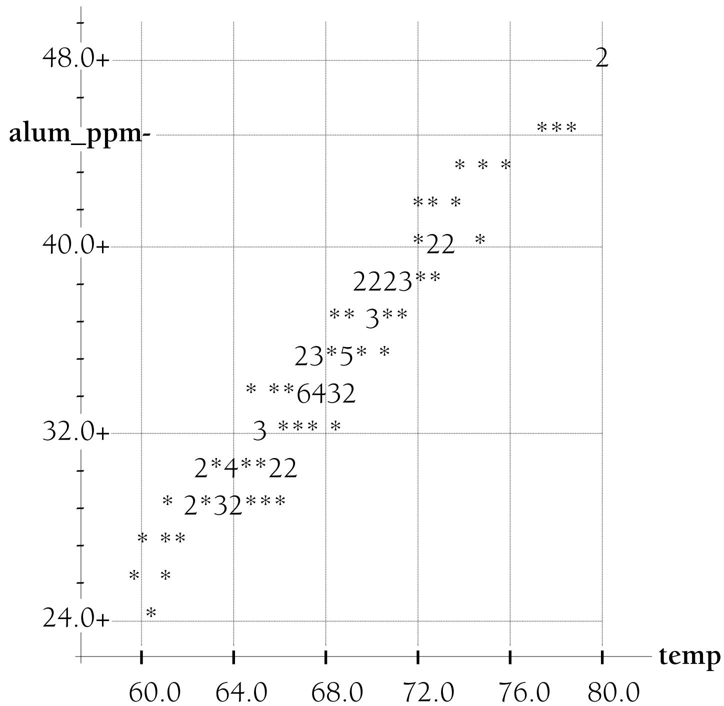
C. Power = $1-\beta = 1 - 0.3351 = 0.6649$

To increase the power without increasing the probability of other errors, you could increase the sample size *n*. The possible drawback with using this method is the cost involved with gathering additional samples.

REQUIREMENT 4.

A.

Plot of amount of aluminum pollutants versus sample temperature.



The plot appears linear with some random variation about an unknown true regression line.

B. The correlation between the amount of aluminum pollutants and the temperature of a weekly test sample is 0.991. This value tells you the degree of linear association between the two variables. In this case there is a high degree of linearity and the slope of the line is positive.

C. The equation of the line that best fits the plotted data is :

$$\text{alum\_ppm} = -37.9 + 1.07 \bullet \text{temp}$$

At 65 degrees Fahrenheit you would expect 31.65 ppm of aluminum pollutants in the sample.

REQUIREMENT 5.

A. The simple linear regression model is:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

The assumptions of the model are:

ε are i.i.d. Norm(μ =0, constant variance = σ^2) random variables

B. The value of the coefficient of determination is 0.982. This tells you that your model is accounting for 98.2% of the variation in the data and the remaining 1.8% of variation is due to error. Therefore, your linear model is good assuming you have not violated your assumptions.

C. From the statistical output shown below, the predictor "temp" has a p-value of 0.000 which is less than the level of significance of 0.05. This indicates that the regression line is statistically significant and reinforces the conclusions that we reached in part (B.).

The regression equation is

alum_ppm = - 37.9 + 1.07 temp

Predictor	**Coef**	**Stdev**	***t*-ratio**	**p**
Constant	-37.9283	0.9638	-39.35	0.000
temp	1.07397	0.01426	75.31	0.000

s = 0.6578 R-sq = 98.2% R-sq(adj) = 98.2%

Analysis of Variance:

SOURCE	**DF**	**SS**	**MS**	**F**	**p**
Regression	1	2454.1	2454.1	5672.07	0.000
Error	102	44.1	0.4		
Total	103	2498.2			

Unusual Observations:

Obs.	**temp**	**alum_ppm**	**Fit**	**Stdev.Fit**	**Residual**	**St.Resid**
14	80.4	48.5467	48.4111	0.1957	0.1356	0.22 X
78	80.3	47.5382	48.2871	0.1941	-0.7489	-1.19 X

X denotes an observation whose X value gives it large influence.

Lake Pollution

The Great Lakes provide the main water supply for many people in the United States and Canada. Additionally, they are commercially fished, provide transportation, and are a source of recreational facilities. Unfortunately, a considerable amount of waste and sewage is dumped into these lakes. This waste includes a massive amount of phosphates which have been traced to detergents, insecticides, and other chemicals, such as DDT and mercury. Extensive pollution kills off the fish, as well as other forms of animal and plant life.

Due to the immense sizes of these lakes, it is extremely difficult to locate all the causes of pollution. Pollution control cannot be performed without consideration of the economics and politics of the situation. For example, the detergent industry has spent vast amounts of money to change to biodegradable detergents, with little apparent affect on overall pollution. We must rely on natural processes to help with the clean up. Rivers usually clean themselves quickly once pollution is stopped, but large lakes are slower to become decontaminated because of the vast amount of polluted water already present. The average retention time of water in Lake Michigan is over 30 years. The average in Lake Superior is 189 years. It will take a long time to make a significant improvement to the cleanliness of the Great Lakes, even after the pollution has ceased.

It is interesting to attempt to investigate the costs involved in pollution reduction, and what might be gained as a result. You need to balance the cost of pollution against the cost of storing or dumping waste elsewhere. Such a complicated model is beyond the scope of this problem, but some of the political aspects could be discussed and taken into account.

Students need time to understand this type of problem. First, they need to appreciate that there are processes whereby pollutants are transported into, and out of, the lakes. That should lead to the assumption that the pollution in the lake is a function of time. The balancing principle (Change = Input - Output) must be introduced. Students probably need to be reminded that it may be helpful to include dimensions wherever variables are introduced.

MODELING

As part of the model construction, the important features and relationships involving the pollutant must be identified and assumptions made about them. If we ignore variations due to different types of pollutants, it seems fair to assume that the contaminants are dumped directly into the lake. It is common for quantities of pollution to be recorded in parts/unit volume, so, at times, it may be convenient to measure pollution density in terms of mass/volume. Deriving the pollution change in the lake in appropriate mathematical terms, and then converting the balance equation to a first order difference equation is likely to be demanding. The following ideas may help this process.

POSSIBLE ASSUMPTIONS:

Perfect mixing:

The lake pollution density does not depend on position in the lake.

Single inflow, single outflow:

A large number of inputs and outputs would tend to improve the validity of our perfect mixing assumption. However, unless we are interested in pollution from a particular source, it seems reasonable only to concern ourselves with the net pollution volume and mass flows, in and out.

Volume of the lake is constant:

Assuming the volume of the lake is constant clearly implies that the inflow rate is balanced by the outflow. This also means ignoring seasonal variations.

VALIDATION:

Be sure the students question themselves. Is the model a good one? Are the assumptions reasonable? Of the assumptions made, that of perfect mixing may be the least plausible.

What is the purpose of the analysis? Who is going to read and act on the report? Is ithe reader an ecologist, a politician, or a manufacturer accused of causing pollution? Within a group of students it may be constructive to give subgroups different roles to perform.

DATA ON THE GREAT LAKES SYSTEM

CHARACTERISTIC	LAKE HURON	LAKE SUPERIOR	LAKE MICHIGAN	LAKE ERIE	LAKE ONTARIO
Length (km)	331	560	490	385	309
Breadth (km)	294	256	188	91	85
Area (km)					
Water surface, US	23,600	53,618	58,016	12,898	9,324
Water surface, Canada	36,000	28,749		12,768	10,360
Drainage basin land, US	187,567	43,253	117,845	46,620	39,370
Drainage basin land, Canada	123,206	81,585		12,224	31,080
Drainage basin land, total	310,773	124,838	117,845	58,793	70,448
Drainage basin (land & water), total	370,373	207,200	175,860	87,434	90,132
Maximum depth (m)	229	406	281	60	244
Average depth (m)		148	84	17	86
Volume of water (km)		12,221	4,871	458	1,636
Mean outflow (litre/sec)		2,067,360	5,012,640	5,550,720	6,626,880
Average water retention time (yr)		189	30.8	2.6	7.8

REFERENCE:

Le Masurier, David. "Pollution of the Great Lakes," *Mathematical Modeling - A source Book of Case Studies*, Oxford University Press (1990), pp.181-193.