CASE STUDY (CS 2)

REGRESSION ANALYSIS AS AN AUDIT TOOL: A CASE STUDY

DIVISION OF FINANCIAL AND GENERAL MANAGEMENT STUDIES
UNITED STATES GENERAL ACCOUNTING OFFICE

JUNE 1972

CASE STUDY (CS 2)

REGRESSION ANALYSIS AS AN AUDIT TOOL: A CASE STUDY

DIVISION OF FINANCIAL AND GENERAL MANAGEMENT STUDIES

UNITED STATES GENERAL ACCOUNTING OFFICE

JUNE 1972

Contents

CHAPTER 1

INTRODUCTION

This report describes an auditing situation in which the Kansas City Regional Office staff used regression analysis to confirm its questioning of an agency position.

It is hoped that this report can be a useful reference to aid further use of this analytical technique.

Regression analysis is a statistical technique used to measure the extent to which a change in one quantity (variable) is accompanied by a change in some other quantity (variable). For example, increases in years of education received tend to be accompanied by increases in annual income earned. If there is evidence that change in one variable causes change in the second variable, the relationship disclosed by the regression technique can be used to estimate the magnitude of the second variable from information concerning the magnitude of the first variable.

CHAPTER 2

THE AUDIT SITUATION

The audit staff was confronted by conflicting opinions regarding the degree of urgency involved in the need for upgrading municipal sewage treatment facilities along the Missouri River. Officials of the Environmental Protection Agency favored immediate action. Many State and local officials believed that upgrading municipal sewage treatment facilities would have little impact on Missouri River water quality and that for this reason, available water pollution abatement funds should be applied to other measures which State and local officials thought offered greater water quality benefits.

Our review was primarily directed toward evaluating whether it would be to the advantage of the Federal Government to invest available funds in upgrading sewage treatment plants or to invest in other pollution abatement projects preferred by State and local officials. The review involved the main stem of the Missouri River from Gavin's Point, South Dakota, to St. Louis, Missouri.

Municipal sewage treatment

The principal purposes of sewage treatment are to keep enough dissolved oxygen in the water into which the sewage is to be discharged to support aquatic life and to prevent nuisance conditions. By removing oxygen-demanding materials from sewage prior to discharging sewage into the waterway, sewage treatment plants attempt to minimize the impact of municipal sewage on dissolved-oxygen levels in river water.

The sewage treatment plants now located along the Missouri River provide what is termed "primary" treatment. An advanced method of treatment, referred to as "secondary," is approximately twice as effective as primary treatment in removing oxygen-demanding materials and disease-carrying bacteria. The Agency favored upgrading all plants to this latter method for municipalities located along the main stem of the Missouri River.

Position of state and local officials

As indicated previously State and local officials concerned with water pollution in the area believed that constructing secondary sewage treatment plants as advocated by the Agency would have limited effect on water quality in the Missouri River. This opinion was based upon their belief that, with existing primary treatment plants, the oxygen level of the river was usually sufficiently high to support aquatic life and that, for those periods when oxygen concentrations fell below the desired level, secondary sewage treatment could not be expected to remedy the situation.

Because of its great turbulence, the Missouri River absorbs oxygen at an extremely rapid rate. It is for this reason that the amount of oxygen in the river usually is considerably above the minimum level to support aquatic life, even though the municipal sewage discharged by the communities along the river receives only primary treatment prior to discharge and thus contains a relatively large amount of oxygen-demanding material. During some periods, however, significant reductions in oxygen levels occur. These reductions occur after heavy rains or during periods when snow has rapidly melted. Oxygen is reduced during these periods because (1) large amounts of rainwater enter the sewer system, causing the municipal treatment plants to overload, therefore permitting sewage to enter the stream untreated and (2) feedlot wastes and various kinds of land runoffs enter the river.

Secondary treatment is preceded by primary treatment. If secondary sewage treatment plants were constructed, their capacities presumably would be limited to the same flow levels as those of the existing primary treatment plants. State and local officials correctly stated that the overload problem would not be solved by upgrading the treatment capability to secondary. They also believed that, since land runoff was primarily associated with rural agricultural areas, water pollution emanating from this source would not be reduced by upgrading the level of treatment of sewage originating in municipalities located along the river.

They believed available funds should be allocated to other pollution abatement efforts offering greater immediate benefits than would the construction of secondary treatment plants. Other suggested uses of funds included constructing separate sewer systems to handle rainwater, thus minimizing the likelihood of overloading sewage treatment plants during heavy rainfall, and researching into improved methods of handling feedlot waste and other agriculture-related runoff.

Supporting evidence

During its review the Kansas City Regional Office staff learned that the Agency had conducted a water quality survey of the Missouri River. The survey data tended to support the opinion of State and local officials. The survey data showed that, except during a heavy rainfall that occurred during the survey, oxygen levels were considerably above the required minimum but that during the heavy rainfall period oxygen levels were significantly reduced.

Anticipated Agency position

Although the available data seemed to support prevalent opinion regarding the limited impact of secondary sewage treatment on water quality during a period of heavy rainfall or snow melting and the adequacy of oxygen concentrations in the river during most other periods, the Agency continued to advocate construction of secondary treatment plants. We believed the Agency might, in defending its position, argue that secondary sewage treatment was needed because future growth in the municipalities located on the Missouri River might result in sewage volumes sufficient to reduce oxygen concentrations to less than acceptable levels, if only primary treatment continued to be provided. To evaluate the reasonableness of such a position, GAO sought a method of predicting the oxygen levels of the water that would be found if increased amounts of sewage, and thus oxygen-demanding materials, were to be discharged into the river.

CHAPTER 3

GAO ANALYSIS

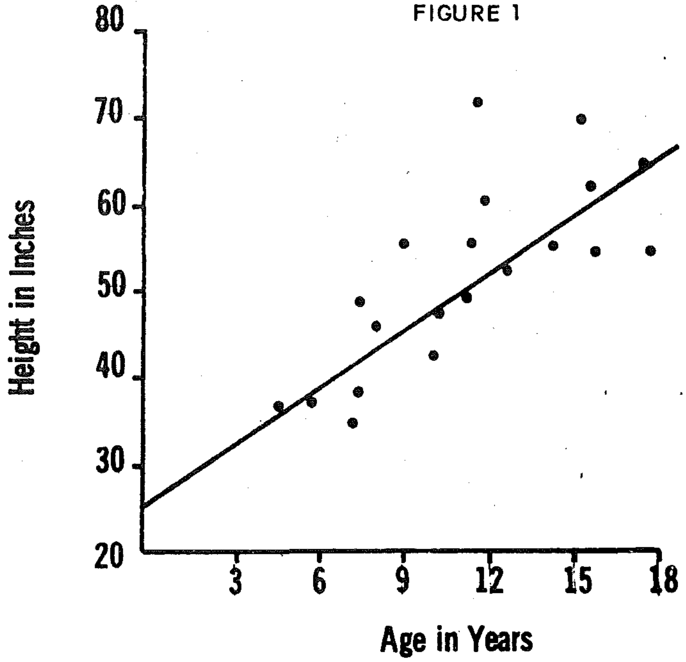
The Kansas City Regional Office staff used regression analysis to predict oxygen levels that might result from increased municipal sewage volumes along the Missouri River's main stem. As defined in the GAO glossary for systems analysis, dated October 1969, regression analysis is:

> "Analysis undertaken to determine the extent to which a change in the value of one variable, the independent variable, tends to be accompanied by a change in the value of another variable, the dependent variable. Where only one independent variable is involved in the analysis the technique is known as simple regression analysis; where two or more independent variables are involved the technique is called multiple regression analysis. ***"

If the relationship between the independent and dependent variables disclosed by the analysis is one that can be represented by a straight line, the relationship is referred to as being linear. A graphic illustration of a simple linear regression is shown in figure 1 on page 6. The plotted data points in the illustration represent the heights of girls of various ages. The straight line in the illustration represents the relationship between height (the dependent variable) and age (the independent variable) as disclosed by regression analysis.

LINEAR REGRESSION

FIGURE 1



In the case of the Missouri River, the audit staff wished to examine the relationship between the volume of oxygen-demanding material, such as that emanating from municipal sewage, in the water, and the level of oxygen in the water. GAO hoped that the statistical relationship between the two variables, as determined through regression analysis, would be one that would permit reasonably reliable predictions of expected oxygen levels for a broad range of assumed volumes of oxygen-demanding materials. The ability to make such predictions would enable the staff to assess the likelihood that future increases in urban populations, and thus in municipal sewage volumes, along the river, would necessitate the installation of secondary sewage treatment facilities to ensure acceptably high dissolved-oxygen levels in the water.

The historical data on which the regression analysis was performed was that collected by the Agency during the water survey referred to earlier in this report. This data consisted of measurements of oxygen-demanding material[1] concentrations and dissolved-oxygen concentrations gathered over a 10-day period at 21 gauging stations covering a 360-mile segment of the river from Gavin's Point Dam near Yankton, South Dakota, to St. Joseph, Missouri. Two conditions were encountered during the survey: (1) normal autumn weather and navigation flows for the first 8 sampling days of the survey and (2) 2 days of a heavy rainfall causing nearly double the normal navigation flows in the St. Joseph part of the river.

The Kansas City staff furnished this survey data to the systems analysis group in Washington so that the computer service available there could be used to perform the regression analysis. According to the terms of the contract under which it obtains access to the time-sharing computer service, GAO has the right to use a large variety of computer programs developed by the contractor. Among these programs are several that can be used to perform regression analysis. The systems analysis group used one of these computer programs to perform the regression analysis requested by the Kansas City staff.

Results

Two regression analyses were performed. The first was on the 8-day average measurements for each of the two variables taken at each point during the days of normal river flow and the second was on the 2-day average measurements taken at each point during the days of heavy rainfall. Each analysis disclosed a sufficiently close relationship between the two variables of interest that the staff was able to use the results for predictive purposes.

---

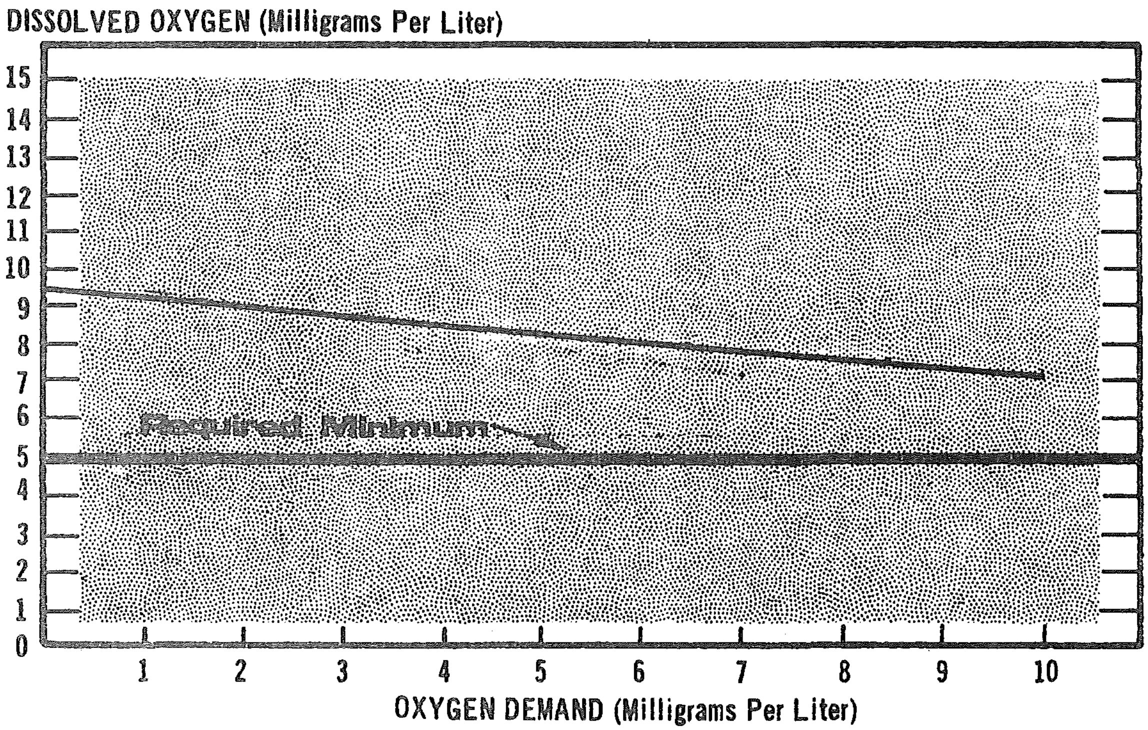
1 Actually, the measurements are in terms of oxygen demand produced by the material, rather than in terms of the material itself.

Dry weather flows

The analysis performed on the dry weather, or normal flow, measurements showed that, under these conditions, concentrations of oxygen demand far in excess of the highest found during the Agency's survey (5.8 milligrams per liter), would not reduce dissolved-oxygen levels to below the minimum of 5 milligrams per liter considered safe for the propogation of aquatic life. The linear relationship disclosed by GAO's regression analysis is represented in figure 2 below.

DRY WEATHER (8-DAY AVERAGES)

FIGURE 2

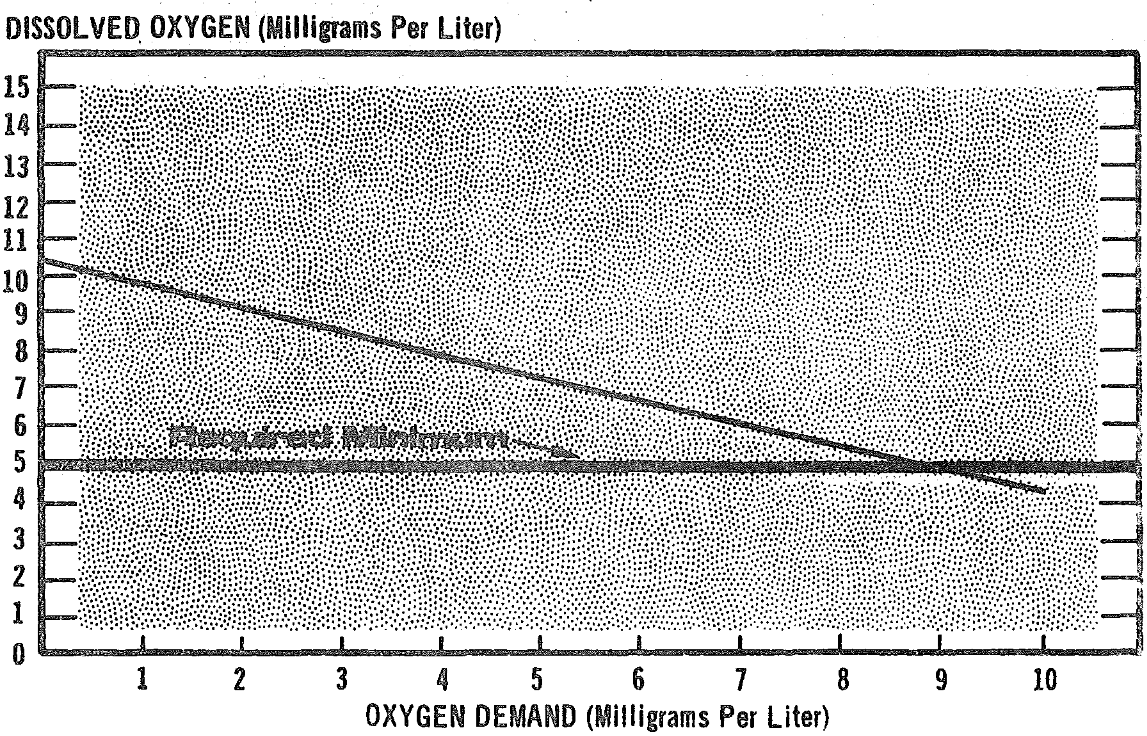


Rainy weather flows

Regression analysis on the rainy weather flow measurements, which showed significantly higher concentrations of oxygen-demanding material and lower dissolved-oxygen levels than were present in the dry weather samples, indicated that oxygen-demand concentrations 18 percent higher than the highest found in the Agency's survey (7.6 milligrams per liter) would have to occur before dissolved-oxygen levels would decrease to below the acceptable minimum. The linear relationship that the audit staff found is illustrated in figure 3 below.

RAINY WEATHER (2-DAY AVERAGES)

FIGURE 3



CHAPTER 4

USE OF THE ANALYSIS

After the staff had performed the regression analysis described in this report, the Agency offered reasons other than the traditionally cited danger of oxygen depletion for advocating the construction of secondary sewage treatment plants. The Agency then undertook a study to determine whether disease-causing bacteria or viruses emanating from human wastes were present to a significant degree in the river, thus perhaps indicating a need for improved sewage treatment.

Had the Agency chosen to raise the oxygen depletion issue, GAO would have used the results of the regression analysis to support its questioning of the merits of the Agency's position. By having these results available and by informing the Agency of them, GAO was able to focus the discussion of the question of the need for secondary sewage treatment on the possibly relevant issue of public health rather than on the apparently less relevant issue of oxygen depletion.

APPENDIX

DATA USED IN THE REGRESSION ANALYSES

Gauging station	Dry weather 8-day averages: Oxygen demand (milligrams per liter)	Dry weather 8-day averages: Dissolved oxygen (milligrams per liter)	Rainy weather 2-day averages: Oxygen demand (milligrams per liter)	Rainy weather 2-day averages: Dissolved oxygen (milligrams per liter)
1	0.9	9.5	1.2	10.0
2	1.1	9.4	1.2	9.7
3	1.5	9.3	1.6	9.4
4	1.6	9.2	1.8	9.2
5	1.4	9.2	1.8	9.0
6	1.5	9.1	4.0	8.0
7	1.9	9.1	4.9	8.0
8	2.0	8.9	4.2	7.8
9	5.8	8.8	5.2	7.6
10	3.4	8.8	5.4	7.8
11	3.4	8.6	5.6	7.5
12	3.3	8.6	6.8	6.8
13	2.8	8.5	6.7	7.0
14	3.2	8.4	7.0	6.1
15	2.6	8.6	5.8	6.6
16	2.6	8.6	5.8	6.0
17	3.2	8.3	7.7	5.2
18	3.4	8.6	5.6	5.6
19	3.0	8.6	(a)	(a)
20	3.0	8.7	(a)	(a)
21	(a)	(a)	(a)	(a)

[a] Data unavailable.