

# **APPLYING THE TOOLS**

## **EXAMPLE THREE**

- 1. Problem Statement**
- 2. Analyst Statement**
- 3. Solution**

## NOTES TO STUDENTS:

A. Some of the regression analyses appear in the heart of the statistical report. When you run the data provided for regression your output will contain 'tones' of results and only some are presented to "DAVE".

B. Many people believe that statistical analysis is heavy-duty number crunching and little else. As many of our former students have told us, however, this is definitely not true. They continually testify to the importance of written reports (and oral presentations) in their jobs. In fact, we believe that many of you will be judged more by the quality of your *writing* (and speaking) than by the quality of your quantitative analysis. Therefore, keep the suggestions and examples ~~in this chapter~~ <sup>handy</sup> -- you might need them more than you realize. Just remember that well-designed studies and careful statistical analysis are often useless unless they are communicated clearly and effectively to the audience that needs them.

### EXAMPLE 3

#### **Problem Statement**

I am a statistical consultant, and I have been hired by the Bendrix Company, a manufacturing company, to analyze its overhead data (attached). The company has supplied me with historical monthly data from the past three years on overhead expenses, machine hours, and the number of production runs. My task is to develop a method for forecasting overhead expenses in the future months, given estimates of the machine hours and number of production runs that are expected in these months. My contact, Dave Clements, is in the company's finance department. He obtained an MBA degree about 10 years ago, and he vaguely remembers some of the statistics he learned at that time. However, he does not profess to be an expert. The more I can write my report in layman's terms. The more he will appreciate it.

The Bendrix Company manufactures various types of parts for automobiles. The manager of the factory wants to get a better understanding of overhead costs. These overhead costs include supervision, indirect labor, supplies, payroll taxes, overtime premiums, depreciation, and a number of miscellaneous items such as insurance, utilities, and janitorial and maintenance expenses. Some of these overhead costs are "fixed" in the sense that they do not vary appreciably with the volume of work being done, whereas others are "variable" and do vary directly with the volume of work. The fixed overhead costs tend to come from the supervision, depreciation, and miscellaneous, categories, whereas the variable overhead costs tend to come from the indirect labor, supplies, payroll taxes, and overtime premiums categories. However, it is not easy to draw a clear line between the fixed and variable overhead components.

The Bendrix manager has tracked total overhead costs over the past 36 months. To help "explain" these, he has also collected data on two variables that are related to the amount of work done at the factory. These variables are:

- MachHrs: number of machine hours used during the month
- ProdRuns: the number of separate production runs during the month

The first of these is a direct measure of the amount of work being done. To understand the second, we note that Bendrix manufactures parts in fairly large batches. Each batch corresponds to a production run. Once a production run is completed, the factory must "set up" for the next production run. During this setup there is typically some downtime while the machinery is

reconfigured for the part type scheduled for production in the next batch. Therefore, the manager believes both of these variables might be responsible (in different ways) for variations in overhead costs. Do scatterplots support this belief?

**Figure 1 Data for Bendrix Overhead Example**

Monthly data on manufacturing overhead costs			
Months	MachHrs	ProdRuns	Overhead
1	1539	31	99798
2	1284	29	87804
3	1490	27	93681
4	1355	22	82262
5	1500	35	106968
6	1777	30	107925
7	1716	41	117287
8	1045	29	76868
9	1364	47	106001
10	1516	21	88738
34	1723	35	107828
35	1413	30	88032
36	1390	54	117943

## Solution

The data appear in Figure 1. Each observation (row) corresponds to a single month. We want to investigate any possible relationship between the Overhead variable and the MachHrs and ProdRuns variables, but because these are time series variables and the Month variable. That is, we should investigate any time series behavior in these variables.

This data set illustrates, even with a modest number of variables, how the number of potentially useful scatterplots can grow quickly. At the very least, we should examine the scatterplot between each potential explanatory variable (MachHrs and ProdRuns) and the response variable (Overhead). These appear in Figure 2 and 3. We see that Overhead tends to increase as either MachHrs increases or ProdRuns increases. However, both relationships are far from perfect.

To check for possible time series patterns, we can also create a time series plot for any of the variables. (Actually, this is equivalent to a scatterplot of the variable versus Month, with the points jointed by lines.) One of these, the time series plot for Overhead, appears in Figure 4. It shows a fairly random pattern through time, with no apparent upward trend or other obvious time series pattern. You can check that time series plots of the MachHrs and ProdRuns variables also indicate no obvious time series patterns.

**Figure 2 Scatterplot of Overhead versus Machines Hours**

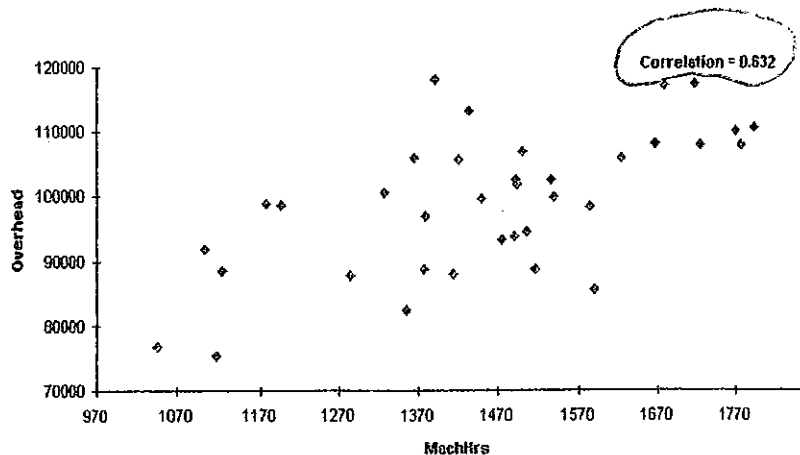


Figure 3 Scatterplot of Overhead versus Production Runs

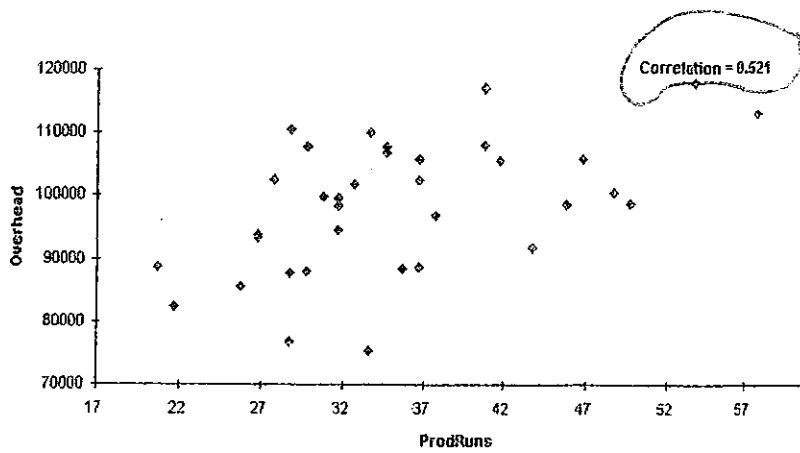
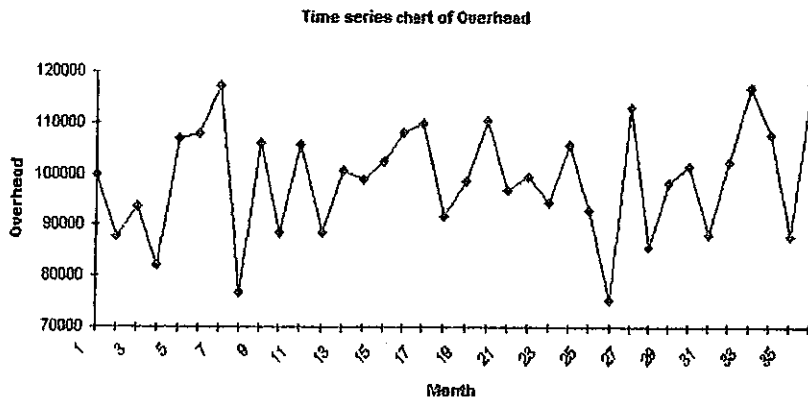
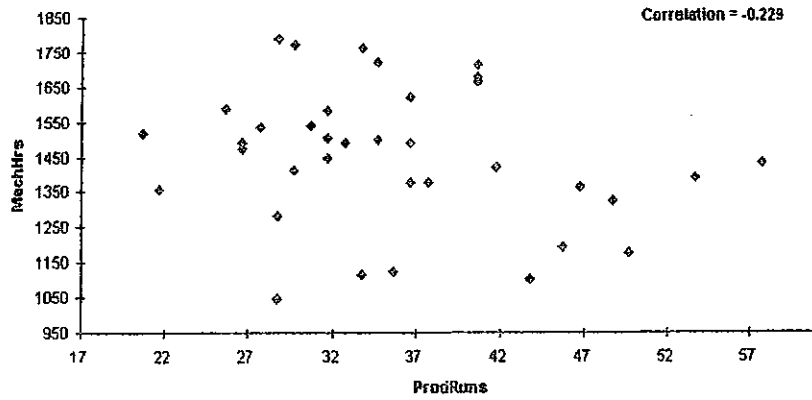


Figure 4 Time Series Plot of Overhead versus Month



Finally, when there are multiple explanatory variables, we can check for relationships among them. The scatterplot of MachHrs versus ProdRuns appears in Figure 5 (Either variable could be chosen for the vertical axis.) This “cloud” of points indicates no relationship worth pursuing.

**Figure 5 Scatterplot of Machine Hours versus Production Runs**



In summary, the Bendrix manager should continue to explore the positive relationship between Overhead and each of the MachHrs and ProdRuns variables. However, none of the variables appears to have any time series behavior, and the two potential explanatory variables do not appear to be related to each other.