**Forecasting for Business and using Linear Regression**

**Why is Forecasting important?**

The future is uncertain. Some events do have a very small probability of happening, like an asteroid destroying the earth. So we accept that tomorrow will come as a certain event. But future demand for a business’s goods and services is very uncertain. Yet, the management of a company wants to have some idea of the survival (or growth) of the company in the future. Should they expect to hire more people or let some go? Should they plan to increase capacity? How much investment is needed for future assets, or should they down size?

Forecasting provides some ideas about the future, but how this is accomplished can vary from company to company. And one key factor is how accurate the forecast is. Generally, the further into the future one looks, the more uncertain the information is. How do forecasters reduce their forecasting errors? How much error is tolerable?

Another key factor in forecasting is data availability. Data processing and storage capability have become extremely available and inexpensive. Software and computing power is also very cheap. Collecting real-time sales data via point-of-sales systems is now common at most retail establishments. But couple this with a situation in companies that have a large number of products, such as a retail store or a large manufacturing company with hundreds or thousands of product numbers and/or product lines and forecasting becomes complicated.

**Forecasting methods**

There are two main types or genres of forecasting methods, qualitative and quantitative. The former consists of judgment and analysis of qualitative factors, such as scenario building and scenario analysis. The latter is obviously based on numerical analysis and that is the focus of this course. This genre of forecasting includes such methods as time series analysis, causal methods, and combination methods. This module will look at a causal method, linear regression. Module 4 focuses on time series analysis using exponential smoothing.

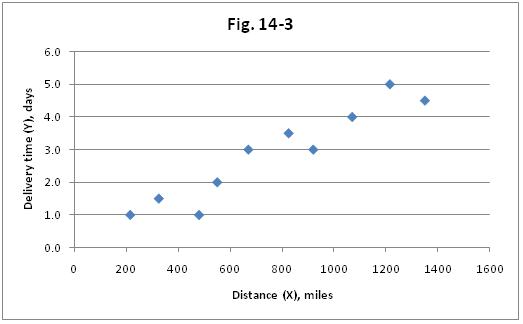
**Linear Regression**

Let’s start with an example. As discussed in the Home page, we are considering a linear relationship between shipment distance and delivery time. Table 1 and Figure 1 show the Data for the shipping distance and delivery situation and the Scatter Plot.

Table 1: Data for shipping distance and delivery situation

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sampled Shipment | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Distance (X) in miles | 825 | 215 | 1070 | 550 | 480 | 920 | 1350 | 325 | 670 | 1215 |
| Delivery time (Y) in days | 3.5 | 1.0 | 4.0 | 2.0 | 1.0 | 3.0 | 4.5 | 1.5 | 3.0 | 5.0 |

Figure 1: Scatter Plot of Shipment Distance and Delivery Time



Since the highway distance is to be used as the independent variable, this selection of trips of specific distances is acceptable. On the other hand, the dependent variable of delivery time is a random variable in this study, which conforms to the assumption underlying regression analysis. Whether or not the two variables have a linear relationship would generally be investigated initially by constructing a scatter plot, as is shown above. Such diagrams also are used to observe whether the vertical scatter (variance) is about equal along the regression line.

Note that there are a few pitfalls when using Linear Regression.

1. In regression analysis a value of Y cannot be legitimately estimated if the value of X is outside the range of values that served as the basis for the regression equation.
2. If the estimate of Y involves the prediction of a result that has not yet occurred, the historical data that served as the basis for the regression equation may not be relevant for future events.
3. The use of a prediction or a confidence interval is based on the assumption that the conditional distributions of Y, and thus of the residuals, are normal and have equal variances.
4. For both regression and correlation analysis, a linear model is assumed. For a relationship that is curvilinear, a transformation to achieve linearity may be available. Another possibility is to restrict the analysis to the range of values within which the relationship is essentially linear.

To learn the details of doing a linear regression analysis, read the specified resources and watch the videos. Also, download the Excel file with Examples and a practice problem.