As the Quality Manager for Excellent Manufacturing Company you have received the Quarterly Production Report. This shows that the latest 13 weeks of the production output and hours worked in each department. The Assembly Department Manager would like to know if there is a way that the data can be used to create a prediction for the number of hours to work given a specific amount of output needed for a week. You have decided to use linear regression to see if there is a relationship between the weekly output and the amount of hours worked.

|  |  |  |
| --- | --- | --- |
| Week | Output | Hours Worked |
| 1 | 908 | 380 |
| 2 | 986 | 373 |
| 3 | 1056 | 403 |
| 4 | 855 | 351 |
| 5 | 1046 | 425 |
| 6 | 1072 | 410 |
| 7 | 983 | 383 |
| 8 | 896 | 380 |
| 9 | 1035 | 410 |
| 10 | 992 | 367 |
| 11 | 1023 | 374 |
| 12 | 1115 | 443 |
| 13 | 909 | 352 |

**Case Assignment:** Using Excel:

a. Enter the data  
b. Generate a scatter plot  
c. Calculate the regression equation using the formula approach  
d. Add the equation to the graph using the "add equation" option and include the equation formula. Verify that the equation in the plot is the same as that you calculated.  
e. Estimate a prediction for the number of hours required to work if the desired output for the week is 1000.  
f. Determine the prediction interval with a 5% alpha.

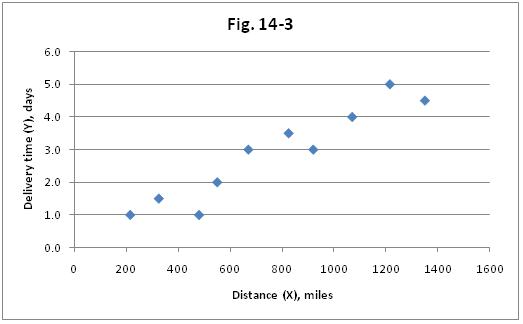
**Assignment Expectations:** Do the activities as described above. Write a two to three page paper explaining what you did and interpret the results of the regression and the prediction. Include references if you use any. Submit the paper as the case and upload the Excel file into Additional **files.**

**REQUIRED READING:**

The Data for the shipping distance and delivery situation are shown below along with the Scatter Plot.

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

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 by constructing a scatter plot (see Section 14.2) or a residual plot (see Section 14.4). Such diagrams also are used to observe whether the vertical scatter (variance) is about equal along the regression line.



**REQUIRED READING**

The assigned reading is Schaum's Outline of Business Statistics, 4th Edition, Ch. 14,

Sections 14.1 - 14.8, and Chapter Problems 14.1 to 14.11 at the end of the chapter. This text is available on ebrary; Kazmier, Leonard J. *Schaum's Outline of Business Statistics, 4th Edition.* Blacklick, OH, USA: McGraw-Hill Trade, 2003. p 271. Retrieved from 11/14/09 <http://site.ebrary.com/lib/touro/Doc?id=10051516&ppg=289>.

Read the text section by section and try to do the sample problems.   
[Download the Excel file](http://cdad.tuiu.edu/Presentation.aspx?course=975&term=94&presentation=20456) and review the problem solutions in Excel.

14.1 - the objectives of Linear Regression and how to determine the dependent and independent variables

14.2 - the Scatter plot and how to visually see the relationship between the independent and dependent variables. See Problem 14.1.

14.3 - explains the mathematics of the method of least squares for fitting a regression line. See Problem 14.2

14.4 - Residuals and Residual Plots explains how to use residuals. See Problem 14.3

14.5 - the Standard Error of Estimate is an important statistic in testing hypotheses regarding goodness of fit. It is similar to the Std Error of Estimate from ANOVA. See Problems 14.5 and 14.6.

14.6 - Inferences concerning the slope of the linear regression line discusses the tests for the estimated slope of the regression line. See Problems 14.7, 14.8, and 14.9.

14.7 - Confidence Intervals for the Conditional Mean - **This section is Optional**

14.8 - Prediction Intervals for Individual estimates of the dependent variable discusses the probability interval of a prediction when using a regression equation. See Problem 14.11.

There are some pitfalls to consider when using regression analysis. These are discussed in section 14.14 along with pitfalls of correlation (Module 3). Those pitfalls are shown here.

(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 con?dence interval is based on the assumption that the conditional distributions of Y, and thus of the residuals, are normal and have equal variances.

 (7) 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.

(Kazmier, Leonard J. *Schaum.* Blacklick, OH, USA: McGraw-Hill Trade, 2003. p 271. Retrieved from <http://site.ebrary.com/lib/touro/Doc?id=10051516&ppg=289>. **Copyright © 2003.  McGraw-Hill Trade.  All rights reserved.)**