Spreadsheet Case 7

Lightning Communications Outlet

Problem:	Determine the relationship between rebates and level of sales
Management skills:	Controlling Planning
PC skills:	Regression analysis Graphics
File:	REBATE_Q.XLS

Lightning Communications Outlet is a distributor of radios, calculators, paging devices, and telephones in the upper Midwest. Telephone sales have recently been falling, so senior management decided to experiment with using cash rebates to stimulate sales. To test the effectiveness of rebates, Lightning Communications offered different rebate amounts for phones retailing for $99.00 in each of its 16 retail outlets. The results were so varied that Robert Whittaker, the firm's president, is not sure that the rebates were effective in stimulating sales.

Larry Gorman, Lightning Communication's director of sales and marketing, believes otherwise. He is convinced that even though sales fluctuated at Lightning's various stores, the sales increases were not random. The greater the rebate, the higher the sales increase. Gorman needs a way of showing that the amount of cash is directly related to the percentage of increase in sales.

Load the data file REBATE_Q.XLS from your data diskette. This file shows the data on the amount of rebate and the corresponding increase in sales in 16 stores. Gorman wants to use this data to determine the relationship between rebates and increase in sales to help management determine whether rebates should be used to increase sales in the future. He believes the results will be more convincing if they are displayed graphically.

Gorman can do this by performing a regression analysis, which is a statistical method for measuring the relationship between two or more variables. If the relationship is only between two variables, the method is called simple regression. If the relationship is between more than two variables, the method is called multiple regression. This case has been simplified for instructional purposes so that only two variables, amount of rebate, and percent increase in sales, will be analyzed.

A regression analysis results in an equation that describes the behavior of one dependent variable in terms of other variables, called independent variables. In this case, there is only one independent variable, amount of rebate, and one dependent variable, percent increase in sales. The regression analysis also produces statistics that measure the strength of the relationship between the independent and dependent variables. Regression analysis can be visualized as a way of drawing the "best line" through a series of data points.

You can then use the regression equation to forecast the impact on sales of various rebates. Some popular business applications for regression analysis are determining the

relationship between a product's price and cost of production or determining the relationship between incidence of defects in manufactured goods and production volume.

Tasks

There are 4 tasks in this case:

1. Carefully examine the template for the data file, REBATE_Q.XLS which you have just loaded. Note that percentages are entered and displayed as 3.2 rather than .032 in order to appear appropriately in the graph you will construct later. Use the regression analysis commands of your spreadsheet software to perform a simple regression analysis. Select the amount of rebate range as the independent variable and the percent increase in sales range as the dependent variable. Specify for your Output Range an unused area of your worksheet because the output will be written over any existing cell contents.

2. Use the results of the regression analysis to construct a Regression Line. The Regression Line can be calculated by multiplying the value of the X Coefficient by the each value of the independent variable (amount of rebate) and then adding the value of the constant.

3. In the regression output area, R squared values range from 0 to 1. An R squared of 0 means the relationship between the independent and dependent variables is entirely random, whereas an R squared of 1 indicates that the dependent variable moves in a lockstep relationship with the independent variable. The X Coefficient tells how much the dependent variable changes for each unit of change in the independent variable. The standard error of the X Coefficient indicates how precise the relationship is between independent and dependent variables. If the error is great, the X Coefficient isn't a good predictor of the value of the dependent variable, but if the standard error is small, the X Coefficient is a good predictor. To judge whether the error is great or small, you must add a formula for the t-statistic. The t-statistic can be calculated by dividing the X Coefficient by the standard error of the coefficient. A t-statistic of 2 means that one can be 95% certain that the X-coefficient is an accurate predictor for the behavior of the dependent variable. A lower t-statistic indicates that the independent variable may not be an accurate predictor.

Print out the worksheet and the Regression Output. In a few sentences state whether you believe that increasing rebates will increase sales and why.

4. To demonstrate the relationship between rebates and sales visually, construct an XY type graph showing the independent variable (amount of rebate) on the X-axis, the dependent variable (percent increase in sales) on the Y-axis, and the Regression Line. Use the X data range for the independent variable, the A data range for the dependent variable and the B data range for the Regression Line. Add appropriate legends and graph titles. Print out this graph to show to management.

Additional Problem

Management wants to be able to experiment with various rebate amounts and determine the effect on sales. In other words, if management asks how much of sales increase can be expected if all stores offer a $10 rebate, they should be able to receive a solid answer. Hint: Create an output area where you can display various rebate amounts and the corresponding increase in sales for each amount. You will need to use the formula you developed to calculate the Regression Line. This is necessary because even though there may be one data point for a $10 rebate, it is an isolated observation. One can be much more confident in predicting the relationship between amount of rebate and sales increases by using the regression formula.

Time Estimates

Expert: 45 minutes
Intermediate: 1.5 hours
Novice: 2.5 hours

Excel Tutorial For Spreadsheet Case 7

This case expands on graphics and other spreadsheet skills acquired previously and introduces the use of a Excel command for regression analysis.

You can use your sample student roster (COURSE.XLS) if you expand it to include a column for the students' ages. (If you saved COURSE.XLS with the formula to place asterisks in column F, you can erase range F15:F18 and then enter student age data in this column. You should also erase the range name table.)

The professor and school administration want to see if there is any correlation between student age and maturity and academic performance. Let's assign the age of 25 to James Jackson, 22 to Andrew Reynolds, 23 to Steven Parker, and 19 to Joyce Winters. You can place ages in corresponding cells in range F15:F18. Enter the column heading AGE in cell F14. You will need to widen column E to 12 positions. You may want to reduce the widths of columns B, D and F so that you can display columns A-H on your worksheet screen.

The number of students on this worksheet is actually too small a sample to be statistically valid in real life, but it will illustrate the concept of regression analysis and the Excel **Regression** Analysis Tool. Excel provides a variety of Analysis Tools that can be accessed through **Tools/Data Analysis**.

If **Data Analysis** does not appear in the **Tools** menu, the Analysis ToolPak add-in has not been loaded. An add-in is a file which provides additional functions, commands and menus. Select **Tools/Add-Ins**. In the Add-Ins dialog box, choose Analysis ToolPak and select the OK Button. Now select **Tools/Data Analysis**.

Now choose Regression from the Data Analysis dialog box by moving down the list using the scroll bar, highlighting "Regression"and pressing OK. To determine the relationship

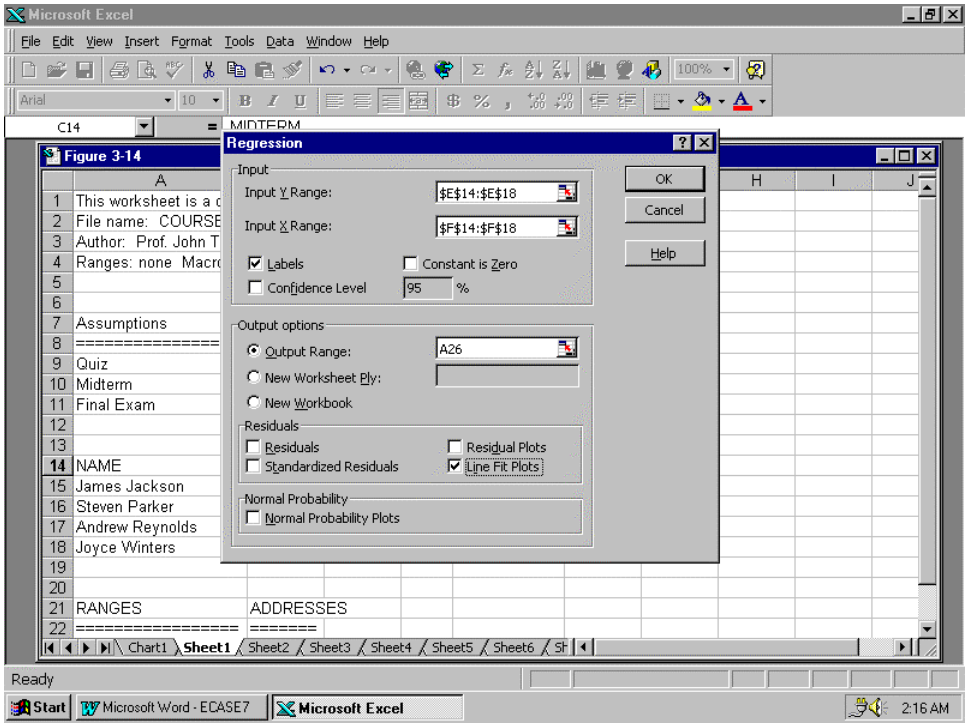
between students' ages and final grades, you must make age the independent variable and final grade the dependent variable.

The Regression dialog box (see Figure 3-8) requires that you specify the ranges that holds the dependent values and the independent values.

At the **Input Y-range** setting specify your dependent variable. In the window provided enter the range E14:E18 (the range for final grades). This dialog box permits you to select ranges on the worksheet. Select the Final Grades range using your mouse, or type the range E14:E18.

At the **Input X-Range** setting specify one or more independent variables. In the window provided enter the range F14:F18 (the range for students' ages). As with the Y-range, ranges on the worksheet can be selected with the mouse. Select the Students' Ages range or type the range F14:F18.

Figure 3-8



Check the **Labels** Check Box to indicate the selected ranges include the Labels in the first row. Another input setting available is the **Constant is zero** Check Box which lets the user force the Y intercept to zero. We will not use this setting for this problem. The **Confidence Level** Check Box can be selected if you want confidence intervals applied to the regression in addition to the default 95% levels.

The remaining settings concern the output options of the regression analysis. Three output options are available: firstly, **Output Range** which deposits the output on the same sheet and asks for the upper-left cell reference of the output range; secondly, **New Worksheet Ply** which creates a new worksheet in the same workbook and (optionally) asks you to name the new worksheet; and thirdly, **New Workbook** which creates an entirely new workbook and places the output in cell A1 of its first worksheet. For our purposes, select **Output Range** and place the cell reference A26 in the available window. Generally it is wise to allocate a worksheet area that is clear.

Excel provides four options for residual values: **Residuals**, **Standardized Residuals**, **Residual Plots**, and **Line Fit Plots**. Residuals are the differences between the actual values and predicted values using the regression coefficients for the same dependent values. The plots will be embedded charts in the worksheet where the output tables will appear. Choose the Line Fit Plots setting to achieve a comparison between the actual values and the predicted values.

The final option is for **Normal Probability Plots**. Select the OK Button to start the regression calculations.

The output tables (see Figure 3-9) contain: Regression Statistics, ANalysis Of VAriance (ANOVA) table, and Regression Coefficients. If selected, Residual Tables are also delivered. The Regression Coefficients are the values which permit you to construct the Regression line. For example, the Y-intercept of the regression you created is 40.53 and the gradient is 1.92. If you placed the following formula in cell G15, you would receive the predicted Student Grade for the age in cell F15:

```
=$B$42+F15*$B$43
```

A set of values to construct a line plot of the predicted values could then be produced. The chart could be created in a way described in the Tutorial for Spreadsheet Case 5. Excel's Regression Tool creates such a chart for you. If you selected the Line Fit Plots Check Box in the Regression dialog box, you would have a similar chart to that in Figure 3-10. We have simply adjusted the X-Axis scale to amplify the line.

You may save your worksheet, but the changes you made to COURSE.XLS for the regression analysis will not be utilized by subsequent cases.

Figure 3-9

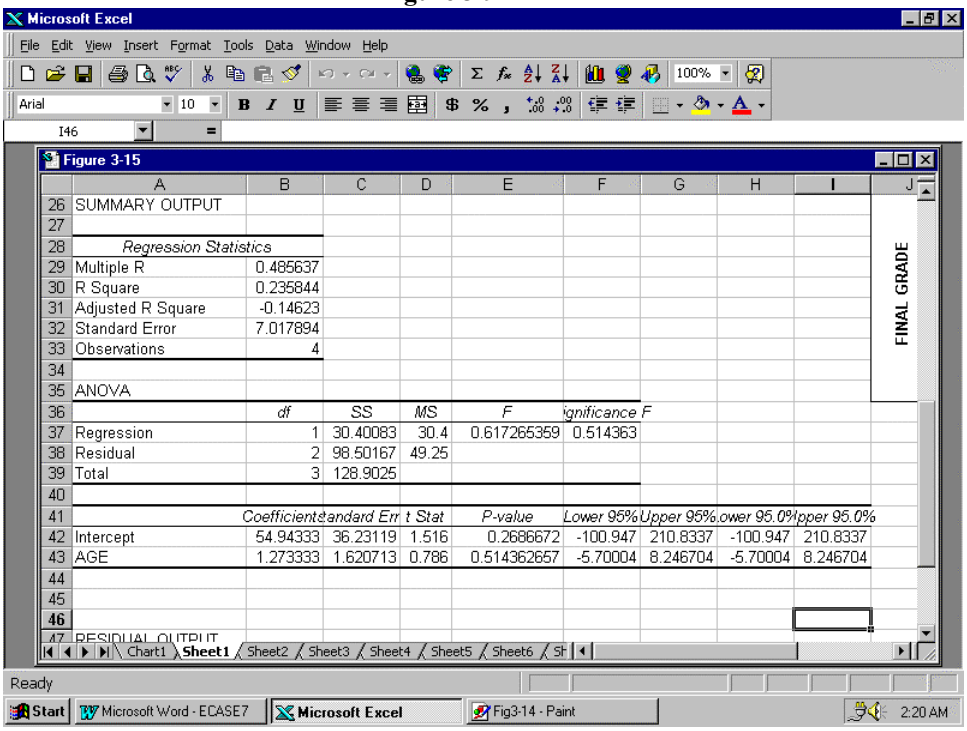


Figure 3-10