Spreadsheet Case 7

Precision Cartridges

Problem:	Determine a relationship between defects and production volume
Management skills:	Controlling Planning
PC skills:	Regression analysis Graphics

File: Precis_q.xls

Precision Cartridges is a small manufacturer of cartridges for computer laser printers in San Jose, California. Precision has been able to take advantage of a growing market and its business has expanded dramatically. Since opening twelve years ago, the company has doubled its workforce and more than tripled production.

Precision's customers are computer supply companies who sell the cartridges to owners of laser printers. Many defective cartridges go undetected until people try to use their printers. Precision's customers feel that any incidence of defects threatens their reputation and credibility.

A zero defect rate is an impossibility, but one below one per thousand is acceptable. Unfortunately, Precision Cartridges's defect rate has been steadily rising. George Freeman, Precision's director of production, wonders the rise in defects is related to the company's spurt of growth and production volume.

To meet production demand, Precision Cartridges had to double its labor force. Its recent recruits tend to be very young, fresh from high school, and largely untrained. Precision's internal training programs have been minimal because many of its positions, such as wrappers and loaders, do not involve high technology or special skills. Freeman feels these younger workers do not pay as close attention to quality control procedures, nor do they operate equipment as carefully as veteran workers. There are also issues of scale and complexity: many more workers and work groups to supervise.

Freeman is under strict orders from senior management to "do something" to curb the rise in disk defects. Precision Cartridges's targets are to produce 1,000,000 cartridges by 2001 and 1,300,000 by 2002. Freeman wants to institute a stronger training program and quality controls but needs some data to convince senior management that if unchecked, the incidence of defects will worsen. In addition to predicting future defect levels, Freeman would also like to use this data as a base line to gauge the success of his quality control problems.

Load the data file Precis_q.xls from your data diskette. It shows Precision's production volume and defects per 1000 cartridges from 1991 to 2000. This data will be used to determine the historical relationship between incidence of defects and production volume which will help Precision Cartridges predict the level of defects in the future.

You 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, incidence of defects, and production volume, 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, production volume, and one dependent variable, incidence of defects. 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 future incidence of defects given projected production volumes. Some popular business applications for regression analysis are determining the relationship between a product's price and cost of production or determining the level of sales that will be generated by an advertising or sales promotion campaign.

Tasks

There are 4 tasks in this case:

1. Carefully examine the template for the data file, Precis_q.xls, which you have just loaded. Use the regression analysis commands of your spreadsheet software to perform a simple regression analysis. Select the production volume range as the independent variable (X -Range) and the level of defects range as the dependent variable (Y-Range). 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 (production volume) and then adding the value of the constant. Print out the worksheet and the Regression Output.

3. Construct an XY (Scatter) type chart showing the independent variable (production volume) on the X-axis, the dependent variable (defects per 1000) on the Y-axis, and the Regression Line. (Be sure to delete the row with the dotted lines under the column labels before you start creating your chart.) Add appropriate legends and graph titles. Print out this graph.

4. Extend the independent variable range with values of 1,000,000 for 2001 and 1,300,000 for 2002. Then extend the regression line to predict future levels of defects. Revise your graph to incorporate this data and print again.

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 may need to widen column E. 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.

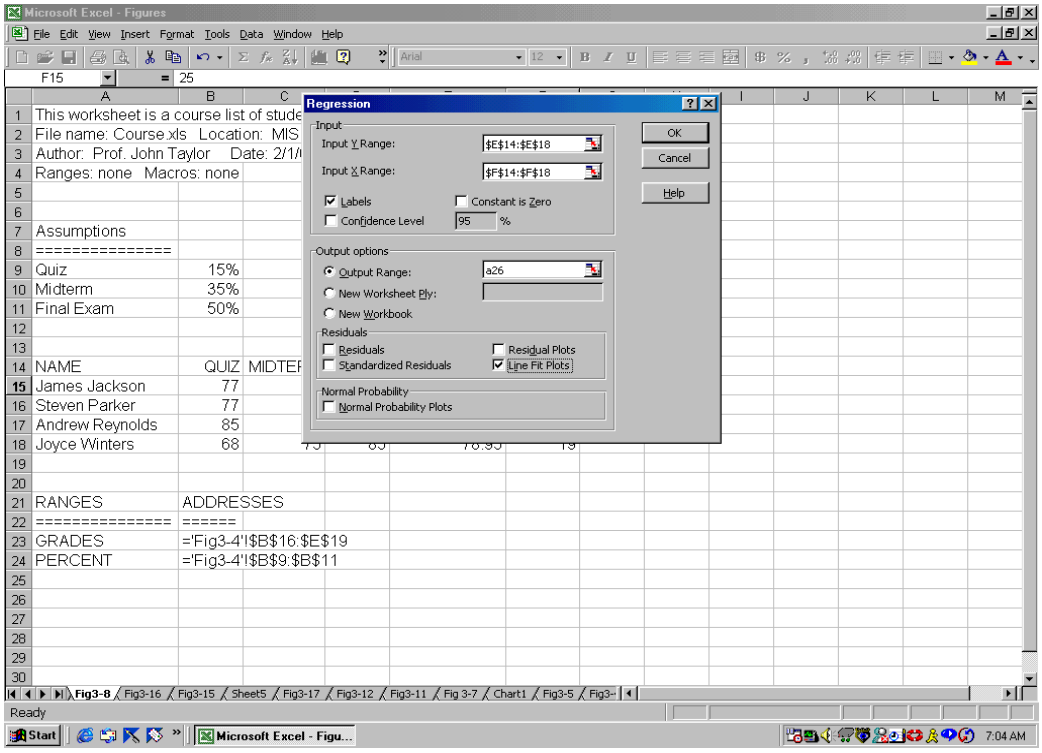
At the **Input X-Range** setting specify one or more independent variables. In the window provided enter the range F14:F18 (the range for students' ages). As with the Y-range, ranges on the worksheet can be selected with the mouse. Select the Students' Ages range or type the range F14:F18.

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.

Figure 3-8



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

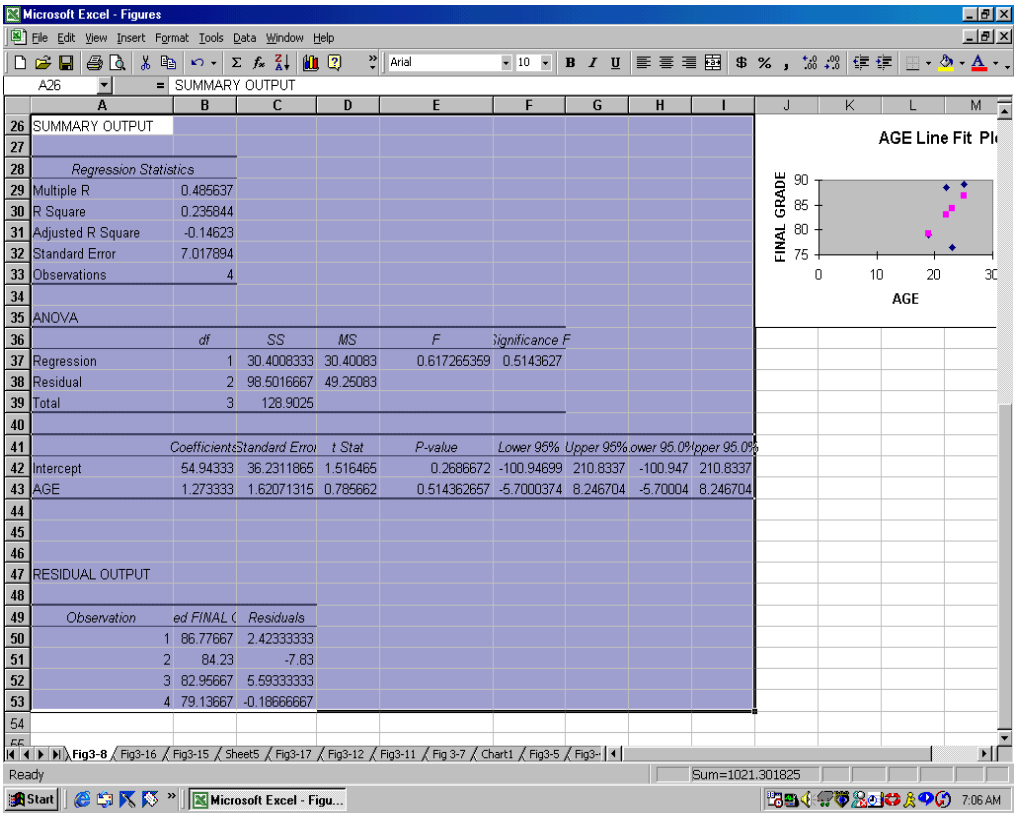
delivered. The Regression Coefficients are the values which permit you to construct the Regression line. For example, the Y-intercept of the regression you created is 54.94 and the gradient is 1.273. If you placed the following formula in cell G15, you would receive the predicted Student Grade for the age in cell F15:

=B42+F15*B43

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.

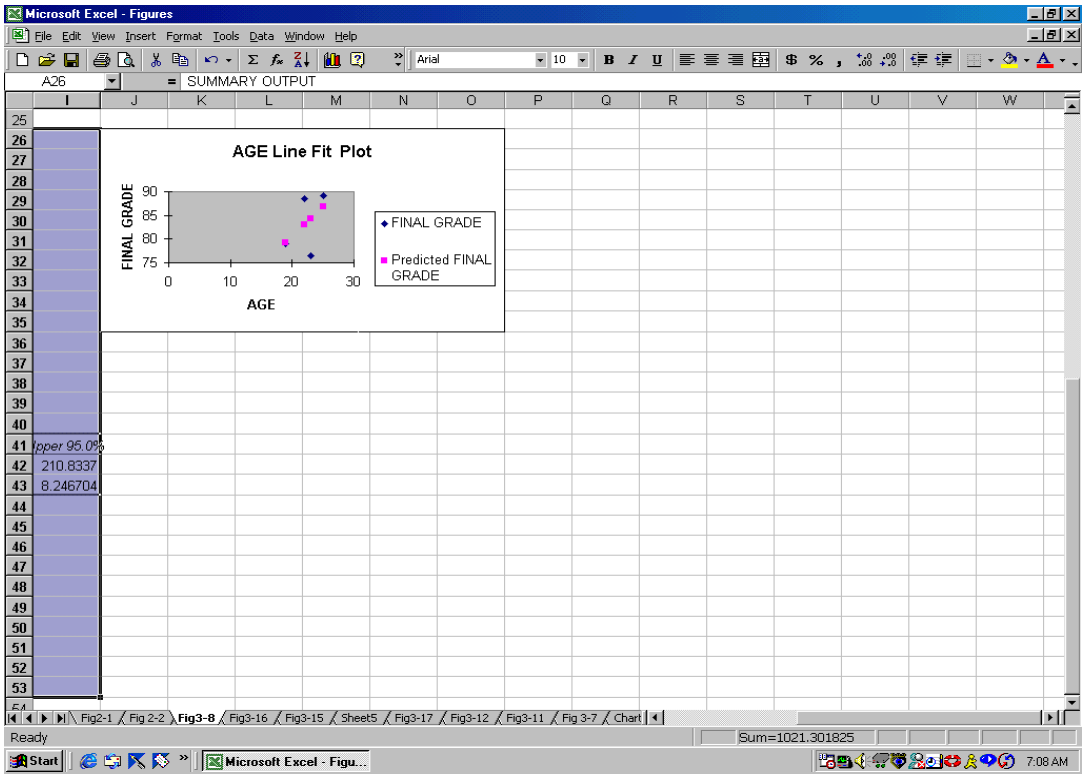
You may save your worksheet, but the changes you made to Course.xls for the regression analysis will not be utilized by subsequent cases.

Figure 3-9



	A	B	C	D	E	F	G	H	I
26	SUMMARY OUTPUT								
27									
28	*Regression Statistics*								
29	Multiple R	0.485637							
30	R Square	0.235844							
31	Adjusted R Square	-0.14623							
32	Standard Error	7.017894							
33	Observations	4							
34									
35	ANOVA								
36		*df*	*SS*	*MS*	*F*	*Significance F*			
37	Regression	1	30.4008333	30.40083	0.617265359	0.5143627			
38	Residual	2	98.5016667	49.25083					
39	Total	3	128.9025						
40									
41		*Coefficients*	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
42	Intercept	54.94333	36.2311865	1.516465	0.2686672	-100.94699	210.8337	-100.947	210.8337
43	AGE	1.273333	1.62071315	0.785662	0.514362657	-5.7000374	8.246704	-5.70004	8.246704
44									
45									
46									
47	RESIDUAL OUTPUT								
48									
49	*Observation*	*ed FINAL G*	*Residuals*						
50	1	86.77667	2.42333333						
51	2	84.23	-7.83						
52	3	82.95667	5.59333333						
53	4	79.13667	-0.18666667						
54									

Figure 3-10


Microsoft Excel - Figures
File Edit View Insert Format Tools Data Window Help
A26
SUMMARY OUTPUT
AGE Line Fit Plot
FINAL GRADE
90
85
80
75
0
10
20
30
AGE
FINAL GRADE
Predicted FINAL GRADE
Upper 95.0%
210.8337
8.246704
Fig2-1 Fig 2-2 Fig3-8 Fig3-16 Fig3-15 Sheet5 Fig3-17 Fig3-12 Fig3-11 Fig 3-7 Chart
Ready
Sum=1021.301825
Start
Microsoft Excel - Figu...
7:08 AM