# CHAPTER 6: COST ESTIMATION

## QUESTIONS

**6-1** Cost estimation is the process of developing a well-defined relationship between a cost object and its cost driver for the purpose of predicting the cost. The cost predictions are used in each of the management functions:
Strategic Management: Cost estimation is used to predict costs of alternative activities, predict financial impacts of alternative strategic choices, and to predict the costs of alternative implementation strategies.
Planning and Decision Making: Cost estimation is used to predict costs so that management can determine the desirability of alternative options and to budget expenditures, profits, and cash flows.
Management and Operational Control: Cost estimation is used to develop cost standards, as a basis for evaluating performance.
Product and Service Costing: Cost estimation is used to allocate costs to products and services or to charge users for jointly incurred costs.

**6-2** The assumptions used in cost estimation are:
a. Linear functions can estimate cost behavior within a relevant range
b. Other assumptions are specific to the estimation method chosen, for example, the assumptions of regression are covered in Appendix B.

**6-3** The three methods of cost estimation are:
a. High-Low. Because of the precision in the development of the equation, it provides a more consistent estimate than the visual fit and is not difficult to use. Disadvantages: uses only two selected data points and is, therefore, subjective.
b. Work Measurement. The advantage is accurate estimates through detailed study of the different operations in the production process, but like regression, it is more complex.
c. Regression. Quantitative, objective measures of the precision and reliability of the model are the advantages of this model; disadvantages are its complexity: the effort, expense, and expertise necessary to utilize this method.

**6-4** Implementation problems with cost estimation include:
a. Cost estimates outside of the relevant range may not be reliable.
b. Sufficient and reliable data may not be available.
c. Cost drivers may not be matched to dependent variables properly in each observation.
d. The length of the time period for each observation may be too long, so that the underlying relationships between the cost driver and the variable to be estimated is difficult to isolate from the numerous variables and events occurring in that period of time; alternatively the period may be too short, so that the data is likely to be affected by accounting errors in which transactions are not properly posted in the period in which they occurred.
e. Dependent variables and cost drivers may be affected by trend or seasonality.

f. When extreme observations (outliers) are used the reliability of the results will be diminished.
g. When there is a shift in the data, as, for example, a new product is introduced or when there is a work stoppage, the data will be unreliable for future estimates.

**6-5** The six steps in cost estimation are as follows:
a. Define the cost to be estimated.
b. Determine the cost drivers.
c. Collect consistent and accurate data.
d. Graph the data.
e. Select and employ the appropriate estimation method.
f. Assess the accuracy of the cost estimate.

Choosing the cost drivers is often the most important step since the model's accuracy is based upon selecting the relevant and appropriate cost drivers.

**6-6** The contrast between regression analysis and high-low analysis is as follows:
a. Regression analysis estimates the cost function by using a statistical model that relates the average change in the dependent variables to a unit change in the cost driver(s).
b. The high-low method estimates the cost function by determining the line that connects the highest and lowest values for the cost driver in the relevant range.

**6-7** Cost estimation methods could be used to help identify activity cost drivers in activity-based costing. For example, if a firm using activity-based costing is looking for an appropriate cost driver for materials handling, regression can be used to determine if the best cost driver is number of parts in the product, the weight of parts in the product or some other measure of materials-handling activity. The R-squared of the regression would provide a useful means to determine which of the cost drivers is a better fit.

**6-8** The dependent variable is the cost object of interest in the cost estimation. An important issue in selecting a dependent variable is the level of aggregation in the variable. For example, the company, plant, or department may all be possible levels of data for the cost object. The choice of aggregation level depends on the objectives for the cost estimation, data availability, reliability, and cost/benefit considerations. If a key objective is accuracy, then a detailed level of analysis is often preferred. The detail cost estimates can then be aggregated if desired.
To identify the independent variables, management accountants consider all the financial and operating data that might be relevant for estimating the dependent variable and choose a subset of those that both appear to be the most relevant and do not duplicate other independent variables.

**6-9** Nonlinear cost relationships are cost relationships that are not adequately explained by a single linear relationship for the cost driver. In accounting data, a common type of nonlinear relationship is trend and seasonality. For a trend example, if sales increase by 8% each year, the plot of the data for sales will not be linear with the driver, the number of years. Similarly, sales which fluctuate according to a seasonal pattern will have a nonlinear behavior. A different type of nonlinearity is where the cost driver and the dependent variable have an inherently nonlinear relationship. For example, payroll costs as a dependent variable estimated by hours worked and wage rates is nonlinear, since the relationship is multiplicative and therefore not the additive linear model assumed in regression analysis.

**6-10** The advantages of using regression analysis include that it:
a. Provides an estimation model with best fit (least squared error) to the data
b. Provides measures of goodness of fit and of the reliability of the model which can be used to assess the usefulness of the specific model, in contrast to the other estimation methods which provide no means of self-evaluation
c. Can incorporate multiple independent variables
d. Can be adapted to handle non-linear relationships in the data, including trends, shifts and other discontinuities, seasonality, etc.
e. Results in a model that is unique for a given set of data

**6-11** A dummy variable is an independent variable assigned the value of 0 or 1 in the regression analysis. It can improve the accuracy of the regression analysis if it is used to represent special conditions unique to a data point, such as seasonal factors or changes in production technology.

**6-12** High correlation exists when the changes in two variables occur together. It is a measure of the degree of association between the two variables. Because correlation is determined from a sample of values, there is no assurance that it measures or describes a cause and effect relationship between the variables.

**6-13** The coefficient of determination (R-squared) measures the degree to which changes in the dependent variable can be predicted by changes in the independent variable(s).

## BRIEF EXERCISES

**6-14** $\frac{\$25,830 - \$18,414}{3,495 - 1,958} = \frac{\$7,416}{1,537} = \$4.82$ per hour

**6-15** The r-squared statistic indicates the degree to which changes in the dependent variable can be predicted by changes in the independent variable. The t-value indicates the statistical reliability of each independent variable. Based on this information, Carter Dry Cleaning should choose Regression B as it has the better R-squared and t-values.

**6-16**

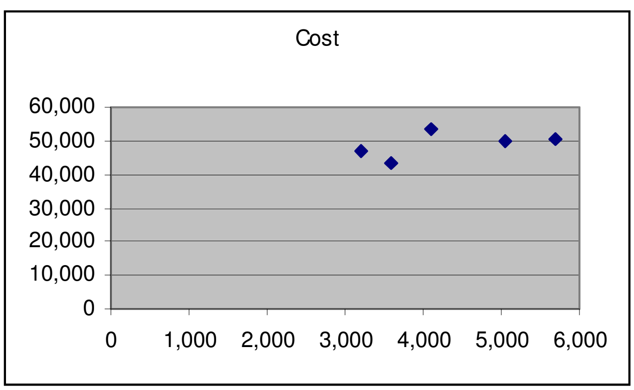
a = Y - (b x X)
a = $10,000 - ($1 x 7,000)
a = $3,000
OR: a= $3,000 = $6,000 – $1 x 3,000

**6-17**

Possible independent variables for analysis of financial data include wage rate, sales, and units produced.

**6-18**

Smith should use year 2003 for the high point and 2004 for the low point, by inspection of the data on hours. Note that the high and low point using cost would be different, 2007 as low and 2005 as high. However, a graph of the data (admittedly not available to the students in a brief exercise) would make it more clear that the best choice would be 2007 and 2008.



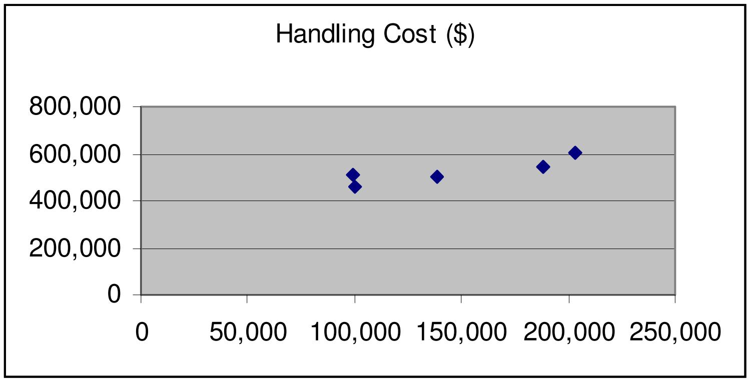

**6-19**

$5,000 ± $400 provides the 67% confidence interval of $4,600 - $5,400. The r-squared and t-values provided are extraneous information and should be disregarded.

**6-20** $\frac{\$20{,}000 - \$10{,}000}{3{,}000{,}000 - 2{,}500{,}000} = \frac{\$10{,}000}{500{,}000} = \$0.02$ per key

**6-21**

The clear choice for the high point is 2006, but the low point is more difficult to determine; One should pick 2003 even though 2005 has lower hours, because, from the graph, the 2003 point will produce a more representative estimation line




**6-22** Total Cost = 200,000 x \$35 + \$125,000
= \$7,000,000 + \$125,000
= \$7,125,000

**6-23**

The r-squared value of .6 tells you that changes in the independent variable do not predict changes in the dependent variable very well. The t-value of 2.3 indicates there is a strong relationship between the independent and dependent variables. The standard error of \$200,000 on a predicted total cost of \$2,584,072 indicates relatively small variability in predicted data.

**6-24** Variable Cost = $\frac{\$400{,}000 - \$250{,}000}{8{,}000 - 5{,}000} = \frac{\$150{,}000}{3{,}000} = \$50$ per hour

**6-25** a = \$80,000 - (120,000 x \$2)
= \$80,000 - \$240,000
a = \$ -160,000
OR \$40,000 – (100,000 x \$2) = \$ -160,000

## EXERCISES

**6-26**

1. b
2. f
3. e
4. i
5. e
6. h
7. l
8. a
9. j
10. k
11. c or d
12. g

**6-27 Cost Relationships (10 min)**

Expected production:

.80 x 250 = 200 computers per month (P)

The cost function equation is:

y = a + b x P

y = $62,250 + $150 x (production)

Total cost is:

= $ 62,250 + $150 x 200

= $92,250

Average cost per computer is:

= $92,250 / 200

= $461.25

**6-28 Cost Relationships (15min)**

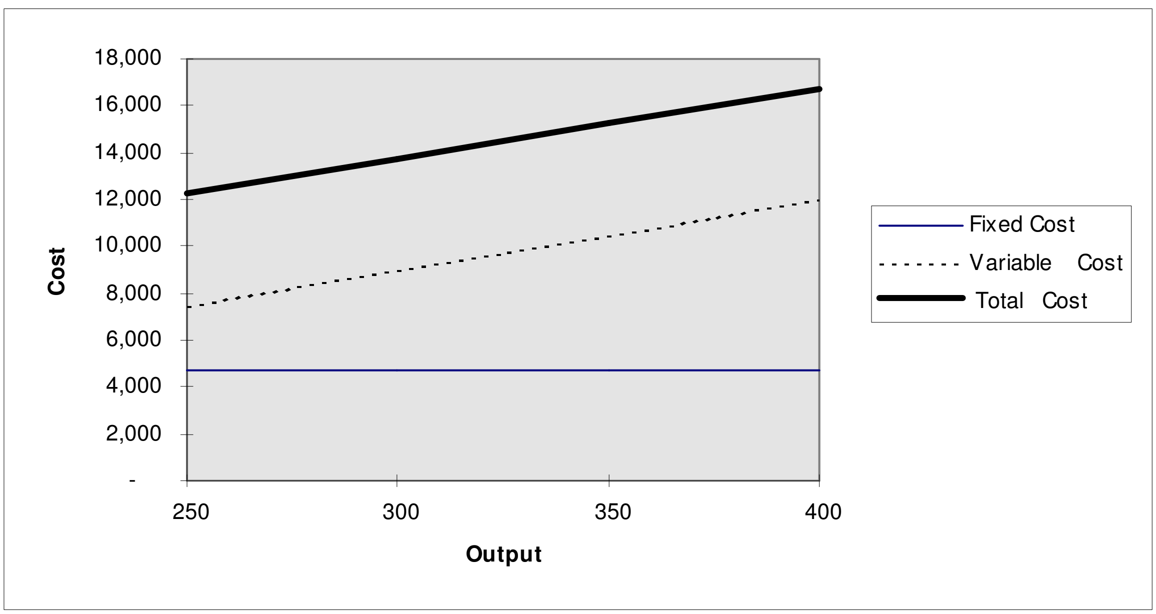
1. Total costs: See Exhibit below

| Output | Total Costs | Total Costs Per Unit |
|---|---|---|
| 250 | $12,250 | $49.00 |
| 300 | $13,750 | $45.83 |
| 350 | $15,250 | $43.57 |
| 400 | $16,750 | $41.88 |

Total variable costs:

| Output | Total Variable Costs |
|---|---|
| 250 | $ 7,500 |
| 300 | $ 9,000 |
| 350 | $10,500 |
| 400 | $12,000 |

Total fixed costs = $4,750



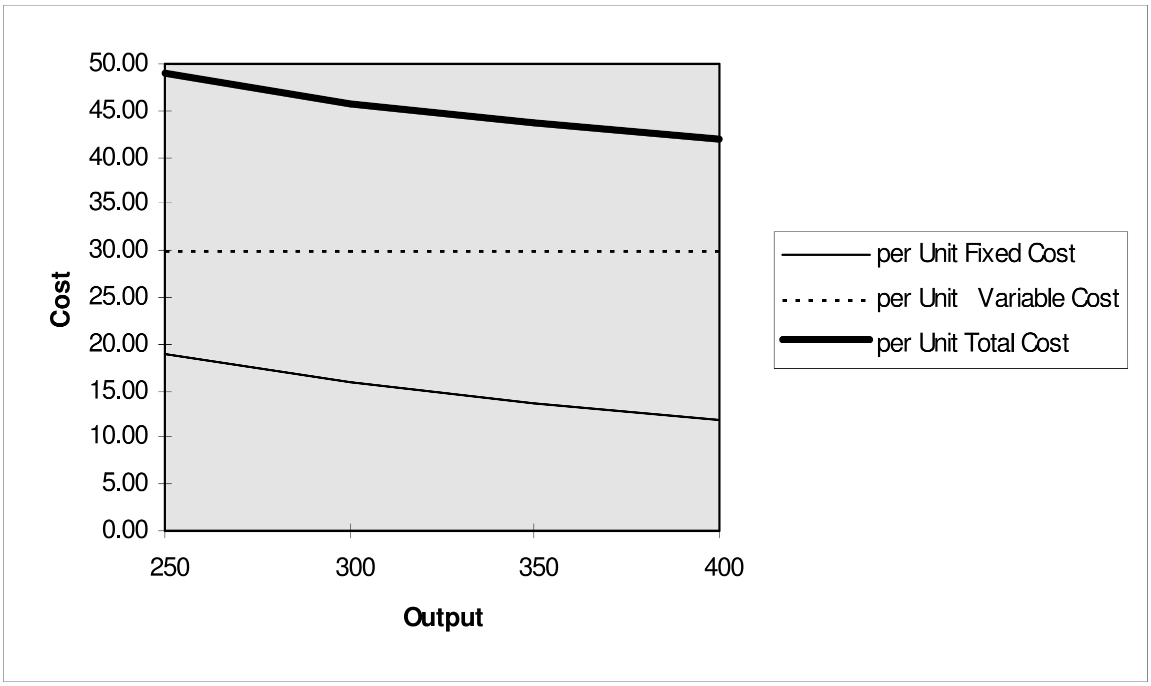

2. Per-unit total cost: See Exhibit below
Per unit variable costs:

$7,500 / 250 = $30

**Exercise 6-28 (continued)**

Per-unit fixed costs:

| Output | Per-Unit Fixed Costs | Total Fixed Costs |
|---|---|---|
| 250 | $19.00 | $4,750 |
| 300 | 15.83 | 4,750 |
| 350 | 13.57 | 4,750 |
| 400 | 11.88 | 4,750 |




3. The important point of these graphs is that total fixed costs are constant while unit fixed costs change as output changes. In contrast, unit variable costs are constant and total variable costs change as output changes.

## 6-29 Cost Estimation; Average Cost (15 min)

Compute total cost for each batch:

| Units | x | Average Cost | = | Total Cost |
|---|---|---|---|---|
| 500 | x | $0.55 | = | $275 |
| 600 | x | $0.50 | = | $300 |

Use high-low analysis and compute cost function:

Slope (b) = $\frac{\$300 - \$275}{600 - 500}$ = $0.25 / unit

Constant (a)   a + $0.25 x (500) = $275
a = $275 - $125
= $150
and/or   a + $0.25 x (600) = $300
= $150

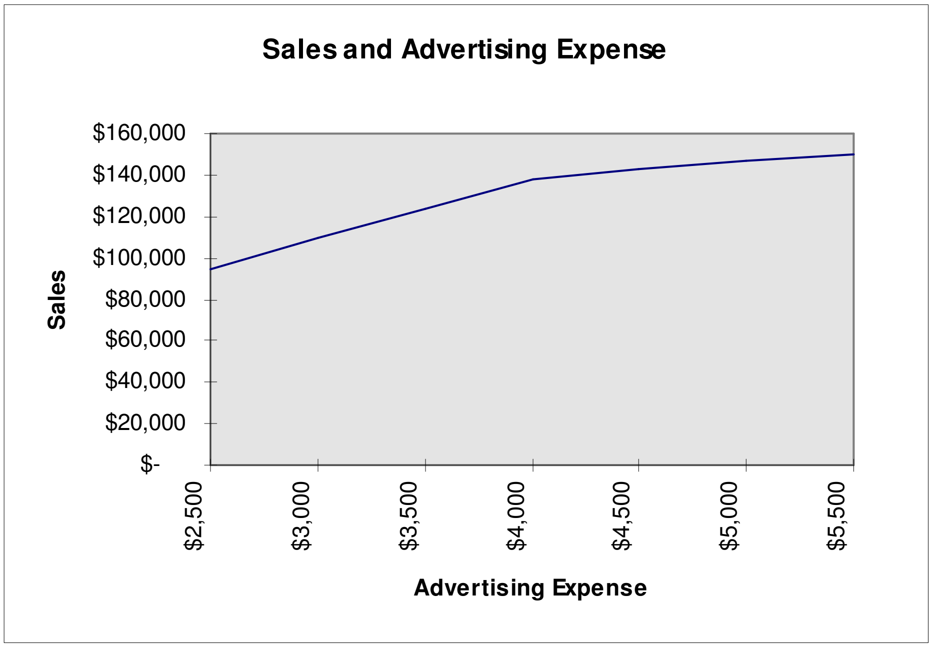
The cost function is:
y = a + bx
y = $150 + $0.25 x (croissants produced)

For 560 croissants:
y = $150 + $0.25 x (560)
= $290 = total costs
Average cost = $290/560 = $.5179

## 6-30 Cost Estimation Using Graphs (15 min)

1.




2. There seems to be a positive linear relationship for the data between $2,500 and $4,000 of advertising expense. Lawson's analysis is correct within this relevant range but not outside of it. Notice that the relationship between advertising expense and sales changes at $4,000 of expense.

## 6-31 Analysis of Regression Results (10 min)

Regression one uses only labor hours, regression two uses only machine hours, and regression three uses them both.

Regression 2 is clearly inferior as it has the lowest R-squared, the highest SE, and an unsatisfactory t-value

Regressions 1 and 2 have comparable SE and R-squared values, though regression 3 is marginally better. Note however, that the t-values show that, in regression 3, labor hours is marginally satisfactory (not quite 2) and machine hours still has an unsatisfactory t. This finding for the t-values in regression three likely indicates that the two variables, labor hours and machine hours, are highly correlated, and the result of combining them shows in the decline of the t-values for both. On the other hand, the addition of another variable to a regression usually improves R-squared and SE (there is more opportunity to explain total variance, because there are more variables available), but in this case R-squared and SE increase only marginally. If the focus for Wang is purely on estimation, then either regression one or three will work, but if there is a plan to use the coefficients of the two variables to approximate unit labor costs or unit machine time costs, then regression three should not be used – the poor t-values indicate that the relationships developed in the regression for these two independent variables are not statistically significant (the t-values are less than 2), and moreover, there is evidence of multicollinearity between these two variables. The regression 1 coefficient for labor hours could be used for approximating unit labor costs, since it has a satisfactory t, but the machine hours variable does not have a significant t (in regressions 2 or 3) and its coefficients cannot be interpreted in this way.

Note also that the overall values for R-squared are relatively low, so that limited confidence should be placed in any of the three.

**6-32 Cost Estimation: High-Low method (15 min)**

Model to fit: Maintenance Expense = a + b x M (machine hours)

The highest and lowest points are months 6 and 10, respectively. Note that the point for month 12 is an outlier, and should not be used as the low point; see the graph below.

The high-low method is as follows:

Change in Total Maintenance Expense = \$3,005 - \$2,570 = \$435

Change in Total Machine Hours = 1,880 - 1,410 = 470

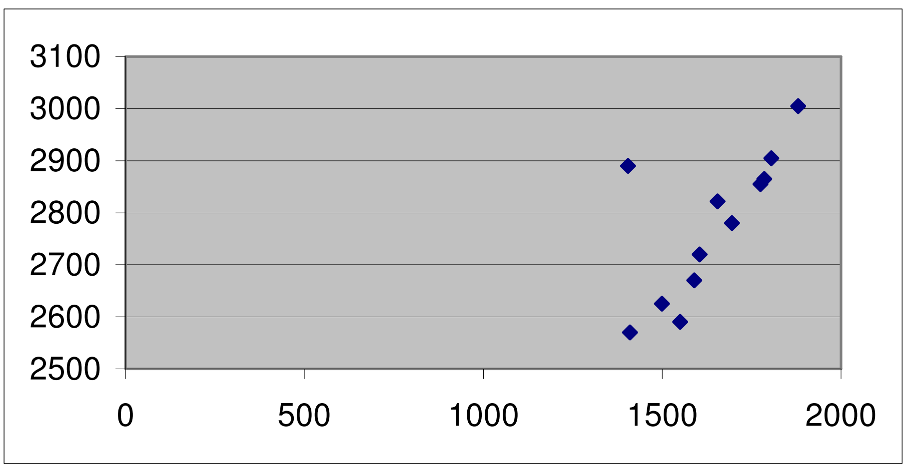
Slope (b) = \$435/470 = \$.9255

Constant (a) = \$3,005 - \$.9255 x (1,880) = \$1,265
and/or = \$2,570- \$.9255 x (1,410) = \$1,265

The equation for maintenance cost is:

Maintenance Costs = \$1,265 + \$.9255 x M (machine hours)

The graph below shows that the selected high and low points are representative of the data, but there is one significant outlier, the point for month 12.



**6-33 Cost Estimation: High-Low method (15 min)**

Model to fit: Maintenance Expense = a + b x M (machine hours)

The highest and lowest points are months 5 and 10, respectively.

the high-low method is as follows:

Change in Total Maintenance Expense = \$3,100 - \$2,220 = \$880
Change in Total Machine Hours = 1,900 - 1,100 = 800

Slope (b) = \$880/800 = \$1.10

Constant (a) = \$3,100 - \$1.10 x (1,900) = \$1,010
and/or
= \$2,220 - \$1.10 x (1,100) = \$1,010

Maintenance Costs = \$1,010 + \$1.10 x M (machine hours)

Note that an alternative solution might be preferred. On the basis of a view of a graph of machine hours versus maintenance expense (see below), it appears that the chosen lowest data point (month 10) is not as representative of the relationships in the data as for month 11 (1,300 hours; \$2,230). The point for month 10 is far to the left of the remaining data points, while the point for month 11 is somewhat closer to the remaining data points. A recalculation of the high-low method with month 11 would reveal:

Change in Total Maintenance Expense = \$3,100 - \$2,230 = \$870
Change in Total Machine Hours = 1,900 - 1,300 = 600

Slope (b) = \$870/600 = \$1.45

Constant (a) = \$3,100 - \$1.45x (1,900) = \$345
and/or
= \$2,230 - \$1.45 x (1,300) = \$345

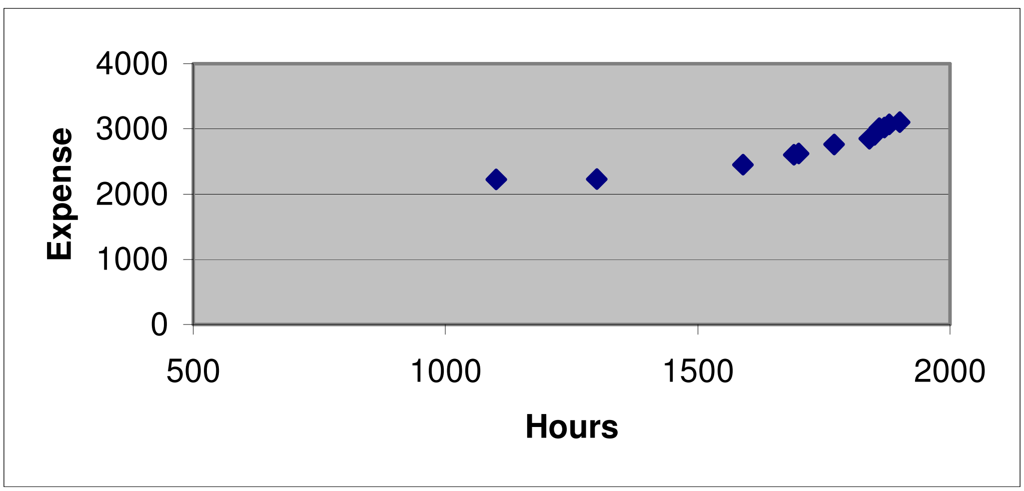
Maintenance Costs = \$345 + \$1.45 x M (machine hours)

**Exercise 6-33 (continued)**

The graph of expense versus hours shows the point for month 10 to be an outlier (to the far left of the graph). One might also argue that the point for month 11 is also an outlier and that the data for month 12 (1,590 hours and $2,450) should be used instead. The model using month 12 as the lowest month would be:

Expense = - $883 + $2.097 x Hours



**6-34 Interpreting Regression Results (10 min)**

1. The estimated cost is:
   $3,719 + 2 x $861 + 1 x $1,986 + 1 x $908 = $8,335

2. There are two dummy variables in this regression:
   - presence of one or more complications.
   - use of a laparoscope (or not)

3. The model has a relatively low r squared of only 53%, but all three independent variables have good t-values (>2.0). Looking at the t-values, it appears that the strongest independent variable is the length of stay, and the weakest is the use of laparoscope.

The exercise is based on information from: "Hospital Costs of Uterine Artery Embolization…" by M Beinfeld, J. Bosch, and G Gazette, *Academic Radiology*, Nov 9, No. 11, November 2002, pp 1300-1304.

**6-35 Analysis of Regression Results (10 min)**

1. The laparoscopic treatment has the better regression result, with a significantly higher R-squared and lower standard error for the number of complications variable.

2. The coefficients tell us the additional cost for an incremement in each of the independent variables. For example, the laparascopic treatment itself costs $908, and each day in the hospital costs $861 in the laparoscipic treatment.
The standard error of each independent variable is used to evaluate the statistical reliability of the independent variable. The t-value is the ratio of the coefficient to the standard error of the independent variable.

3. The t-value is the ratio of the coefficient to the standard error of the independent variable.

| | Not Laparoscopic | t-value | Laparoscopic | t-value |
|---|---|---|---|---|
| Coefficients for Independent Variables | | | | |
| Intercept | $ 8,043 | | $ 3,719 | |
| Length of Stay | | | | |
| Coefficient | Not significant | | 861 | |
| Standard error for the coefficient | Not applicable | | 80 | |
| Number of Complications | | | | |
| Coefficient | 3,393 | | 1,986 | |
| Standard error for the coefficient | 1,239 | **2.74** | 406 | **4.89** |
| Laparascopic | | | | |
| Coefficient | Not applicable | | 908 | |
| Standard error for the coefficient | Not applicable | | 358 | **2.54** |
| R-squared | 0.11 | | 0.53 | |

## PROBLEMS

### 6-36 Cost Estimation; High-Low Method (20 min)

1. Cost equation using square feet as the cost driver:

Variable costs:

$$\frac{\$4{,}700 - \$2{,}800}{4{,}050 - 2{,}375} = \$1.134$$

Fixed costs:

$4,700 = Fixed Cost + $1.134 x 4,050

Fixed Cost = $107

Equation One: Total Cost = $107 + $1.134 x square feet

There are two choices for the High-Low points when using openings for the cost driver. At 11 openings there is a cost of $2,800 and at 10 openings there is a cost of $2,875.

Cost equation using 11 openings as the cost driver:

Variable costs:

$$\frac{\$4{,}700 - \$2{,}800}{19 - 11} = \$237.50$$

Fixed costs:

$4,700 = Fixed Cost + $237.50 x 19

= $187.50

Equation Two: Total Cost = $187.50 + $237.50 x openings

Cost equation using 10 openings as the cost driver:

Variable costs:

$$\frac{\$4{,}700 - \$2{,}875}{19 - 10} = \$202.78$$

Fixed costs:

$4,700 = Fixed Cost + $202.78 x 19

= $847.18

Equation Three: Total Cost = $847.18 + $202.78 x openings

**Problem 6-36 (continued)**

Predicted total cost for a 3,300 square foot house with 14 openings using equation one:

$107 + $1.134 x 3,300 = $3,849.20

Predicted total cost for a 3,300 square foot house with 14 openings using equation two:

$187.50 + $237.50 x 14 = $3,512.50

Predicted total cost for a 3,300 square foot house with 14 openings using equation three:

$847.18 + $202.78 x 14 = $3,686.10

There is no simple method to determine which prediction is best when using the High-Low method. In contrast, regression provides quantitative measures (R-squared, standard error, t-values,...) to help assess which regression equation is best.

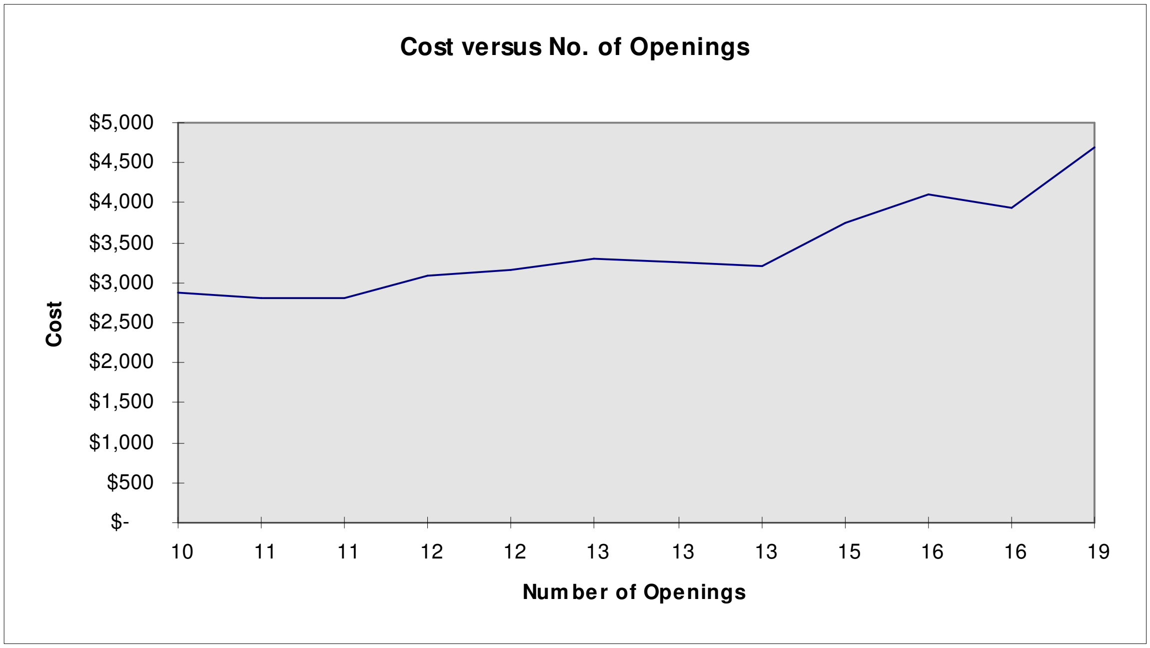
Predicted cost for a 2,400 square foot house with 8 openings, using equation one:

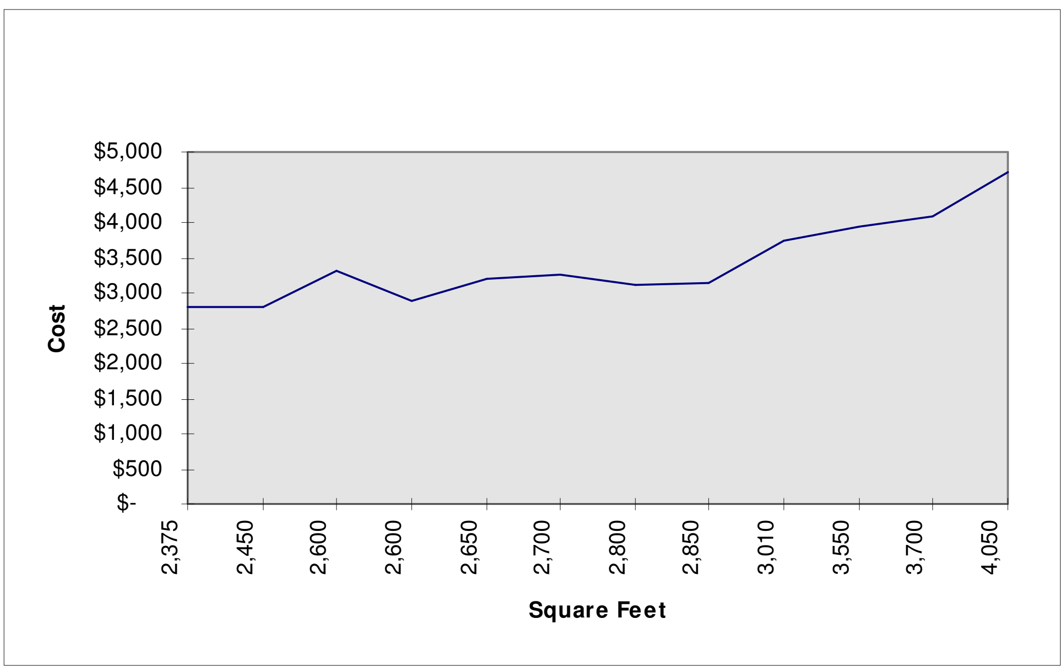
$107 + $1.134 x 2,400 = $2,828.60

We cannot predict with equation 2 or equation 3 since 8 openings are outside the relevant range, the range for which the high-low equation was developed.

2. See accompanying graphs, which show that the relationship between costs and square feet is relatively linear without outliers, as is the relationship between costs and number of openings. From this perspective, both variables are good cost drivers.

## Problem 6-36 (continued)

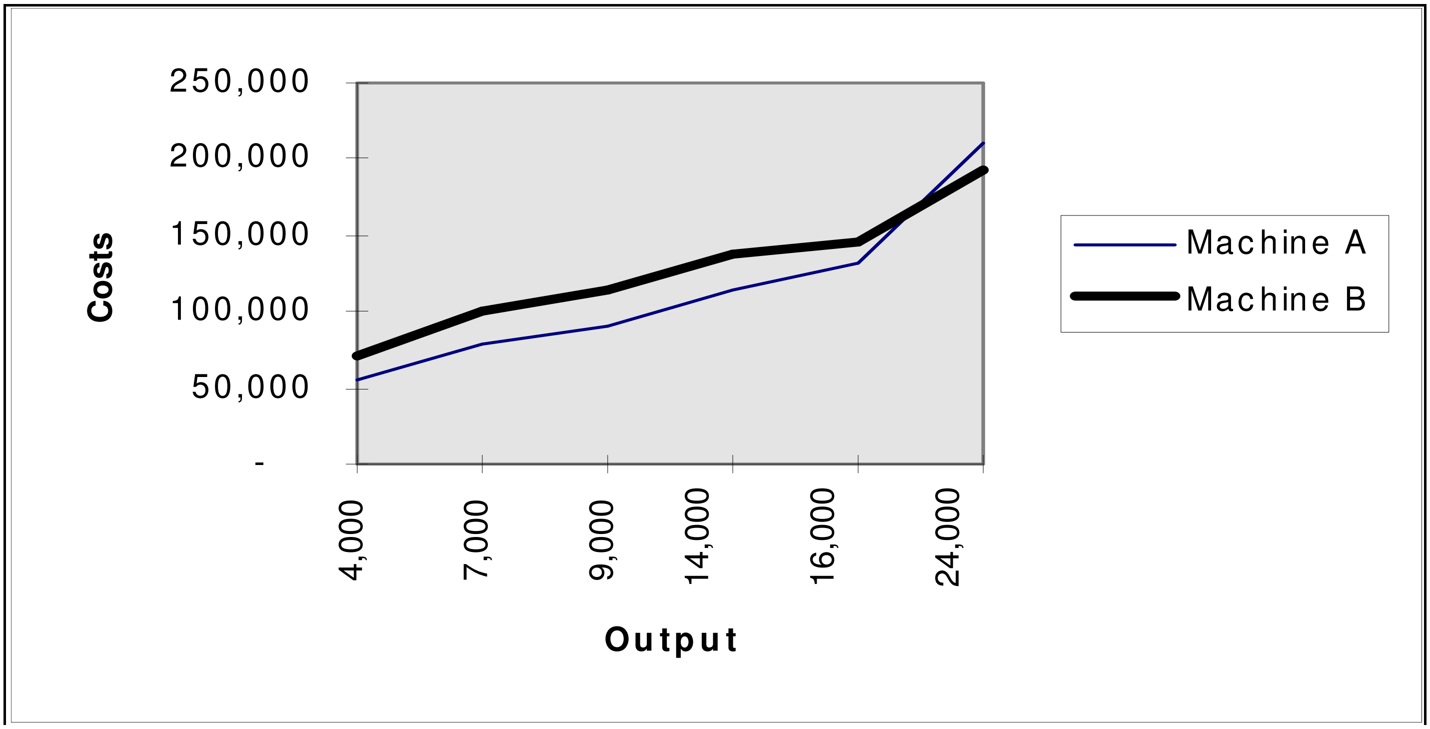


## 6-37 Cost Estimation; Machine Replacement; Ethics (25 min)

1. A graph of the data shows no significant outliers nor nonlinear relationships. See below




Using the High-Low method:

Machine A:

$$\text{slope} = \frac{\$210,000 - \$54,600}{24,000 - 4,000} = \$7.77$$

constant = \$210,000 - (\$7.77 x 24,000)
= \$23,520

or

= \$54,600 - (\$7.77 x 4,000)
= \$23,520

The estimate for total costs at 22,000 square yards is:
\$23,520 + (\$7.77 x 22,000) = \$194,460

At 15,000 yards:
\$23,520 + (\$7.77 x 15,000) = \$140,070

**Problem 6-37 (continued)**

Machine B:

$$\text{slope} = \frac{\$192,000 - \$70,000}{24,000 - 4,000} = \$6.10$$

constant = \$192,000 - (\$6.10 x 24,000) = \$45,600

or

= \$70,000 - (\$6.10 x 4,000) = \$45,600

The estimate for total costs at 22,000 square yards is:
\$45,600 + (\$6.10 x 22,000) = \$179,800

At 15,000 yards:
\$45,600 + (\$6.10 x 15,000) = \$137,100

Costs are lower at both the 22,000 level and the 15,000 level for Machine B.

2. The ethical issue presented in this case should be addressed using the approach described in chapter 1. Here it seems important to consider the nature and extent of the effect of the defect on customers and also SpectroGlass. Since the glass is used in office buildings, and defects are likely to affect the safety of those using the buildings, the cost analyst has a responsibility to make sure that management has a clear picture of the costs of each machine. The cost analysis should be presented to top management in a way that makes the ethical choice apparent and appropriate. For this reason, the calculations should not be modified.

3. In addition to the costs of the machine, SpectroGlass should be aware of any import duties or restrictions for the purchase of the machines from Germany or Canada. How will these restrictions and duties, if any, affect the cost and availability of the machine? Also, will the purchase in either country lead to potentially beneficial business relationships in that country. For example, the purchase in a given country might open up new markets for SpectroGlass.

## 6-38 Cost Estimation; High-Low Method (25 min)

Estimated cost of electricity equals $210 (from information about August)

($870 - $210)/(20 - 60) = - $16.50 /degree

<u>At 20 degrees F:</u>

$870 = intercept + (-$16.50 x 20)
intercept = $1,200

Cost equation: Utilities cost = $1,200 - $16.50 x degrees above zero

A cost estimate for January is not available since the average temperature of 10 degrees is outside the relevant range of the data used to develop the high-low estimate.

The cost estimate for February is: $1,200 - $16.50 x 40 = $540

Note to instructor: the problem can also be solved using regression analysis, as follows:

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.94586791 |
| R Square | 0.8946661 |
| Adjusted R Square | 0.88413271 |
| Standard Error | 80.2291185 |
| Observations | 12 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 1 | 546709.8021 | 546709.8 | 84.93619847 |
| Residual | 10 | 64367.1146 | 6436.711 | |
| Total | 11 | 611076.9167 | | |

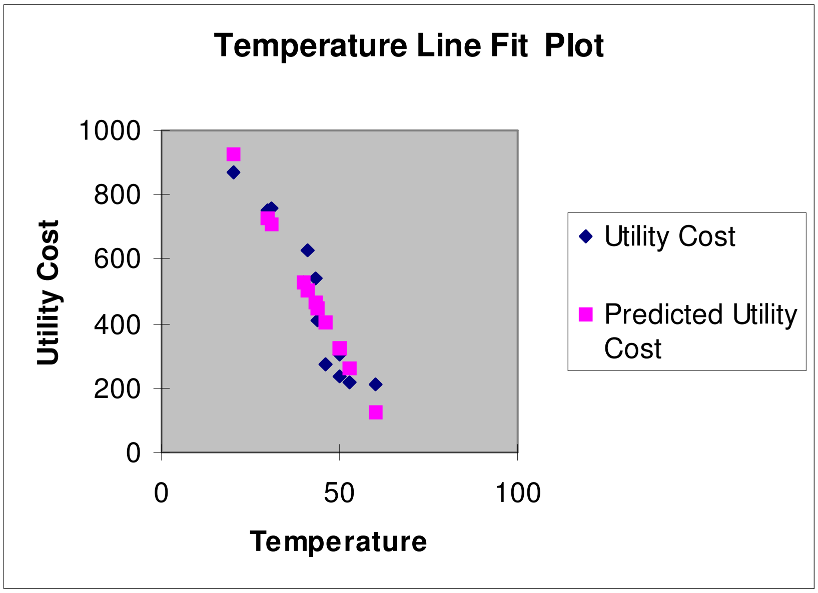
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 1330.88094 | 95.4055084 | 13.94973 | 7.00913E-08 |
| Temperature | -20.1487624 | 2.186260788 | -9.216084 | 3.34144E-06 |

Cost equation: Utility cost = $1,331 - $20.1488 * Temperature

Estimated cost for February with a predicted temperature at 40 degrees:
$1,331 - $20.1488 * 40 = $525

**Problem 6-38 (continued)**

The cost/temperature relationship is shown in the Excel chart below:



**6-39 to 6-43 Regression Analysis (25 min)**

**6-39** b ($4,470 - $2,820)/(520 - 300) = $7.50;
$4,470 - 520 x $7.50 = $570

**6-40** d $684.65 + 420 x $7.2884 = $3,745.78

**6-41** b 0.99724

**6-42** a 99.724%

**6-43** d SE = $34.469;
Predicted Cost at 400 hours = $684.65 + 400 x $7.2884 = $3,600;
confidence range = $3,600 +/- $34.469

**6-44 Regression Analysis (25 min)**

1.

a. Using the regression results relating total workers to the number of orders received (Regression I), the number of temporary workers needed is 17, calculated as follows:

$$
\begin{aligned}
W &= a + bT \\
&= 21.938 + .0043\ (12,740) \\
&= 76.720 \text{ (Round to 77)}
\end{aligned}
$$

| | |
|---|---|
| Total workers needed | 77 |
| Less permanent workers | <u>60</u> |
| Temporary workers | 17 |

b. Using the regression analysis that relates the number of temporary workers to the number of orders received (Regression 2), the number of temporary workers needed is 19, calculated as follows:

$$
\begin{aligned}
W &= a+bT \\
&= -46.569 + .0051\ (12,740) \\
&= 18.405
\end{aligned}
$$

Temporary workers = 19 (rounded)

2. Regression Analysis 2 appears to be better for the following reasons.

- The standard error of the W estimate is lower for Regression 2 (1.495 vs. 3.721) which usually means that the prediction is expected to be more accurate. However, note that the dependent variable in Regression 2 (temporary workers) is on the average somewhat smaller (about 1/4) that of regression 1 (total workers), so that the standard errors are not directly comparable. Because the dependent variable for Regression 2 is smaller, we would expect the standard error also to be smaller. Bloom needs to compare the SE to the mean of the dependent variable (not available in this case) to further understand whether the regressions provide useful predictions.

**Problem 6-44 (continued)**

- The coefficient of determination is higher for Regression 2 (.755 vs. .624) which indicates a more reliable regression model; the R Squared values are OK but not particularly strong.
- The t-value is slightly higher for regression 2 (2.04 vs 1.95 for regression 1); neither t-value is particularly good, an indication of a weak relationship between the independent variable and the dependent variable.
- Since both the regressions are quite similar, Bloom should consider why one should lead to better results than the other. That is, could not Bloom argue that regression 1 is preferred on the logical basis that there should be a stronger relationship between total orders and total number of workers? Even if the statistical measures are more favorable for regression 2, regression 1 might be preferable because it represents a more plausible relationship. Since any regression is based on a sampling of data observations, it might be that if several regressions were done over many time periods, the regression 1 model would show better statistical measures.

3. At least three ways that Peter Bloom could improve his analysis in order to get better predictions are described below.

- Perform regression calculations on a daily rather than weekly basis, since Bloom is analyzing the need for temporary workers on a daily rather than a weekly basis.
- Note the comment (in part 2 above) about the need for further data analysis to examine the value of regression 1 versus regression 2.
- Run the regression on data for more than one year if this data is available. If the historical data is a good indicator of future trends, this will result in better predictions of expected orders. If several months of data are used, Bloom should make sure that there are no shifts or cyclical changes in the data over this longer period of time.

**Problem 6-44 (continued)**

- Graph the data and look for outliers and other signs of non-linearity; note however that the Durbin Watson statistics for both regressions are in the OK range.
- Study the data for seasonality and modify the model if appropriate.

**6-45 Regression Analysis; Evaluating Regression Equations (20 min)**

1. The Pilot Shop should adopt regression 2 to forecast total shipping department costs for the following reasons:

a. R-squared, the coefficient of determination (the proportion of the variance explained by the independent variable), is higher for regression 2.

b. The standard error of the estimate, which is a measure of the precision of the regression, is smaller for Equation 2. The standard error is used to make estimates of confidence in the prediction of the regression equation.

c. The t-value for regression 1 is poor (1.89) while the t-value in regression 2 is OK (3.46).

d. Since we do not have a Durbin-Watson statistic to measure the potential for nonlinearity, and since we do not have a graph of the data to examine for nonlinearity, we are unable to assess the potential for nonlinearity in the two regressions. Similarly, without a graph we are unable to assess the potential for nonconstant variance in the data.

2. Since the number of orders to be shipped next week is given, the appropriate estimation model is regression 2, and the total estimated shipping cost is $2,994.90.

SC = 642.9 + 3.92 x NR
SC = 642.9 + 3.92 x 600
SC = $2,994.90

3. An important limitation of the regression we have chosen is that we have not been able to assess the potential for nonlinearity in the relationships among the variables. The presence of nonlinear relationships can be assessed by examining the Durbin-Watson statistic and/or by examining the graphs of the data. One of Shephard's first tasks should be to examine for potential nonlinearity in the data.

Another limitation Shephard should consider is the potential for unreliability in the data, either due to outliers or to errors in the data used in the regression. Outliers should be removed and errors should be corrected for obtaining the regression results.

An additional limitation is that we are unable to assess the potential for nonconstant error variance since we do not have a graph of the data.

**Problem 6-45 (continued)**

The global nature of the Pilot Shop's operations adds another limitation to the analysis. The purchasing and shipping costs will vary with international business conditions and also with fluctuations in foreign currencies. Moreover, customs restrictions and charges are involved in international trade.

## 6-46 Regression Analysis (20 min)

1.

| Independent Variables | Results | Example | Totals |
|---|---|---|---|
| Regression intercept | 1,224 | | 1,224 |
| | | | |
| Attendance at prior concert | | | |
| Coefficient | 3,445 | 1 | 3,445 |
| t-value | 4.11 | | |
| | | | |
| Spending on advertising | | | |
| Coefficient | 0.113 | 35,000 | 3,955 |
| t-value | 1.88 | | |
| | | | |
| Performer's CD sales | | | |
| Coefficient | 0.00044 | 10,000,000 | 4,400 |
| t-value | 1.22 | | |
| | | | |
| Television appearances | | | |
| Coefficient | 898 | 0 | |
| t-value | 2.4 | | |
| | | | |
| Other Public performances | | | |
| Coefficient | 1,233 | 6 | 7,398 |
| t-value | 3.7 | | |
| | | | |
| R-squared | 0.88 | | |
| Standard error of the estimate | 2,447 | | |
| Total Projected Attendance | | | **20,422** |

2. The overall reliability of the regression, as measured by R-squared is very good, at 88% and the standard error of the estimate, at 2,447 is reasonably small, considering the level of predicted attendance, 20,422. On the other hand, two of the five independent variables have unsatisfactory t-values. The t-values for the advertising variable and the CD sales variables are less than 2.0, indicating these variables have a non-significant relationship to number of ticket holders. Rock n' Roll should consider removing them from the model. Other potentially useful variables include dummy variables for the timing of the performer's appearance (near a holiday weekend, early or late in the season, the prior appearance was on a rainy day, etc), and other variables related to the performer's popularity, such recent appearances in the print media, release of a new CD or single, etc.

## 6-47 Regression Analysis; Appendix

The correlation analysis shows that only one of the correlations is significant at the .05 level – order size and runtime, and the relationship is negative, or inverse. That is, the larger the order size, the smaller the runtime per unit. Based on an actual company, this result is due to the fact that the machine operators slowed the machine time at the start of each order to ensure that the order was running properly before getting the machine up to the normal runtime speed. The effect of this practice is that larger orders saved the company in two ways. First, it should reduce the average per unit setup time (there is evidence of this in the data, but the correlation of -.209 is not statistically significant) since setup time varies per batch (order) and not by units in the order. Second, the larger orders allowed the machine operators to operate the machines at a higher than average speed relative to the operating speeds for the smaller orders, thus saving runtime on the larger orders.

Another informative aspect of the correlation analysis is to show the positive (.452) and marginally significant (p=.08) relationship between complexity and setup time per unit. This means that greater complexity tends to increase setup time, an intuitive result.

2. The information above is particularly useful to PolyChem as it begin to focus on smaller customers in order to find profitable alternatives to the low-cost competition it now faces. The key point is that selling in smaller and more customized orders will increase setup and runtime costs, as illustrated from the correlation analysis. Smaller orders lead to slower runtime, and more complex orders lead to longer setup time, for higher overall unit costs for these smaller orders. The company should consider pricing policies and other cost control measures to ensure the success of this new strategy.

This problem is adapted from the Laurent Company case, which is included in the casebook that accompanies this text.

## 6-48 Regression Analysis (20min)

1. Assuming that all purchases of autos for resale (cost of goods sold) represent variable costs

Price = \$30,000,000/1,500 = \$20,000
Variable cost per unit =
= (\$862,500 + .9 x \$2,300,000 + \$24,750,000) / 1,500
= \$ 18,455
Fixed cost = \$1,854,000 + .1 x \$2,300,000 = \$2,084,000

<u>Profit for 2,000 units sold</u>

| | | |
|---|---|---|
| Sales | 2,000 x \$20,000 = | \$40,000,000 |
| Less Variable costs | 2,000 x \$18,455 = | 36,910,000 |
| Contribution Margin | | 3,090,000 |
| Less Fixed costs | | 2,084,000 |
| Profit | | \$1,006,000 |

2.
a. The relevant range is the band or range of activity within which specified cost relationships (behavioral assumptions) remain valid and fixed costs remain fixed.

b. The R-squared value is a measure of the goodness of fit between the independent and dependent variable, the extent to which the independent variable accounts for the variability in the dependent variable. An R-squared value of 0.60 indicates that 60% of the total variation in mixed expenses is explained by the regression equation.

c. The composite-based relationships may not be realistic and could result in incorrect predictions. Application of these relationships to specific new dealerships may not allow for regional variations in items such as wages and rents and may not include factors that are peculiar to the start-up of a dealership.

d. The standard error of the estimate is the measure of precision of the regression. The standard error of the estimate helps to determine the range of the accuracy of the estimate with a given degree of confidence.

**Problem 6-48 (continued)**

3. Using the regression equation that Jack Snyder developed, the approximate range of sales that could occur during the year is calculated below.

Range of sales = Sales +/- (Standard error x 2)
= $28,500,000 +/- ($4,500,000 x 2)
= $28,500,000 +/- $9,000,000
= $19,500,000 to $37,500,000

Note that this is a relatively wide range for the prediction. Consider that the ratio of the standard error to the amount to be predicted is $4,500/$28,500 = 15.8%, an indication of a relatively poor SE and a relatively poor regression. The mediocre R-Squared of 60% is a further indication of the weakness of the model, and therefore of the lack of precision of the predictions from the model.

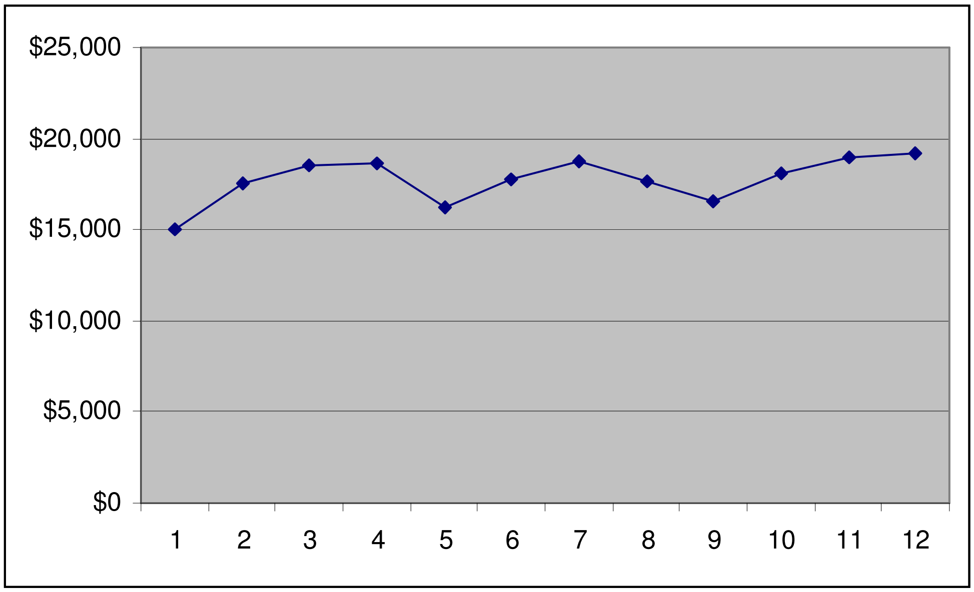
4. A key issue for USMI is the risk of expanding its dealership network. Regression analysis allows financial managers to make predictions about the effect of the proposed expansion on sales and profits. While Jack Snyder's model is not particularly reliable or precise (part 3 above), the approach is certainly worthwhile, especially if Jack can determine a way to modify his model to improve its reliability and precision, perhaps by including better independent variables. Strategically, firms that are better able to predict sales, using regression and/or other methods, will be in a stronger competitive position – the firm's planning will be more focused and effective.

**6-49 Cost Estimation; High-Low Method, Regression Analysis (30min)**

1. High-Low Method

   An examination of the exhibit below indicates that representative high and low points are the last and first data points, respectively, so these points are used to develop the high-low estimate.




Variable cost = ($19,200 - $15,000)/(12 - 1) = $381.82
Fixed cost = $ 15,000 – ($381.82 x 1) = $14,618.18
[also: $19,200 – ($381.82 x 12) = $14,618.16]

Quarterly Predictions are:

$14,618 + $381.82 x 13 = $19,582
$14,618 + $381.82 x 14 = $ 19,963
$14,618 + $381.82 x 15 = $ 20,345
$14,618 + $381.82 x 16 = $ 20,727

Regression

The regression equation from the spreadsheet is:

Return Expense
= $16,559 + (quarter number x 183.22)

**Problem 6-49 (continued)**

Predicted Expense for the next four quarters using regression analysis:

| | | | |
|---|---|---|---|
| 1 | $16,559+ (13 x $183.22) | = | $ 18,940.86 |
| 2 | $16,559 + (14 x $183.22) | = | $ 19,124.08 |
| 3 | $16,559 + (15 x $183.22) | = | $ 19,307.30 |
| 4 | $16,559 + (16 x $183.22) | = | $ 19,490.52 |

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.523289854 |
| R Square | 0.273832272 |
| Adjusted R Squa | 0.201215499 |
| Standard Error | 1128.260621 |
| Observations | 12 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 1 | 4800279.72 | 4800280 | 3.770923 |
| Residual | 10 | 12729720.28 | 1272972 | |
| Total | 11 | 17530000 | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 16559.09091 | 694.39641 | 23.84674 | 3.82E-10 |
| Trend | 183.2167832 | 94.34989292 | 1.941886 | 0.080828 |

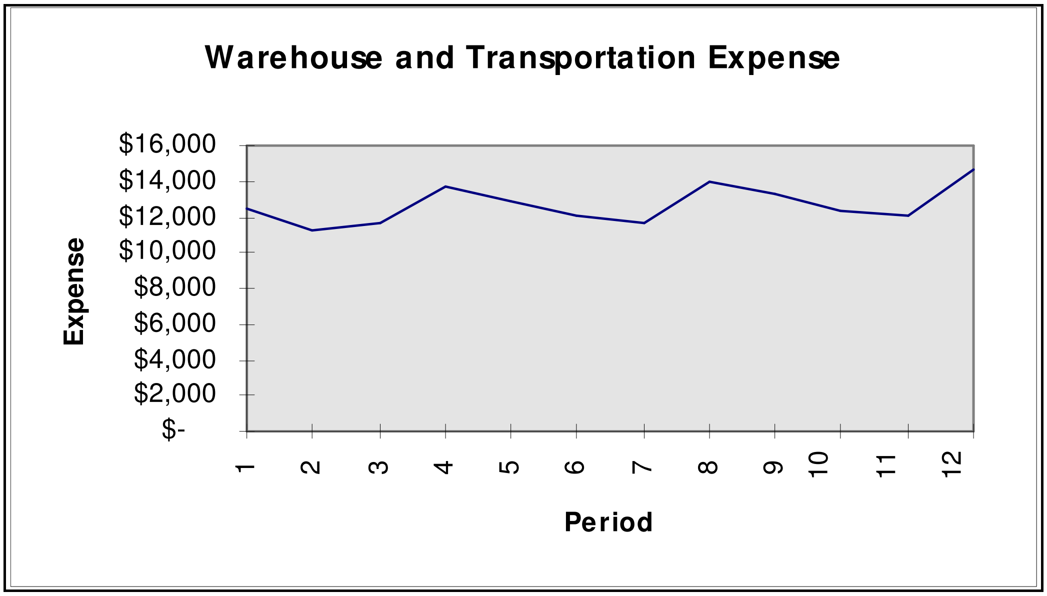
Regression analysis is the best method to recommend, because it gives the best-fitting line. An inspection of the graph would suggest seasonality. The accountant should perform a seasonal adjustment to further improve the model.

2. If DVD Express is involved in global production of its products, then expenses incurred from returns must be analyzed by production facility, as these costs are likely to differ among production facilities due to different equipment used in manufacturing the DVD players.

**6-50 Cost Estimation; High-Low Method; Regression (50 min)**

1. <u>High-Low Method</u>

An examination of the exhibit below indicates that representative high and low points are the last and first data points, respectively, so these points are used to develop the high-low estimate. The second data point is the lowest point but not considered representative of the data given the graph.




Variable cost = ($14,600 - $12,500)/(12 - 1) = $190.91
Fixed cost = $ 12,500 - $190.91 x 1 = $12,309 = $14,600 - $190.91 x 12

Quarterly Predictions are:

$12,309 + $190.91 x 13 = $14,791
$12,309 + $190.91 x 14 = $14,982
$12,309 + $190.91 x 15 = $15,173
$12,309 + $190.91 x 16 = $15,364

**Problem 6-50 (continued)**

Regression

The regression equation from the spreadsheet is:

Warehouse and Transportation Expense
= \$11,855 + (quarter number x \$126.22)

Predicted Expense for the next four quarters using regression analysis:

| | | | |
|---|---|---|---|
| 1 | \$11,855 + (13 x \$126.22) | = | \$13,496 |
| 2 | \$11,855 + (14 x \$126.22) | = | \$13,622 |
| 3 | \$11,855 + (15 x \$126.22) | = | \$13,748 |
| 4 | \$11,855 + (16 x \$126.22) | = | \$13,875 |

| Clothes for U Regression Estimation | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | |
| 1 | 12500 | SUMMARY OUTPUT | | | | | | | | |
| 2 | 11300 | | | | | | | | | |
| 3 | 11600 | *Regression Statistics* | | | | | | | | |
| 4 | 13700 | Multiple R | 0.439734 | | | | | | | |
| 5 | 12900 | R Square | 0.193366 | | | | | | | |
| 6 | 12100 | Adjusted R S | 0.112703 | | | | | | | |
| 7 | 11700 | Standard Err | 974.8929 | | | | | | | |
| 8 | 14000 | Observations | 12 | | | | | | | |
| 9 | 13300 | | | | | | | | | |
| 10 | 12300 | ANOVA | | | | | | | | |
| 11 | 12100 | | *df* | *SS* | *MS* | *F* | *ignificance F* | | | |
| 12 | 14600 | Regression | 1 | 2278339 | 2278339 | 2.397202 | 0.152593 | | | |
| | | Residual | 10 | 9504161 | 950416.1 | | | | | |
| | | Total | 11 | 11782500 | | | | | | |
| | | | | | | | | | | |
| | | | *Coefficients* | *tandard Err* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *ower 95.0%* | *pper 95.0%* |
| | | Intercept | 11854.55 | 600.0051 | 19.75741 | 2.42E-09 | 10517.65 | 13191.44 | 10517.65 | 13191.44 |
| | | X Variable 1 | 126.2238 | 81.52464 | 1.54829 | 0.152593 | -55.4245 | 307.872 | -55.4245 | 307.872 |

Regression analysis is the best method to recommend, because it gives the best-fitting line. An inspection of the graph would suggest seasonality. The accountant should perform a seasonal adjustment to further improve the model.

2. If Clothes for U is involved in global sourcing for its stores, then transportation costs and warehousing costs must be analyzed by country, as these costs are likely to differ significantly among countries. Also, in addition to transportation and warehousing, Clothes for U should consider including in the analysis the additional costs of global business, including customs duties, delays, and other costs.

## 6-51 Learning Curves (20 min)

The average production hours per unit obviously decreased as the output increased. This decrease corresponds very closely to that of a 90% learning curve.

1. An estimate of the hours required to build 16 aircraft is 2,624 hours.

| Output | Avg. Time | Total Time |
|---|---|---|
| 1 | 250 | 250 |
| 2 | 225 (250 x .9) | 450 (2 x 225) |
| 4 | 203 (225 x .9) | 812 (4 x 203) |
| 8 | 182 (203 x .9) | 1,456 (8 x 182) |
| 16 | 164 (182 x .9) | 2,624 (16 x 164) |

2. The role of learning curves is to help predict future costs when significant learning takes place in the work. When learning is present, unit costs increase at a nonlinear, decreasing rate, so that linear estimation methods such as regression and the high-low method are not as appropriate. The learning curve method takes into account the nonlinear learning behavior in the situation.

### 6-52 Learning Curves (25 min)

| Number of jobs | Average labor hrs | Total labor hrs |
|---|---|---|
| 1 | 17 | 17 |
| 2 | 13.6=17x.8 | 27.2=13.6x2 |
| 4 | 10.9 | 43.6 |
| 8 | 8.7 | 69.6 |

The average time per apartment at the end of the summer is:

| | |
|---|---|
| Total time to complete 8 apartments | 69.6 |
| Total time to complete 4 apartments | 43.6 |
| Total time to complete the last 4 apartments | 26.0 |

Average time per apartment = 26/ 4 = 6.5 hours

Yes, they achieved the goal to complete one apartment in less than 8 hours.

**6-53 Learning Curves (20 min)**

| Cum-ulative Output | Average Labor Time | Total Labor Time | Labor Costs/ Unit | Total Labor Costs | Total Direct Costs/Unit | Total Costs/ Unit |
|---|---|---|---|---|---|---|
| 100 | .25 | 25 | $3.75 | $ 375.00 | $20.00 | $30.00 |
| 200 | .2125 | 42.5 | $3.19 | $ 637.50 | $19.44 | 29.44 |
| 400 | .180625 | 72.25 | $2.71 | $1083.75 | $18.96 | 28.96 |
| 800 | .153531 | 123 | $2.30 | $1842.30 | $18.55 | 28.55 |
| 1600 | .130502 | 209 | $1.96 | $3132.04 | $18.21 | 28.21 |
| 3200 | .110926 | 355 | $1.67 | $5324.46 | $17.92 | 27.92 |

Calculation of the first row (100 units) is as follows:
25 hours / 100 hats = .25 hrs/hat
.25 hours x $15/hr = $3.75/unit
100 units x $3.75/unit = $375
$3.75/unit (labor) + $16.25/unit (all else) = $20.00 cost/unit
$8,000 fixed costs/800 hats/month = $10 fixed costs/unit

The selling price is $25.00 ($20 x 1.25; from table and calculations above)

2. Yes, EH can produce 1,600 hats for a unit direct cost of $18.21. The offer of $20.00 for each hat covers direct costs. The offer does not cover full cost, including fixed cost, but in the short-run fixed costs are irrelevant. EH can complete this order in two months (doing no other work).

## 6-54 Learning Curves (30 min)

Assuming an 80% learning curve, the production time will likely follow the schedule below:

| Batch Size | Average Time | Total Time | Increase in Time | Time Per Unit |
|---|---|---|---|---|
| 80 | 60.00 | 4,800.00 | 4,800.00 | 60.00 |
| 160 | 48.00 | 7,680.00 | 2,880.00 | 36.00 |
| 320 | 38.40 | 12,288.00 | 4,608.00 | 28.80 |

[Note to instructor: the total time of 12,288 can also be derived by using the power function in Excel (one of the "Math and Trig." functions). Use the formula $Y=ax^{-b}$ , where Y = average time per unit, a=$60, and x = the number of batches for which learning improves, and b = .322 for a learning rate of 80%. Set x=4 and b=.322 in the Excel function to find .63993, and 60 hours x .63993 = 38.396 hours. Since Y = the average time per unit, then 320 x 38.396 = 12,287 hours. ]

The total production of 1,000 units will require 320 units affected by learning (for a cumulative total time of 12,288 hours) plus an additional 680 units with no learning. The last 680 units will require 28.8 hours each:

| | |
|---|---|
| Total time for 920 additional units is: | |
| For first 320 units | 12,288.00 |
| For last 680 units x 28.8 hrs= | 19,584.00 |
| Total time for 1,000 units | 31,872.00 |
| Less time spent for first 80 units | 4,800.00 |
| Total hours for future production | 27,072.00 |

Future direct labor cost: 27,072 x $14.50 = $392,544

2. The 75% learning rate is faster than the 80% rate used in the above analysis, so that the labor hours and direct labor costs would be lower with a 75% rate. The fastest learning rate is a rate approaching 50% while the slowest rate is close to 100%.

**Problem 6-54 (continued)**

Estimated Production Time at 75% Learning Rate

| Batch Size | Average Time | Total Time | Increase in Time | Time Per Unit |
|---|---|---|---|---|
| 80 | 60.00 | 4,800.00 | 4,800.00 | 60.00 |
| 160 | 45.00 | 7,200.00 | 2,400.00 | 30.00 |
| 320 | 33.75 | 10,800.00 | 3,600.00 | 22.50 |

At 75% learning rate, the total time for 920 additional units is:

| | |
|---|---|
| Total time for 920 additional units is: | |
| For first 320 units | 10,800.00 |
| For last 680 units x 22.5 hrs= | 15,300.00 |
| Total time for 1,000 units | 26,100.00 |
| Less time spent for first 80 units | 4,800.00 |
| Total hours for future production | 21,300.00 |

Future Direct labor cost: 21,300 x $14.50 = $308,850

At 75% learning rate the firm may enjoy a saving of $83,694 in direct labor cost, a decrease of approximately 20% of the direct labor cost estimated at 80% learning rate.

3. Conditions that might reduce the potential for the benefits from learning curve analysis include:

- a simple task that is quickly learned, so that there is little to be gained from forecasting the rate of learning over time
- a poor work environment, so that workers do not have the motivation to achieve the expected learning rate
- ineffective or lacking incentive programs to provide the desired motivation
- high employee turnover, so that little of the learning is effectively used
- a task that is less labor intensive, so that direct labor is only a relatively small part of total costs

**Problem 6-54 (continued)**

Strategically, firms like Hauser that are better able to predict costs using learning curves and/or other methods will also be in a stronger competitive position – the firm's planning will be more focused and effective. For example, firms that determine whether to manufacture or outsource parts and components incorporating expected learning effects into the analysis, as in this case for the Hauser company, will improve their chances to implement world-class cost efficient manufacturing.

## 6-55 Cost Estimation; Regression Analysis (50 min)

1. The spreadsheet regression output for Plantcity is shown in Exhibits 6-55A, B and C. Exhibit 6-55A shows the regression which includes both predictors, sales dollars and sales units, while Exhibit 6-55B shows sales dollars only, and Exhibit 6-55C shows sales units only.

Exhibit 6-55A (Units and Dollars)

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.836460729 |
| R Square | 0.699666551 |
| Adjusted R Square | 0.678214162 |
| Standard Error | 356.8016909 |
| Observations | 31 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 2 | 8304227.689 | 4152114 | 32.61485 |
| Residual | 28 | 3564608.505 | 127307.4 | |
| Total | 30 | 11868836.19 | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 1720.993363 | 410.3481754 | 4.193983 | 0.000249 |
| Units | 1.663079443 | 0.351697453 | 4.728722 | 5.82E-05 |
| Dollars | 0.212611616 | 0.214591377 | 0.990774 | 0.330281 |

**Problem 6-55** (continued)

Exhibit 6-55B (Dollars)

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.678100395 |
| R Square | 0.459820145 |
| Adjusted R Square | 0.441193254 |
| Standard Error | 470.1909447 |
| Observations | 31 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 1 | 5457529.985 | 5457530 | 24.68582 |
| Residual | 29 | 6411306.209 | 221079.5 | |
| Total | 30 | 11868836.19 | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 650.5468079 | 451.0275889 | 1.442366 | 0.159913 |
| Dollars | 0.956144724 | 0.192441985 | 4.968483 | 2.77E-05 |

Exhibit 6-55C (Units)

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.830142974 |
| R Square | 0.689137358 |
| Adjusted R Square | 0.678417956 |
| Standard Error | 356.6886878 |
| Observations | 31 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 1 | 8179258.414 | 8179258 | 64.28879 |
| Residual | 29 | 3689577.78 | 127226.8 | |
| Total | 30 | 11868836.19 | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 2112.01648 | 112.3290416 | 18.80205 | 8.7E-18 |
| Units | 1.918401433 | 0.239260971 | 8.018029 | 7.66E-09 |

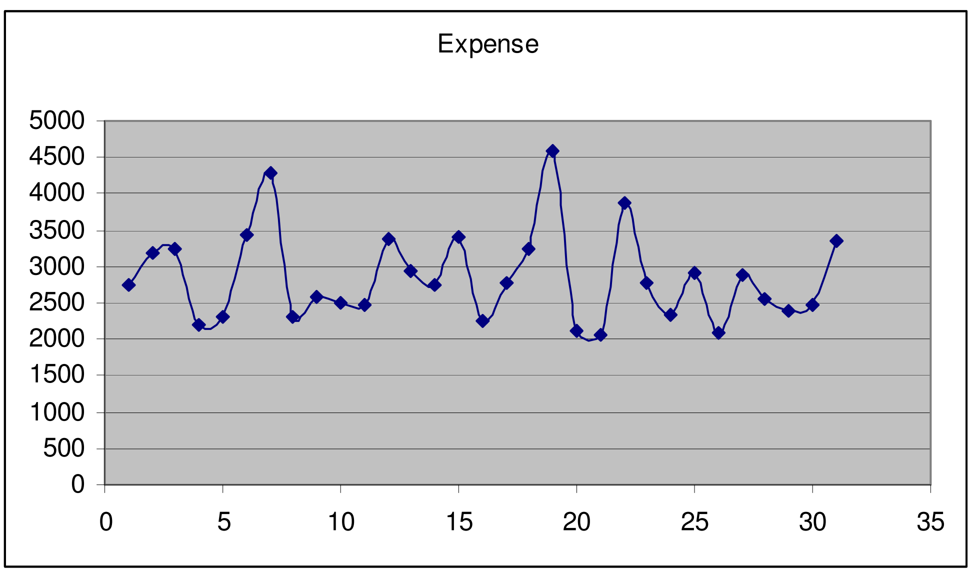
 

**Problem 6-55 (continued)**

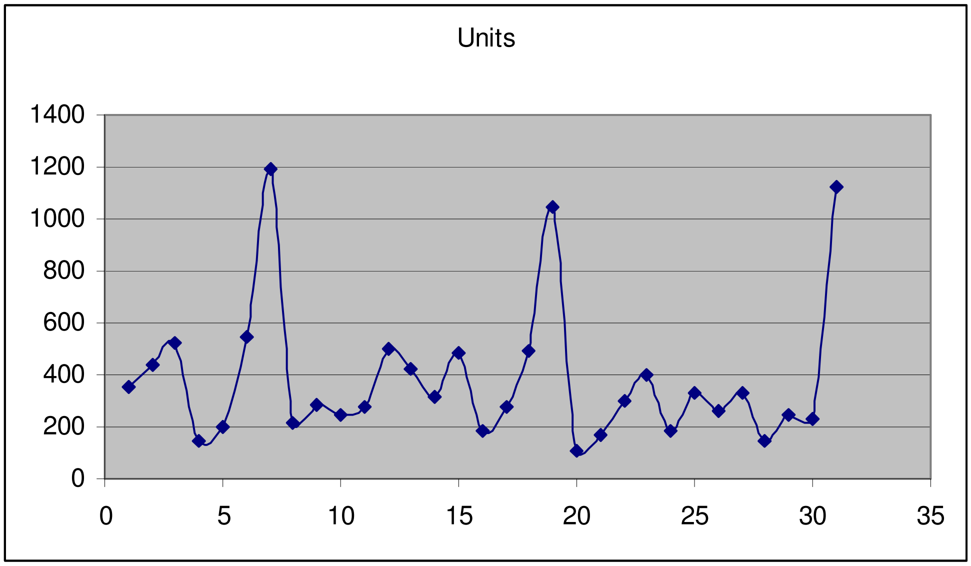
The precision of the regression shown in 6-55A is good, with a standard error of the estimate of 356 relative to a dependent variable with values averaging about 3,000. Also, the reliability of the model is quite good, with an R-squared of 68%, an F value of 32.6 and a t-value on sales units of 4.7. However, the t-value on the sales dollars variable is
poor, as shown by the low t-value (.99).

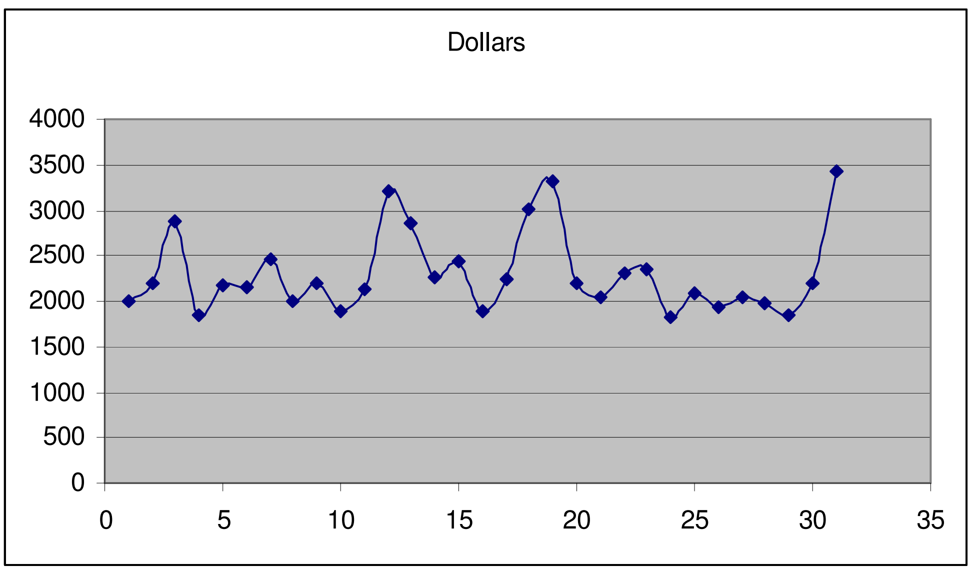
The regression using sales dollars only (Exhibit 6-55B) is somewhat worse while the regression on sales units (Exhibit 6-55C) gives almost equivalent R-squared and standard errors values to the model with both units and dollars. Because the regression on sales units only is simpler and has a lower standard error and higher R-squared, the model using only sales units is a logical choice for the cost estimation model in this case.

For further regression analysis on this data, consider the graphs below which shows evidence of seasonality in the data.



## Problem 6-55 (continued)




Since the graphs show clear evidence of seasonality, another try of the model with seasonality included would be a useful next step. The addition of a seasonal variable for the month of December improved the model in Exhibit 6-55C substantially.

**Problem 6-55 (continued)**

The seasonal model for sales dollars is shown in Exhibit 6-55D. Note the substantial improvement in R-squared; also note that the seasonal variable is significant. The coefficient on the seasonality variable is negative because supplies expense does not rise as fast as units sold in December.

Exhibit 6-55D

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.859051742 |
| R Square | 0.737969895 |
| Adjusted R Square | 0.719253459 |
| Standard Error | 333.2733966 |
| Observations | 31 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 2 | 8758843.802 | 4379422 | 39.42898 |
| Residual | 28 | 3109992.392 | 111071.2 | |
| Total | 30 | 11868836.19 | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 1815.233657 | 167.0183648 | 10.86847 | 1.48E-11 |
| Units | 2.949465938 | 0.503693179 | 5.85568 | 2.7E-06 |
| Season | -1042.036219 | 456.1679284 | -2.284326 | 0.030136 |

2. Predicted monthly figures for supplies expense using the regression in Exhibit 6-55D:

| | **Units** | **Seasonality** | **Predicted Expense** |
|---|---|---|---|
| Jan | 180 | 0 | 2346 |
| Feb | 230 | 0 | 2494 |
| Mar | 190 | 0 | 2376 |
| Apr | 450 | 0 | 3142 |
| May | 350 | 0 | 2848 |
| Jun | 350 | 0 | 2848 |
| Jul | 450 | 0 | 3142 |
| Aug | 550 | 0 | 3437 |
| Sep | 300 | 0 | 2700 |
| Oct | 300 | 0 | 2700 |
| Nov | 450 | 0 | 3142 |
| Dec | 950 | 1 | 3575 |

## 6-56 Cross-Sectional Regression (30 min)

1.

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.976518934 |
| R Square | 0.953589229 |
| Adjusted R Square | 0.95001917 |
| Standard Error | 25458.32309 |
| Observations | 15 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 1.73119E+11 | 1.73E+11 | 267.1074 | 4.77053E-10 |
| Residual | 13 | 8425640789 | 6.48E+08 | | |
| Total | 14 | 1.81545E+11 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | -5225.263287 | 10780.40244 | -0.4847 | 0.635954 |
| TPD | 157.5079291 | 9.637390778 | 16.34342 | 4.77E-10 |

Construction Cost Equation

Cost = -$5,225 + $157.508 x TPD

For Babylon, NY:

Cost = -$5,225 + $157.508 x 750 = $112,906,000

2. The regression has strong statistical measures. The R-squared is relatively high at 95.35%; the t-value for the independent variable TPD is high and the risk level (p) is low; the standard error of the estimate, at 25,458, is relatively small given the amounts predicted for the dependent variable, so overall, the regression looks very strong, and the management accountant should feel comfortable to rely on it in cost estimation. One way it would be improved is to significantly increase the amount of data; the current analysis is based on only 15 plants, and the accuracy of the regression model would improve with a much larger number of data points. Also, it is likely, that since the costs will increase faster for a project than the capacity (TPD) increase, due to engineering and other construction limitations, the analyst should ensure that the data under analysis is reasonably linear. If there is evidence of a non-linear relationship between cost and TPD, then the analyst should use non-linear regression methods. There are a variety of non-linear regression approaches. A practical and simple to apply method is to convert the data by taking the natural log (ln) of each data point and then running the regression with the logged data.

**6-56 (continued)**

The conversion of logs removes the multiplicative type of non-linearity from the equation. To see this, review the discussion in footnote 6 in Appendix A.

Source: Richard K. Ellsworth, "Cost-to-Capacity Analysis for Estimating Project Costs," *Construction Accounting and Taxation*, September/October 2005, pp 5-10.

## 6-57 Cross-sectional Regression (60 min)

1. The first regression we try includes all four independent variable; square feet, number of employees, location type, and sales dollars. Each of these variables has a plausible relationship to inventory spoilage. We find from the regression results (see below) that the R-squared is very good (95%), the SE is relatively poor at 15% = 370.51/(36,758/15). The t-values are good for two of the variables (location and square feet), poor for the sales variable, and marginal for the number of employees variable.

| Variable | t-value |
|---|---|
| Square Feet | 2.86 |
| Employees | -1.89 |
| Location | 3.63 |
| Sales | -.33 |

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.97305677 |
| R Square | 0.94683948 |
| Adjusted R Square | 0.92557527 |
| Standard Error | 370.518856 |
| Observations | 15 |

ANOVA

| | *Df* | *SS* | *MS* | *F* | *Sig F.* |
|---|---|---|---|---|---|
| Regression | 4 | 24451628 | 6112907 | 44.5274 | 2.43E-06 |
| Residual | 10 | 1372842 | 137284 | | |
| Total | 14 | 25824470 | | | |

| | *Coefficients* | *Std Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | -201.784428 | 393.7116 | -0.51252 | 0.61942 | -1079.03 | 675.4598 |
| Footage | 0.62591525 | 0.218101 | 2.86984 | 0.01667 | 0.139956 | 1.111874 |
| Employees | -73.7720711 | 38.88984 | -1.89695 | 0.08706 | -160.424 | 12.8799 |
| Location | 879.37668 | 242.0028 | 3.63375 | 0.00458 | 340.1608 | 1418.593 |
| Sales | -0.00074804 | 0.002235 | -0.33469 | 0.74477 | -0.00573 | 0.004232 |

**Problem 6-57 (continued)**

The negative t-value on “employees” suggests that the spoilage at some stores might be due to under-staffing, but the t-value is not strong enough for strong conclusions. The t-value on “sales” is weak enough that we should consider deleting the variable from the model; moreover, can we explain why the coefficient on this variable is negative? With this thinking, we decide to re-run the model keeping only the two independent variables: location and square feet. The results are shown below, under **“Regression Two.”**

The second regression, shown below, has comparable values for R-squared and SE, but the t-values are improved. Additionally, the F-value almost doubles, meaning a more statistically reliable model. For these reasons we have chosen to rely on this second regression model to complete the analysis for Jim. Note below the regression results there is a residual report which shows the predicted and actual values for spoilage at each store, and the error term (“residual”). A large positive residual is unfavorable while a large negative residual is favorable.

2. Stores 6 and 7 have relatively high spoilage for their given levels of square feet and location type, based upon the relationships for all 15 stores, as captured in the regression model. Why then are these two stores so different? Jim has now a basis for beginning an investigation. Jim might also want to investigate why the level of spoilage is so unexpectedly low at stores 12 and 14, to perhaps discover the factors associated with these stores (beyond the 4 variables already considered) that have contributed to their success.

## Problem 6-57 (continued)

**Regression Two: Square Footage and Location Only**

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.96194073 |
| R Square | 0.92532997 |
| Adjusted R Square | 0.91288496 |
| Standard Error | 400.865101 |
| Observations | 15 |

ANOVA

| | *Df* | *SS* | *MS* | *F* | *Sig F.* |
|---|---|---|---|---|---|
| Regression | 2 | 23896156 | 1.2E+07 | 74.3535 | 1.73E-07 |
| Residual | 12 | 1928314 | 160693 | | |
| Total | 14 | 25824470 | | | |

| | *Coefficients* | *Std Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | -777.895194 | 292.1796 | -2.66239 | 0.0207 | -1414.5 | -141.291 |
| Footage | 0.45825253 | 0.126322 | 3.62765 | 0.00347 | 0.18302 | 0.733485 |
| Location | 926.553993 | 256.4631 | 3.61282 | 0.00356 | 367.7689 | 1485.339 |

RESIDUAL OUTPUT

| *Store Number* | *Predicted Spoilage* | *Residuals* | Actual Spoilage |
|---|---|---|---|
| 1 | 1,248 | 264 | $1,512 |
| 2 | 2,862 | 143 | 3,005 |
| 3 | 1,615 | 71 | 1,686 |
| 4 | 1,707 | 201 | 1,908 |
| 5 | 2,794 | (410) | 2,384 |
| 6 | 4,201 | 605 | 4,806 |
| 7 | 1,753 | 500 | 2,253 |
| 8 | 1,523 | (80) | 1,443 |
| 9 | 3,619 | 136 | 3,755 |
| 10 | 1,180 | (157) | 1,023 |
| 11 | 1,294 | 258 | 1,552 |
| 12 | 2,679 | (560) | 2,119 |
| 13 | 5,439 | 67 | 5,506 |
| 14 | 3,687 | (653) | 3,034 |
| 15 | 1,157 | (385) | 772 |

## 6-58 Regression Analysis in Tax Court Cases (20 min)

The information on this solution was obtained from an article by B. Anthony Billings and D. Larry Crumbly, "the Use of Regression Analysis as Evidence in Litigating Tax-Related Issues," *Journal of Applied Business Research,* Summer 1996.

Reviewing ten tax court cases wherein regression analysis was used, the authors identified the following factors as important to the court:

- sample size; a sample of size three was found to be inappropriate
- the plausibility of the model, which required the regression analysts expertise both in regression analysis and in the knowledge of the relevant content area, the phenomena being examined
- inclusion of all relevant independent variables
- accuracy of the data entered into the model
- proper attention to and disclosure of regression assumptions such as tests for non-linearity, consideration of the acceptability of the values for R-squared, F, SE and t-values
- the independent variables used in the model must have a logical explanation for a relationship to the dependent variable; unexplained, or "spurious" relationships were not acceptable