Using Regression Analysis

Resources Glossary Help

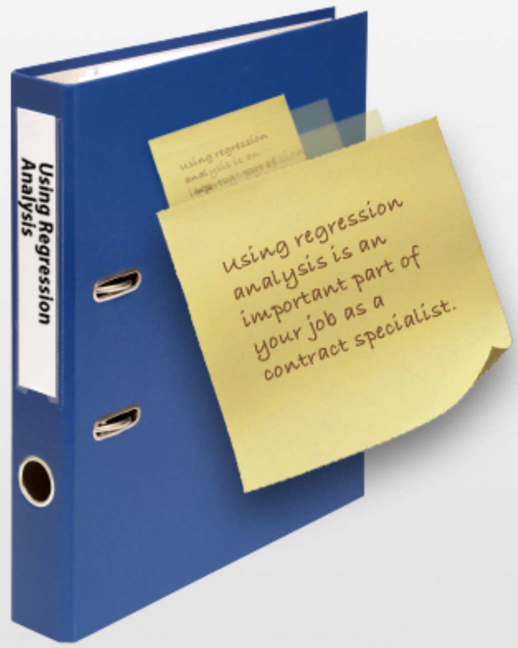
Module Introduction

Welcome to Using Regression Analysis!

You might be wondering, "How will the information in this module help me in my job as a contract specialist?"

As a contract specialist, you may need to use regression analysis in developing cost estimating relationships and other analyses based on a straight-line relationship even when the data points do not fall on a straight line.

This module will provide information on simple regression (2-variable linear regression) in which a single independent variable (X) is used to predict the value of a single dependent variable (Y).



D

User Instructions: Select Next to continue.

1 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Objectives

At the end of this lesson, you will be able to:

- Define regression analysis
- Identify the different regression analysis methods
- Identify contract pricing situations where simple regression analysis should be considered
- Identify the steps for using simple regression analysis

Select Next for a high-level overview of the content that will be presented in this module.

User Instructions: Select Next to continue.

2 of 53 Back Next

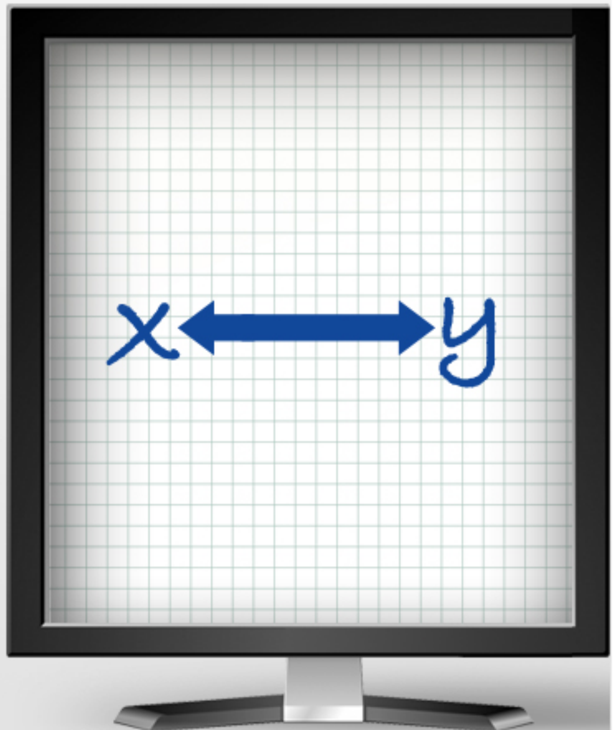
Using Regression Analysis

Resources Glossary Help

Overview - What Is Regression Analysis?

Regression analysis is a statistical technique used to establish the relationship of an independent variable, such as thrust or square footage, and one or more dependent variables, such as the cost of aircraft engines or housing.

Essentially, regression analysis attempts to measure the degree of correlation between the dependent and independent variables. This analysis establishes the independent variables' predictive value.



D

User Instructions: Select Next to continue.

3 of 53 Back Next

Using Regression Analysis

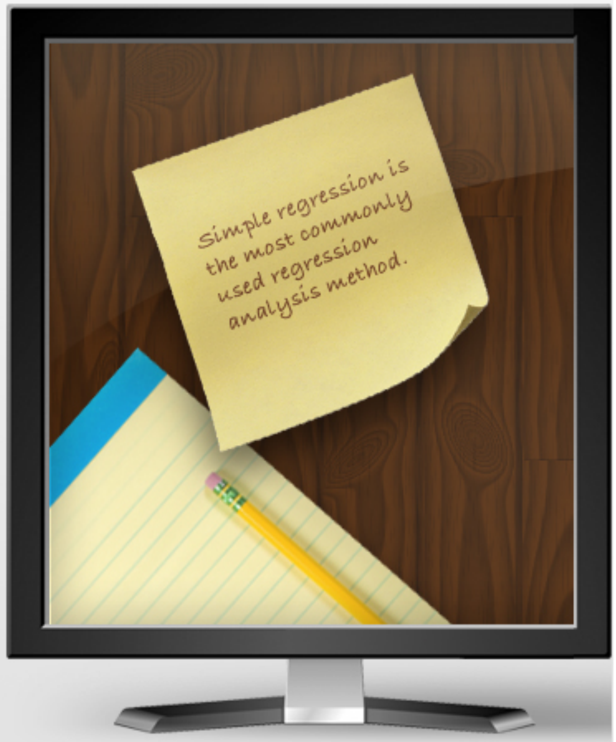
Resources Glossary Help

Overview - Simple vs. Complex Regression

Although this module will focus on simple regression, there are some situations in which you may need more powerful regression analysis tools.

Simple regression, sometimes called 2-variable linear regression, is when a single independent variable (X) is used to predict the value of a single dependent variable (Y).

Multiple regression (multivariate linear regression) and curvilinear regression are variations of simple regression that you may find useful. The general characteristics of both will be addressed later in this module.



User Instructions: Select Next to continue.

4 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Overview - Regression Analysis and Contract Pricing

Regression analysis is an important component of your job as a contract specialist. Remember not to get too caught up in the detailed manual calculations; there are numerous quantitative tools to help you with the "number crunching."

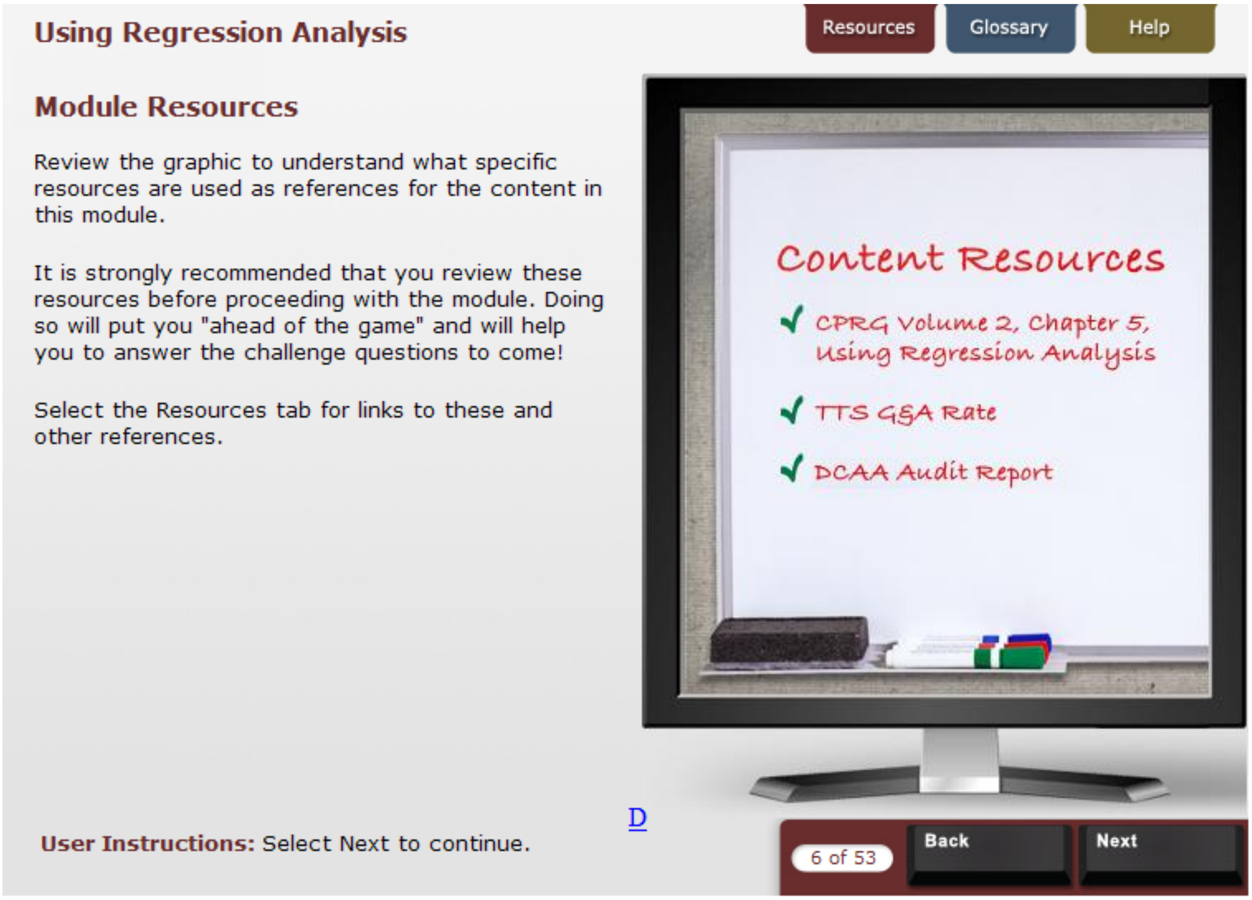
Your job is to determine if the regression equation is appropriate to use in pricing the contract action by analyzing certain statistics.

Regression analysis allows you to predict approximate pricing based on two or more variables. This provides a baseline against which you can judge proposed prices and determine price reasonableness.

D

User Instructions: Select Next to continue.

5 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
Module Resources
Review the graphic to understand what specific resources are used as references for the content in this module.
It is strongly recommended that you review these resources before proceeding with the module. Doing so will put you "ahead of the game" and will help you to answer the challenge questions to come!
Select the Resources tab for links to these and other references.
Content Resources
CPRG Volume 2, Chapter 5, Using Regression Analysis
TTS GSA Rate
DCAA Audit Report
D
User Instructions: Select Next to continue.
6 of 53
Back
Next

Using Regression Analysis

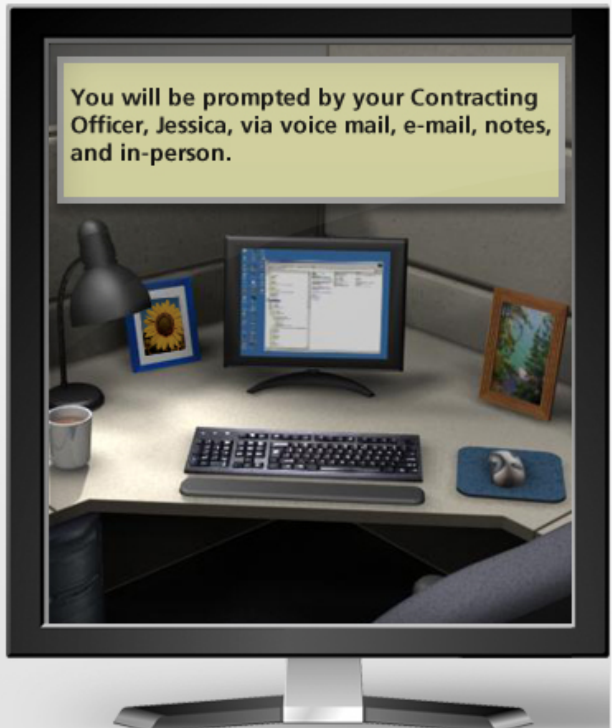
Resources | Glossary | Help

Simulation Overview

In this module, you are going to answer a series of Challenge Questions about using regression analysis. Do not worry if you miss a question - you will be directed to pertinent content and then given a chance to answer the question again.

Even if you get a question right the first time, it is strongly recommended that you elect to review the pertinent content by selecting the Review button. This review will help you answer the practice test questions at the end of the module.

Remember to use the Resources and Glossary tabs at any time during the simulation to access helpful information.



D

User Instructions: Select Next to continue.

7 of 53 | Back | Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Hi, it's Jessica, your contracting officer. I want you to review the DCAA Audit Report on the G&A rate.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CC
CAPTIONS
The contractor proposed a G&A rate of 15% based on the Forward Price Rate Agreement;
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
however, since that time DCAA has concluded an incurred cost audit and found TTS underestimated their G&A base by approximately 24%.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
DCAA is recommending a G&A rate of 11.17% based on the result of regressing the contractor's pool and base amounts.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Based on this information, and your knowledge of the upcoming $20.9M modification of the BBOMS contract, I want you to contact the ACO and verify that the DCAA recommended rate of 11.17% is now the DCMA FPRR.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
I left DCMA Info Memo 11-108 on your desk, as well as a Defense Procurement Memo dated Jan 4, 2011 so you can review the new FPRA and FPRR policy.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CC
CAPTIONS
Before you contact the ACO, I want you to use your regression skills to ensure that you fully understand the DCAA Auditor's position on the G&A rate.
User Instructions: Select Next to continue.
8 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Call me if you have any questions!
User Instructions: Select Next to continue.
8 of 53
Back
Next

Resources | Glossary | Help

Using Regression Analysis

Challenge Question #1

The simple regression equation assumes ____________________. Select all of the correct answers.

- ☐ A. When graphed, the data forms a downward curved line.
- ☐ B. It takes more than one independent variable to estimate new values of the dependent variable.
- ☐ C. There is a linear relationship between the independent and dependent variables.
- ☐ D. The regression equation is used to predict new values of Y based on the known values of X.
- ☐ E. It is easy to predict an accurate value for Y by plotting the data points and approximating the line of best fit visually.

Check Answer

User Instructions: Select the correct answers and then select the Check Answer button.

9 of 53 Back Next

Resources Glossary Help

Using Regression Analysis

Simple Regression

Simple regression is used to predict the value of a dependent variable (Y) given the value of the independent variable (X) based on a straight-line/linear relationship even…

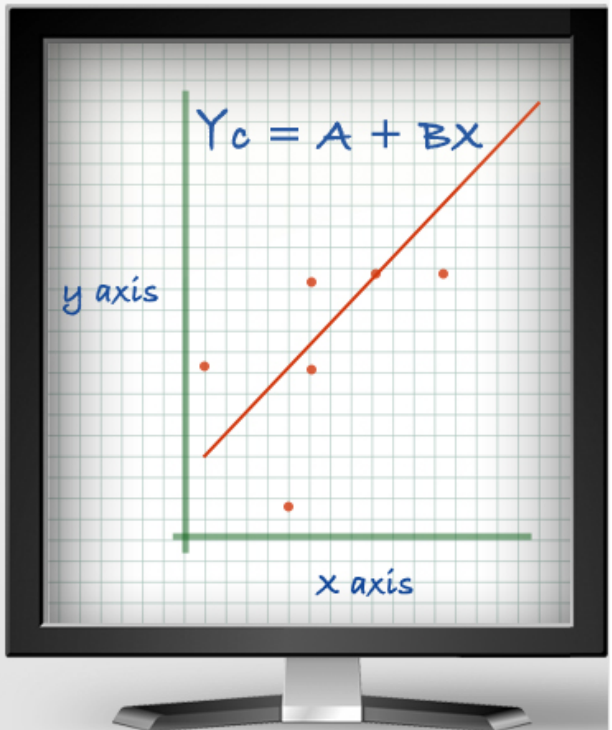
> **when data points do not fall in a straight line.**

The dependent variable is normally price or cost measured in dollars or labor hours.

The independent variable is a measure related to the item or service being acquired:

- Physical characteristic (measures big or small)
- Performance characteristic (measures fast or slow)
- Element of cost (labor hours)

This is the most commonly used regression tool for contract specialists, especially when looking at Cost Estimating Relationships (CERs).



D

User Instructions: Select Next to continue.

9 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Multiple Regression

Multiple regression analysis assumes that the change in Y can be better explained by using more than one independent variable.

For example, suppose you want to determine the relationship between:

1. Main-frame computer hours
2. Field-audit hours expended in audit analysis
3. Cost reduction recommendations sustained during contract negotiations

Multiple regression can involve any number of independent variables. To solve the audit example above, we would use a three-variable linear equation -- two independent variables and one dependent variable.

Regression Results

Predictor Variable	Equation	r^2
Computer Hours	$Y = A + BX_1$	.82
Field Audit Hours	$Y = A + B X_2$	.60
Comp Hrs and Field Audit Hrs	$Y = A + B_1X_1 + B_2X_2$	.88

D

User Instructions: Select Next to continue.

9 of 53 Back Next

Using Regression Analysis

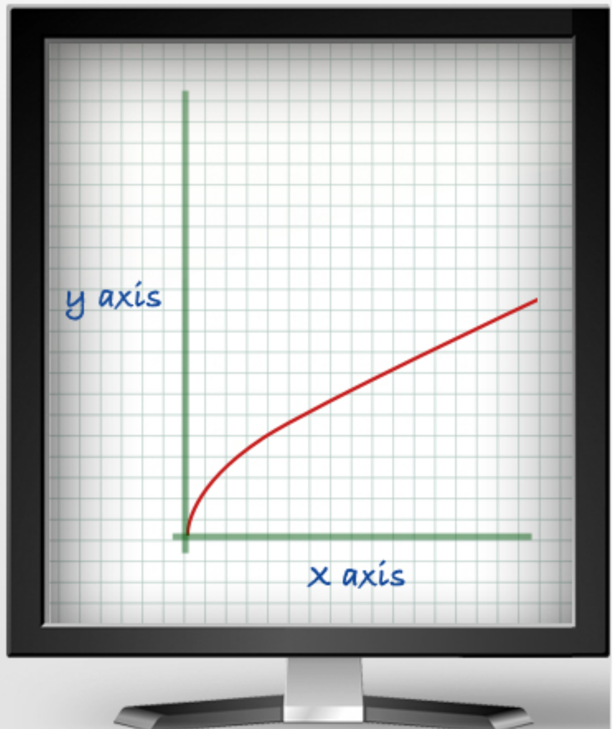
Resources Glossary Help

Curvilinear Regression

In some cases, the relationship between the independent variable(s) may not be linear. Instead, a graph of the relationship on ordinary graph paper would depict a curve.

You cannot directly fit a line to a curve using regression analysis.

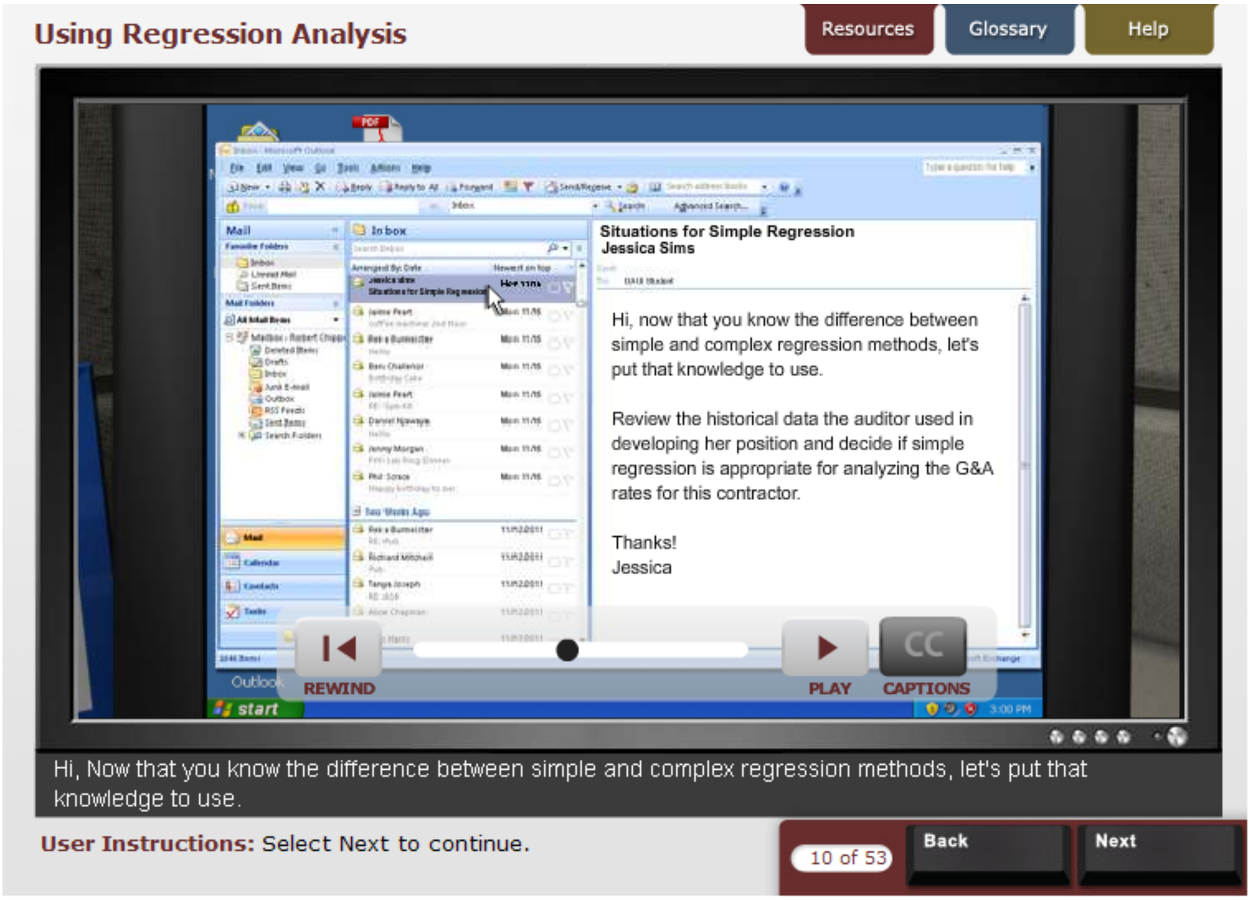
However, if you can identify a quantitative function that transforms a graph of the data to a linear relationship, you can then use regression analysis to calculate a line of best fit for the transformed data.

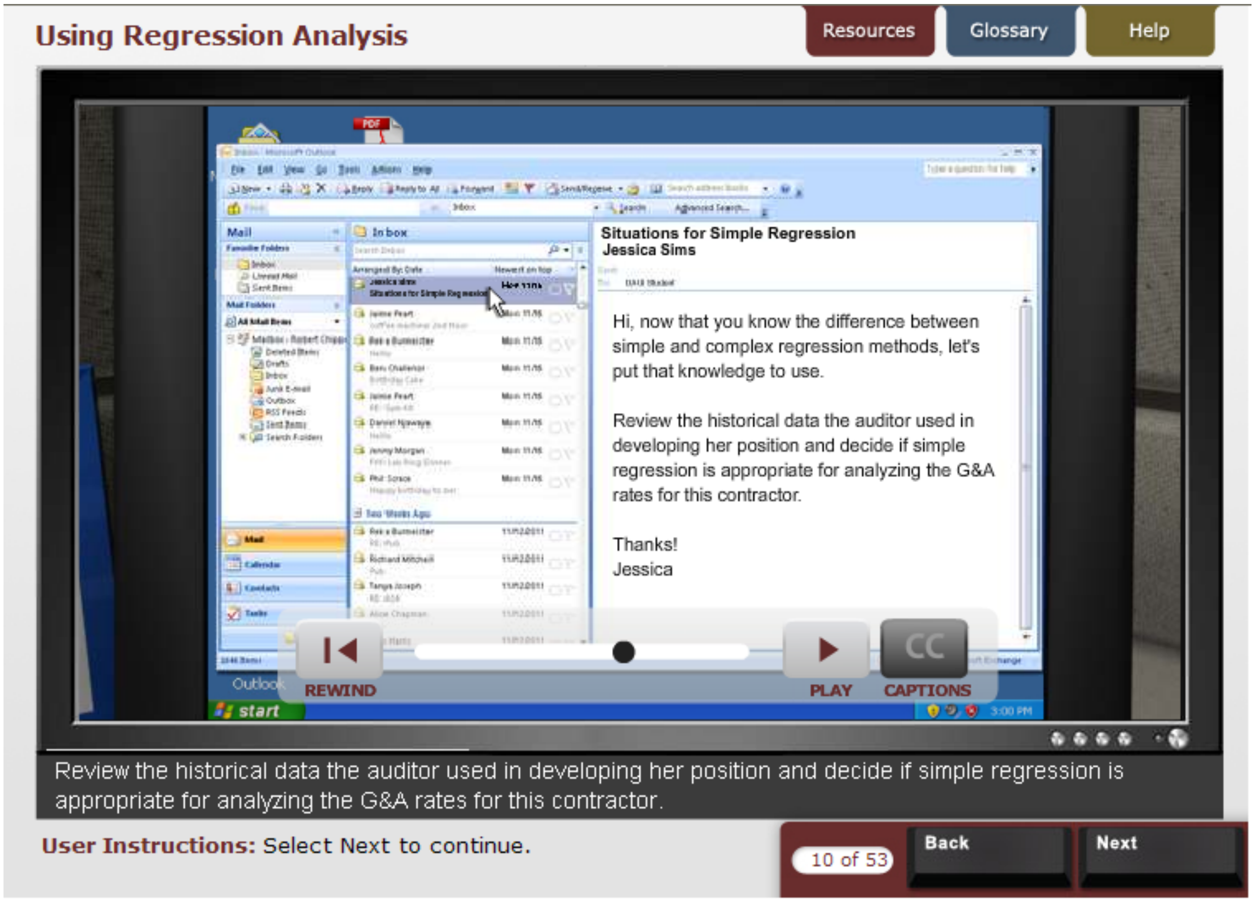


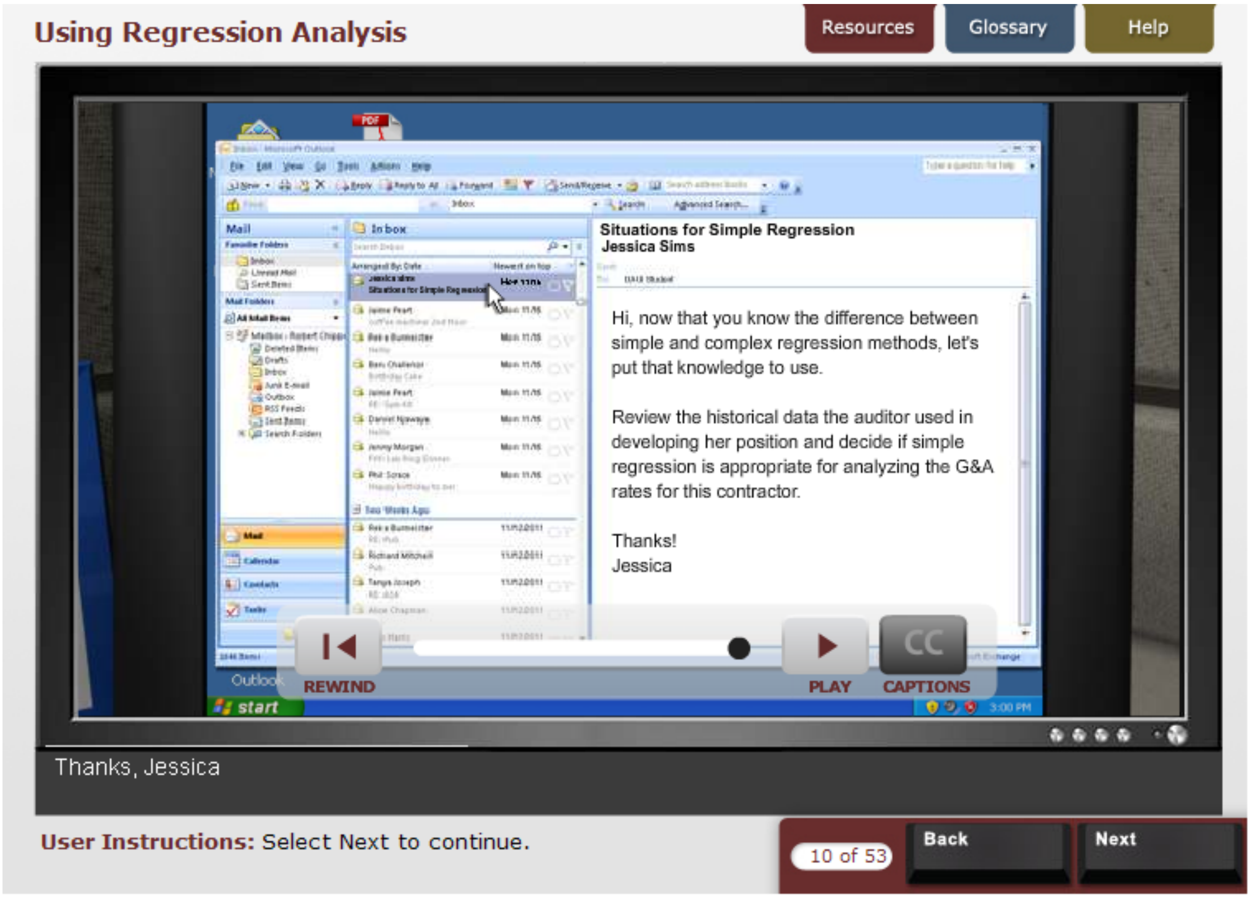
D

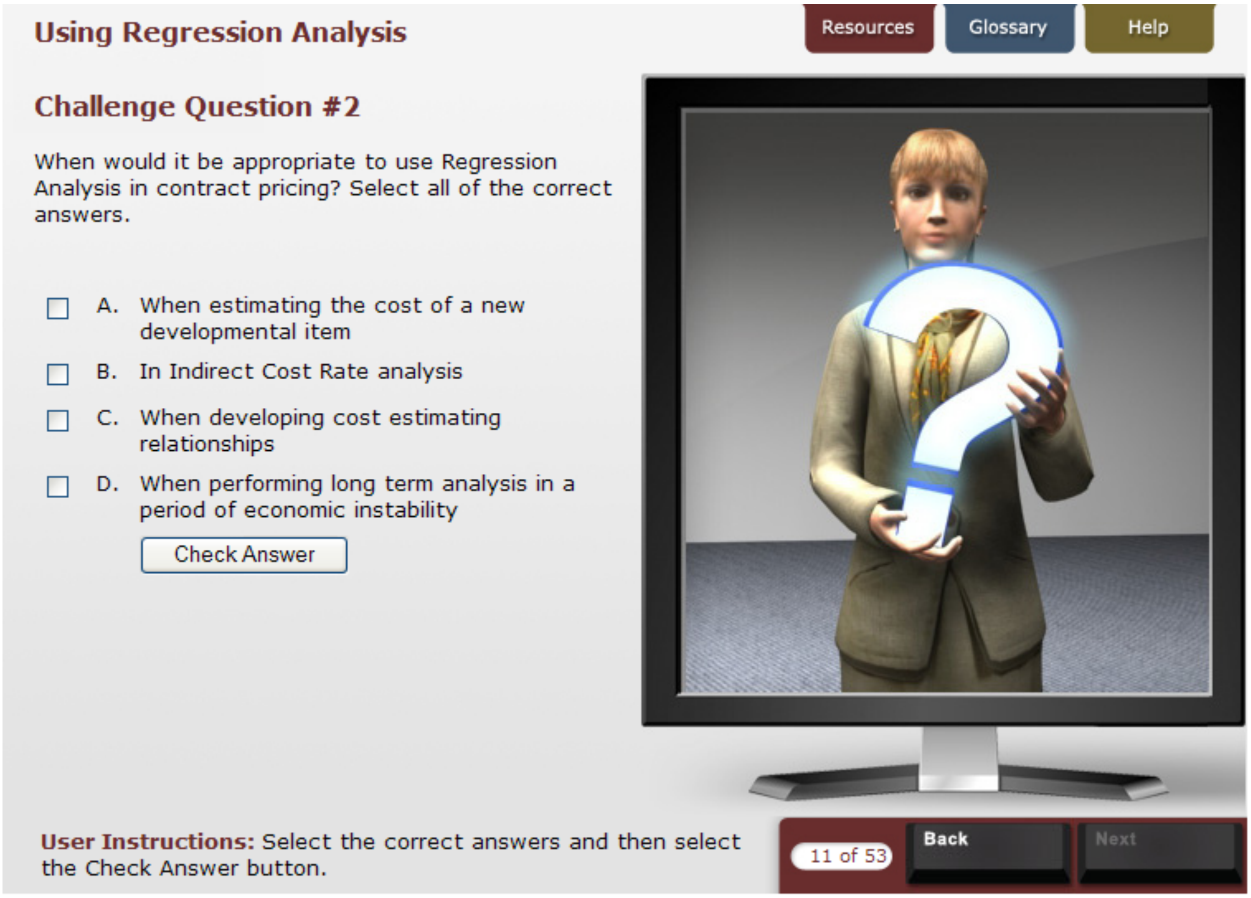
User Instructions: Select Next to continue.

9 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
Situations for Simple Regression
Jessica Sims
Hi, now that you know the difference between simple and complex regression methods, let's put that knowledge to use.
Review the historical data the auditor used in developing her position and decide if simple regression is appropriate for analyzing the G&A rates for this contractor.
Thanks!
Jessica
REWIND
PLAY
CAPTIONS
Hi, Now that you know the difference between simple and complex regression methods, let's put that knowledge to use.
User Instructions: Select Next to continue.
10 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
Situations for Simple Regression
Jessica Sims
Hi, now that you know the difference between simple and complex regression methods, let's put that knowledge to use.
Review the historical data the auditor used in developing her position and decide if simple regression is appropriate for analyzing the G&A rates for this contractor.
Thanks!
Jessica
REWIND
PLAY
CAPTIONS
Review the historical data the auditor used in developing her position and decide if simple regression is appropriate for analyzing the G&A rates for this contractor.
User Instructions: Select Next to continue.
10 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
Situations for Simple Regression
Jessica Sims
Hi, now that you know the difference between simple and complex regression methods, let's put that knowledge to use.
Review the historical data the auditor used in developing her position and decide if simple regression is appropriate for analyzing the G&A rates for this contractor.
Thanks!
Jessica
REWIND
PLAY
CC
CAPTIONS
Thanks, Jessica
User Instructions: Select Next to continue.
10 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
Challenge Question #2
When would it be appropriate to use Regression Analysis in contract pricing? Select all of the correct answers.
A. When estimating the cost of a new developmental item
B. In Indirect Cost Rate analysis
C. When developing cost estimating relationships
D. When performing long term analysis in a period of economic instability
Check Answer
User Instructions: Select the correct answers and then select the Check Answer button.
11 of 53
Back
Next

Using Regression Analysis

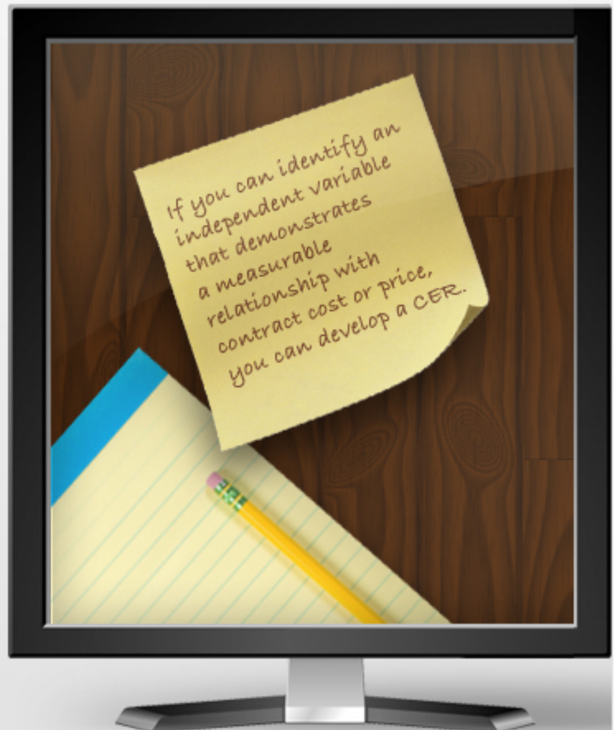
Resources Glossary Help

CER Development and Analysis

Regression analysis is one of the techniques most commonly used to establish cost estimating relationships (CERs) between independent variables and cost or price.

As the name implies, a CER is a technique used to estimate a particular cost or price by using an established relationship with an independent variable.

When you use regression analysis to quantify a CER, you can then use that CER to develop and analyze estimates of product cost or price.



D

User Instructions: Select Next to continue.

11 of 53 Back Next

Resources Glossary Help

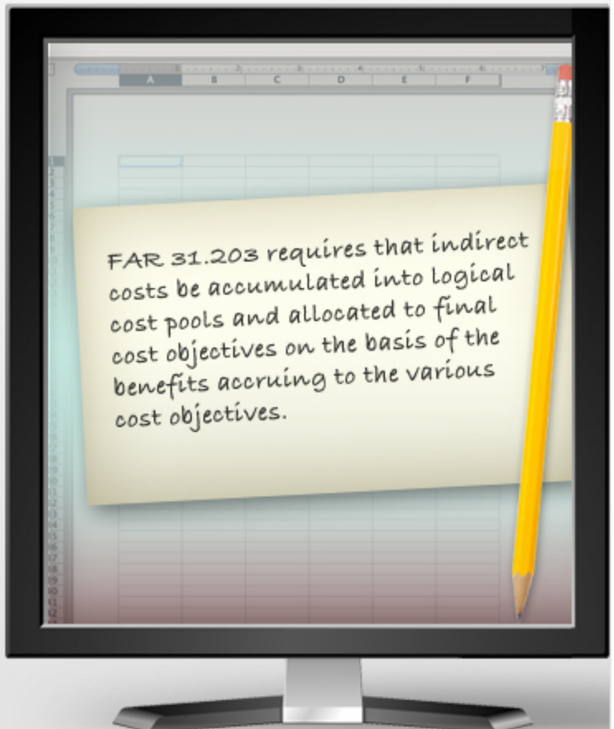
Using Regression Analysis

Indirect Cost Rate Analysis

Indirect costs are costs that are not directly identified with a single final cost objective (e.g., a contract). They are identified with two or more final cost objectives or an intermediate cost objective(s).

Regression analysis is commonly used to quantify the relationship between the indirect cost allocation base and expense pool over time.

If you can quantify the relationship, you can then use that relationship to develop or analyze indirect cost rate estimates.



D

User Instructions: Select Next to continue.

11 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

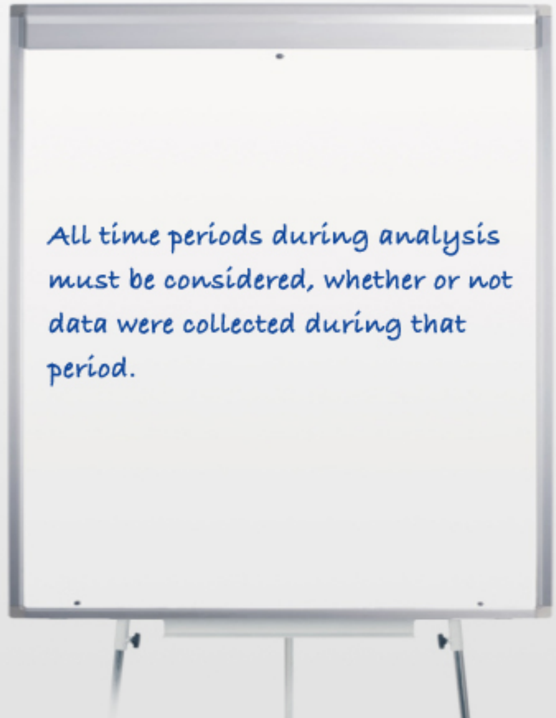
Time-Series Analysis

You can also use regression analysis to analyze trends that appear to be related to time. It is particularly useful when you can identify and adjust for other factors that affect costs or prices to isolate the effect of inflation/deflation for analysis.

The most common applications of this type are forecasting future wage rates, material costs, and product prices.

In time-series analysis, cost or price data are collected over time for analysis. An estimating equation is developed using time as the independent variable.

The time periods are normally weeks, months, quarters, or years. Each time period is assigned a number (e.g., the first month is 1, the fourth month is 4, etc.).



D

User Instructions: Select Next to continue.

11 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

Time-Series Analysis (cont.)

Time does not cause costs or prices to change. Changes are caused by a variety of economic factors.

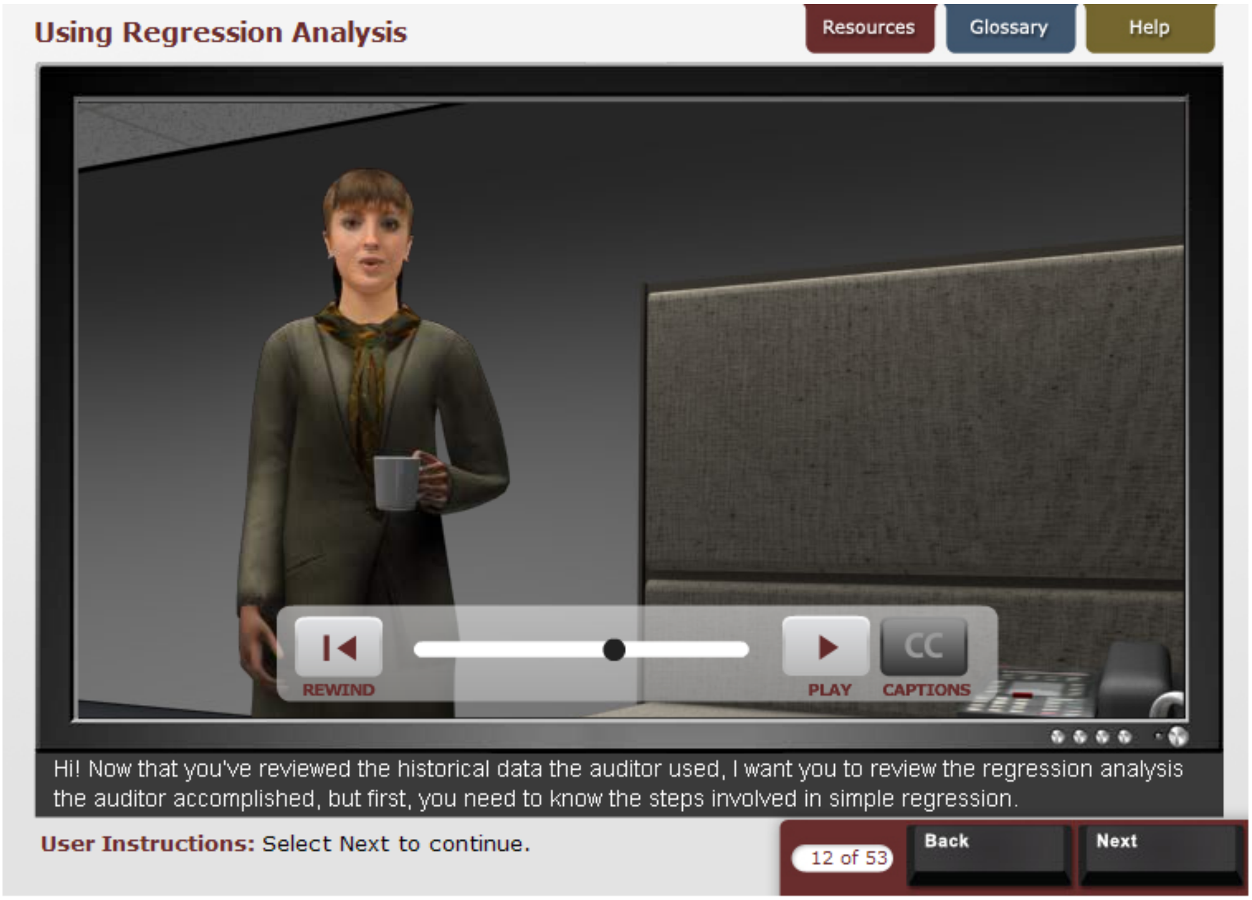
Do not use time-series analysis when you can identify and effectively measure the factors that are driving costs or prices. If you can identify and measure one or more key factors, you should be able to develop a better predictive model than by simply analyzing cost or price changes over time.

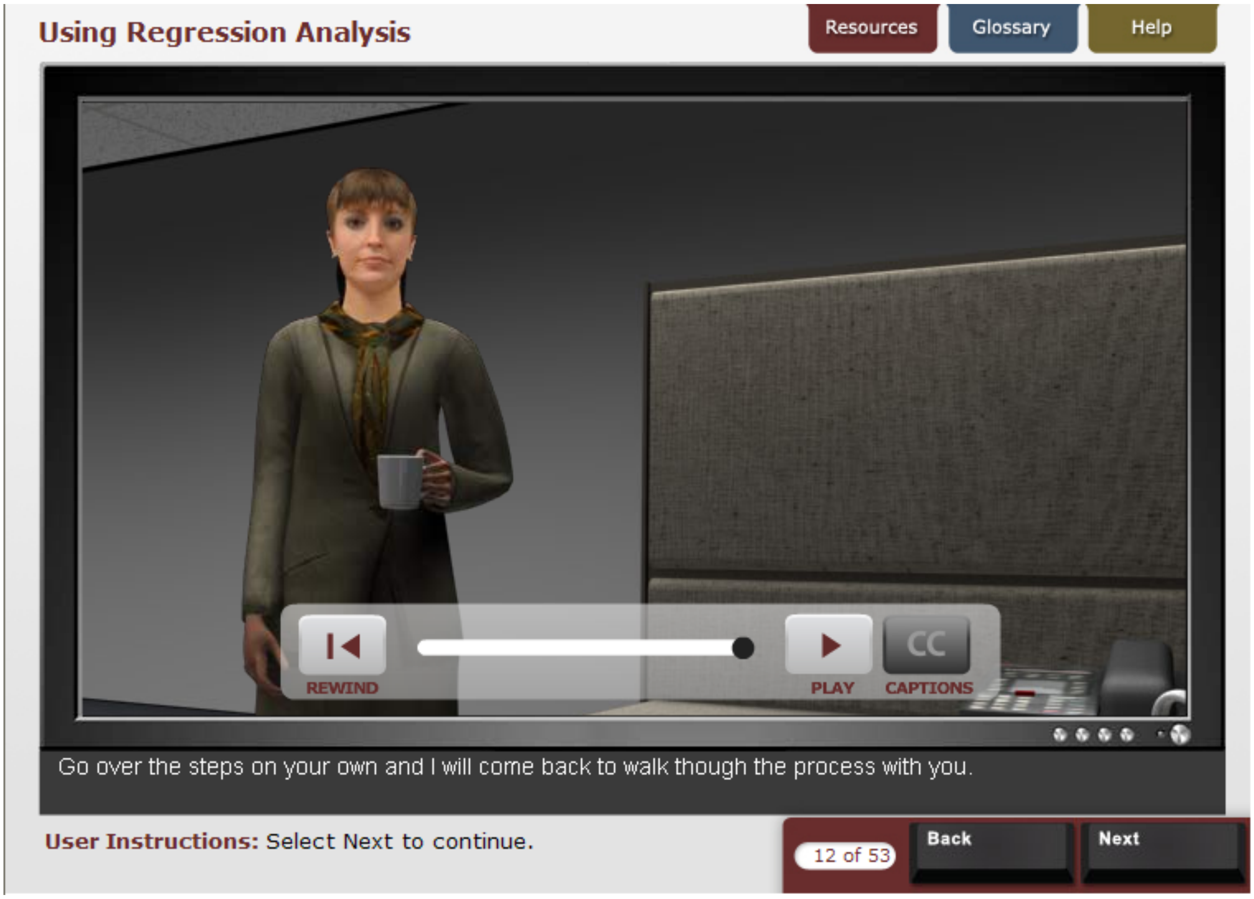
However, if you cannot practically identify or measure such factors, you can often make useful predictions by using regression analysis to analyze cost or price trends over time.

D

User Instructions: Select Next to continue.

11 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
CC
Hi! Now that you've reviewed the historical data the auditor used, I want you to review the regression analysis the auditor accomplished, but first, you need to know the steps involved in simple regression.
User Instructions: Select Next to continue.
12 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Go over the steps on your own and I will come back to walk though the process with you.
User Instructions: Select Next to continue.
12 of 53
Back
Next

Using Regression Analysis

Resources Glossary Help

Challenge Question #3

What is the first step of a simple regression analysis?

- A. Compute the sample mean for X and Y
- B. Put the data in tabular form
- C. Collect the historical data required for analysis
- D. Formulate the estimating equation

Check Answer

User Instructions: Select the correct answer.

13 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

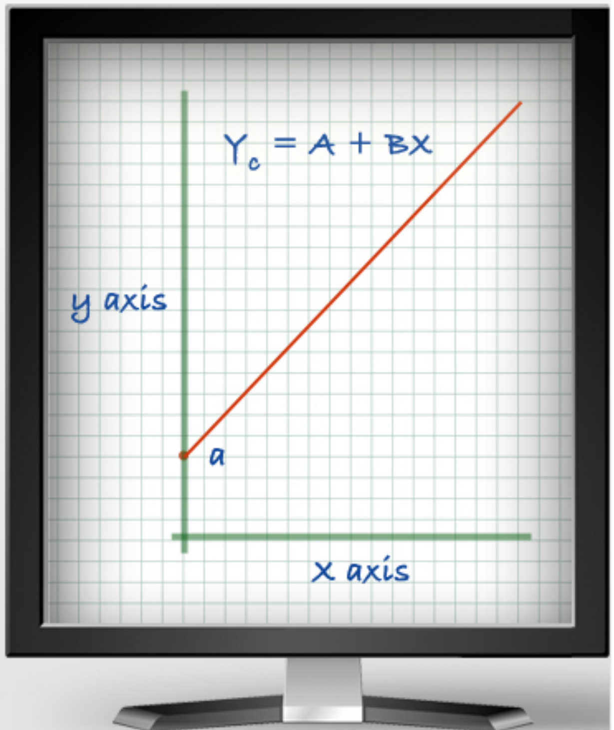
Developing and Using a Simple Regression Equation

The simple regression model is based on the equation for a straight line: $Y_c = A + BX$

- Y_c = The calculated or estimated value for the dependent (response) variable
- A = The Y intercept, the theoretical value of Y when X = 0
- X = The independent (explanatory) variable
- B = The slope of the line (the change in Y divided by the change in X, i.e., the value by which Y changes when X changes by one).

For a given data set, A and B are constants. They do not change as the value of the independent variable changes. Y_c is a function of X.

Specifically, the functional relationship between Y_c and X is that Y_c is equal to A plus the product of B times X.



D

User Instructions: Select Next to continue.

13 of 53 Back Next

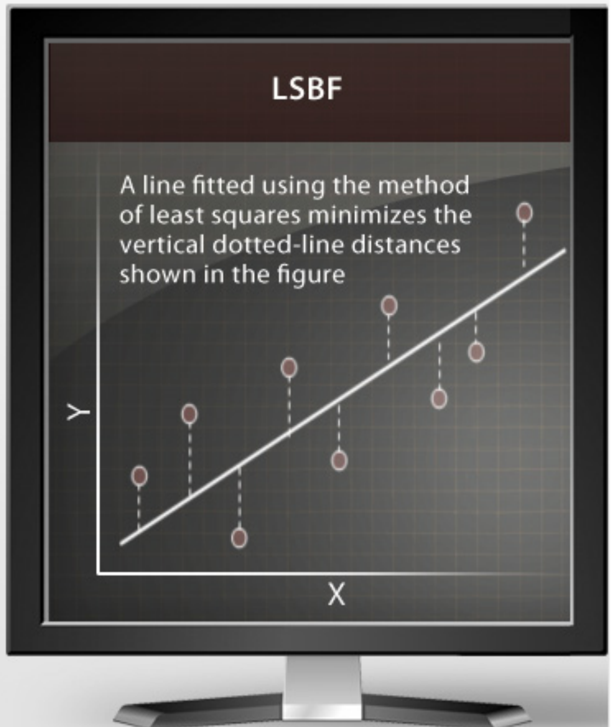
Using Regression Analysis

Resources Glossary Help

Developing a Simple Regression Equation

To develop a 2-variable regression equation for a particular set of data, use the following 5-step least-squares-best-fit (LSBF) process:

1. Collect the historical data required for analysis.
2. Put the data in tabular form.
3. Compute the sample mean for X and Y.
4. Compute the slope (B) and the Y intercept (A).
5. Formulate the estimating equation.



D

User Instructions: Select Next to continue.

13 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

2-Variable Linear Regression Scenario

The following scenario will be continued through the next several screens. It will be used to illustrate the five steps of developing a simple linear regression equation:

> Assume there is a relationship between the Marvin Company's direct labor hours and manufacturing overhead cost. It is based on the use of direct labor hours as the allocation base for manufacturing overhead.
>
> Develop an estimating equation using direct labor hours as the independent variable and manufacturing overhead cost as the dependent variable.
>
> Estimate the indirect cost pool assuming that 2,100 manufacturing direct labor hours will be needed to meet 20x8 production requirements.

The following historical data was gathered for the Marvin Company:

Year	Direct Labor Hours	Overhead
20x2	1,200	$73,000
20x3	1,500	$97,000
20x4	2,300	$128,000
20x5	2,700	$155,000
20x6	3,300	$175,000
20x7	3,400	$218,000
20x8	2,100 (Est)	$?

User Instructions: Select Next to continue.

D

13 of 53 Back Next

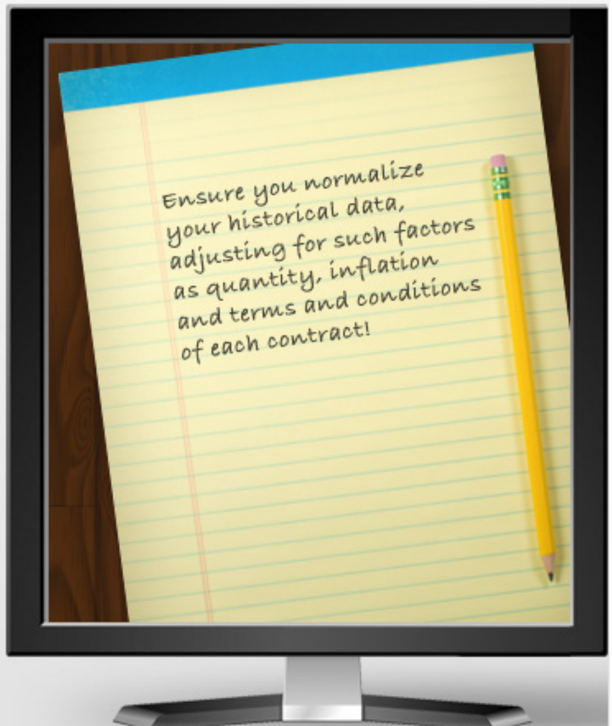
Using Regression Analysis

Resources Glossary Help

Step 1: Collect Historical Data

In this step, you will need to identify the X and Y values for each observation.

- X = Manufacturing Direct Labor Hours (Independent Variable)
- Y = Manufacturing Overhead Cost (Dependent Variable)



D

User Instructions: Select Next to continue.

13 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Step 2: Put Data in Tabular Form

In this step, you will need to enter the historical data gathered in a table.

- X = Manufacturing direct labor hundreds of hours (00s)
- Y = Manufacturing overhead in thousands of dollars ($000s)

Entering the data for the dependent and independent variables and calculating the values for XY, X^2 and Y^2 is necessary in order to compute the sample mean for X and Y in the next step.

	X	Y	XY	X^2	Y^2
	12	73	876	144	5,329
	15	97	1,455	225	9,409
	23	128	2,944	529	16,384
	27	155	4,185	729	24,025
	33	175	5,775	1,089	30,625
	34	218	7,412	1,156	47,524
Totals:	144	846	22,647	3,872	133,296

- X = Manufacturing direct labor hundreds of hours (00s)
- Y = Manufacturing overhead in thousands of dollars ($000s)

D

User Instructions: Select Next to continue.

13 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

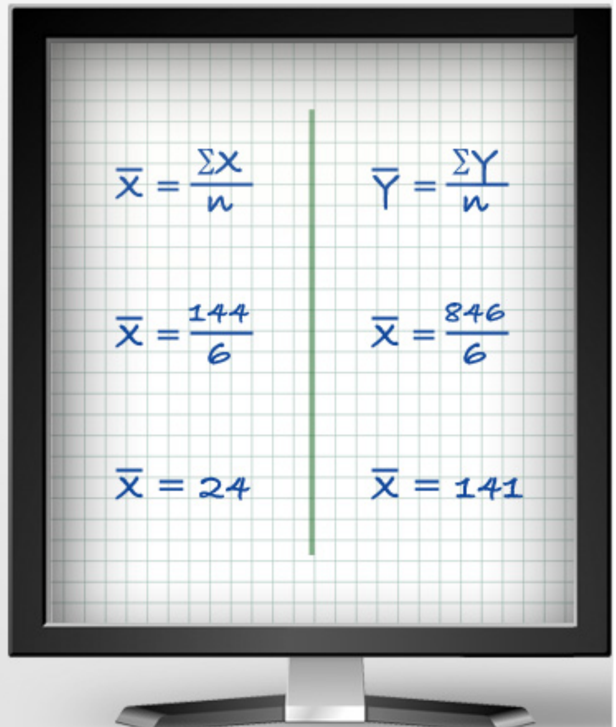
Step 3: Compute Sample Mean X and Y

In this step you must compute the sample mean for X and Y, written as X bar and Y bar, using the following equations:

$$\bar{x} = \frac{\Sigma X}{n} \qquad \bar{y} = \frac{\Sigma Y}{n}$$

- $\bar{x}$ = Sample mean for observations the independent variable
- $\bar{y}$ = Sample mean for observations the dependent variable
- Σ = Summation of all the variables that follow the symbol (e.g., Σ X represents the sum of all X values)
- n = Total number of observations in the sample

$\bar{x}$ = 144/6 = 24 and $\bar{y}$ = 846/6 = 141.



D

User Instructions: Select Next to continue.

13 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Step 4: Compute Slope and Intercept

In this step, you must calculate the slope (B) and the intercept (A) using the following equations:

$$B = \frac{\Sigma XY - n(\overline{X})(\overline{Y})}{\Sigma X^2 - n(\overline{X})^2}$$

$$A = (\overline{Y}) - B(\overline{X})$$

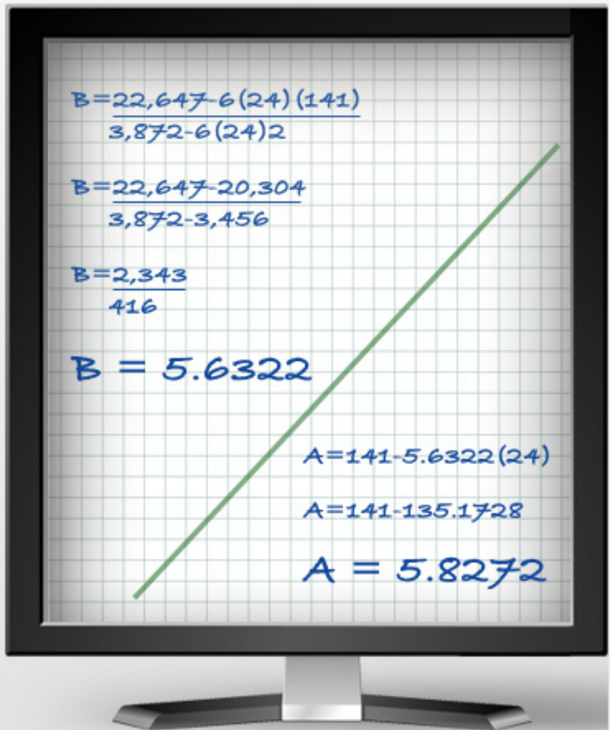
Let's calculate slope first, and then the intercept:

$$B = \frac{22{,}647 - 6(24)(141)}{3{,}872 - 6(24)^2} \qquad A = 141 - 5.6322(24)$$

$$B = \frac{22{,}647 - 20{,}304}{3{,}872 - 3{,}456} \qquad A = 141 - 135.1728$$

$$B = \frac{2{,}343}{416} \qquad A = 5.8272$$

$$B = 5.6322$$



D

User Instructions: Select Next to continue.

13 of 53 | Back | Next

Resources Glossary Help

Using Regression Analysis

Step 5: Formulate the Equation

In this step you must formulate the estimating equation using the following format:

$Y_c = A + BX$

Y_c = Manufacturing overhead ($000's)
X = Manufacturing direct labor hours (00's)

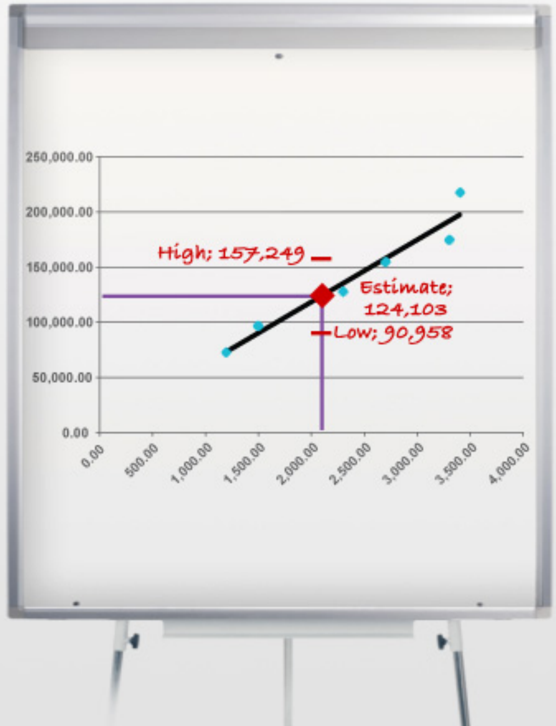
Substitute the calculated values for A and B into the equation:

$Y_c = 5.8272 + 5.6322X$

Estimate manufacturing overhead given the estimate for manufacturing direct labor hours of 2,100:

$Y_c = 5.8272 + 5.6322(21)$
$Y_c = 5.8272 + 118.2762$
$Y_c = 124.1034$

Rounded to the nearest dollar, the estimate



D

User Instructions: Select Next to continue.

13 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Step 5: Formulate the Equation

equation using the following format:

$Y_c = A + BX$

Y_c = Manufacturing overhead ($000's)
X = Manufacturing direct labor hours (00's)

Substitute the calculated values for A and B into the equation:

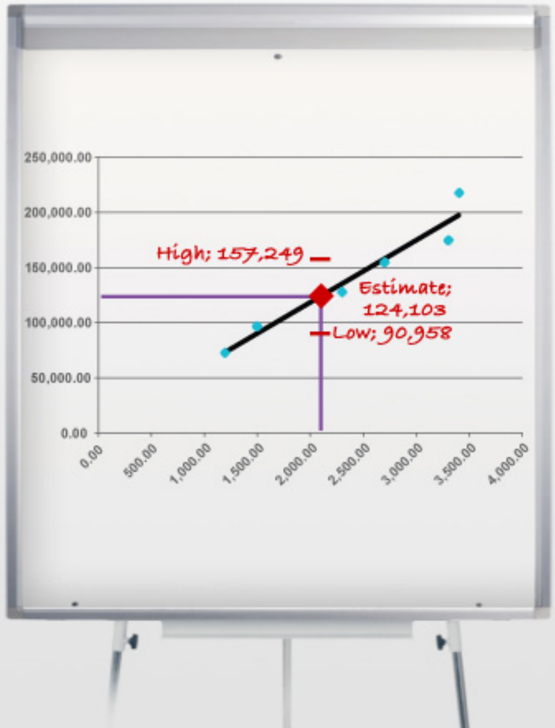
$Y_c = 5.8272 + 5.6322X$

Estimate manufacturing overhead given the estimate for manufacturing direct labor hours of 2,100:

$Y_c = 5.8272 + 5.6322(21)$
$Y_c = 5.8272 + 118.2762$
$Y_c = 124.1034$

Rounded to the nearest dollar, the estimate would be $124,103.

D



User Instructions: Select Next to continue.

13 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Variation in the Regression Model

The purpose of regression analysis is to predict the value of a dependent variable given the value of the independent variable. The LSBF technique yields the best single line to fit the data, but you also want some method of determining how good this estimating equation is.

In order to do this, we must partition the variation into the total variation, variation that is explained by the regression line, and the unexplained variation. We will analyze the variation by summarizing these terms in an Analysis of Variance (ANOVA) Table.

This topic is also discussed in the CPRG, Volume 2, Chapter 5, Section 5.3, Analyzing Variation in the Regression Model.

D

User Instructions: Select Next to continue.

14 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Assumptions of the Regression Model

The assumptions listed on the next screen enable us to calculate unbiased estimators of the population and to use these in predicting values and regression function coefficients (of Y given X).

You should be aware of the fact that violation of one or more of these assumptions reduces the efficiency of the model, but a detailed discussion of this topic is beyond the purview of this module. Assume that all these assumptions have been met.

D

User Instructions: Select Next to continue.

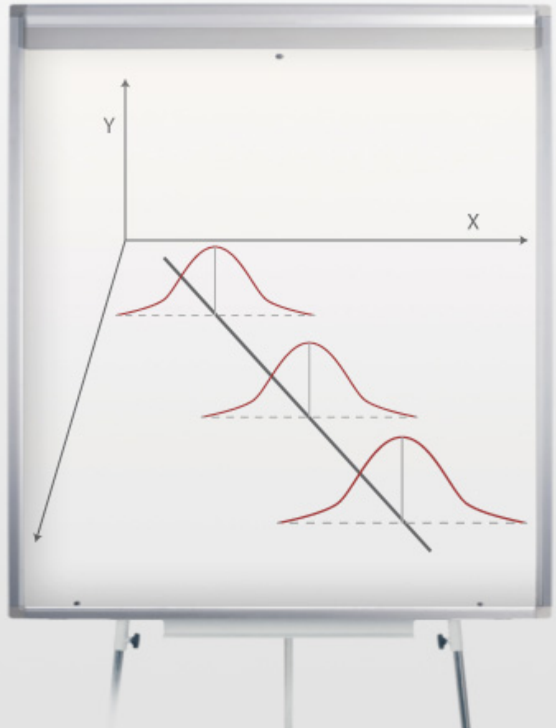
15 of 53 Back Next

Resources Glossary Help

Using Regression Analysis

Assumptions of the Regression Model (cont.)

- For each value of X there is an array of possible Y normally distributed about the regression line.
- The mean of the distribution of possible Y values is on the regression line, i.e., the expected value of the error term is zero.
- The standard deviation of the distribution of possible Y values is constant regardless of the value of X (this is called homoscedasticity).
- The error terms are statistically independent of each other, i.e., there is no serial correlation.
- The error term is statistically independent of X.



D

User Instructions: Select Next to continue.

16 of 53 Back Next

Using Regression Analysis

Variation in the Regression Model - Total Variation

Recall that the purpose of regression analysis is to predict the value of a dependent variable given the value of the independent variable. The LSBF technique yields the best single line to fit the data, but you also want some method of determining how good this estimating equation is. In order to do this, you must first partition the variation.

Total Variation. The sum of squares total (SST) is a measure of the total variation of Y. SST is the sum of the squared differences between the observed values of Y and the mean of Y.

$SST = \Sigma(Y_i - \overline{Y})^2$

Where:

SST = Sum of squared differences
Y_i = Observed value i
$\overline{Y}$ = Mean value of Y

The equation may also be written as: $SST = \Sigma Y^2 - \overline{Y}\Sigma Y$

Total variation can be partitioned into two variations categories: explained and unexplained. This can be expressed as: SST = SSR + SSE

User Instructions: Select Next to continue.

17 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

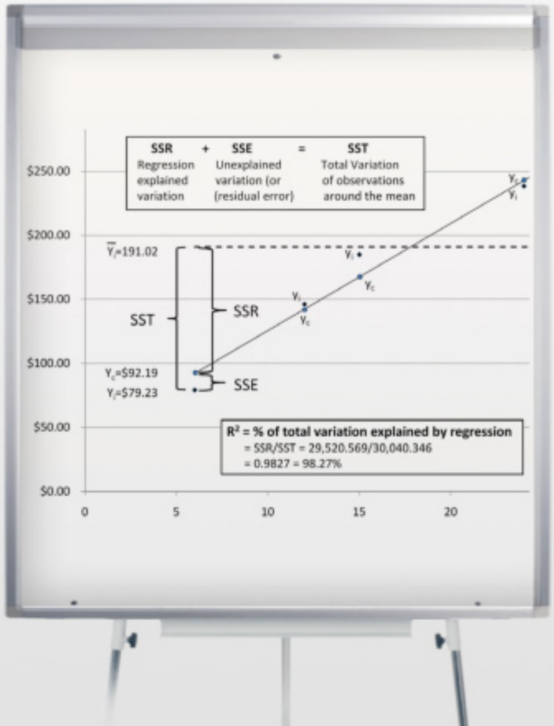
Variation in the Regression Model - Explained Variation

Explained Variation. The sum of squares regression (SSR) is a measure of variation of Y that is explained by the regression equation. SSR is the sum of the squared differences between the calculated value of Y (Yc) and the mean of $\overline{Y}$.

$$SSR = \Sigma(Y_c - \overline{Y})^2$$

You can use the following formula to speed SSR calculation:

$$SSR = B(\Sigma XY - \overline{X}\Sigma Y)$$



D

User Instructions: Select Next to continue.

18 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

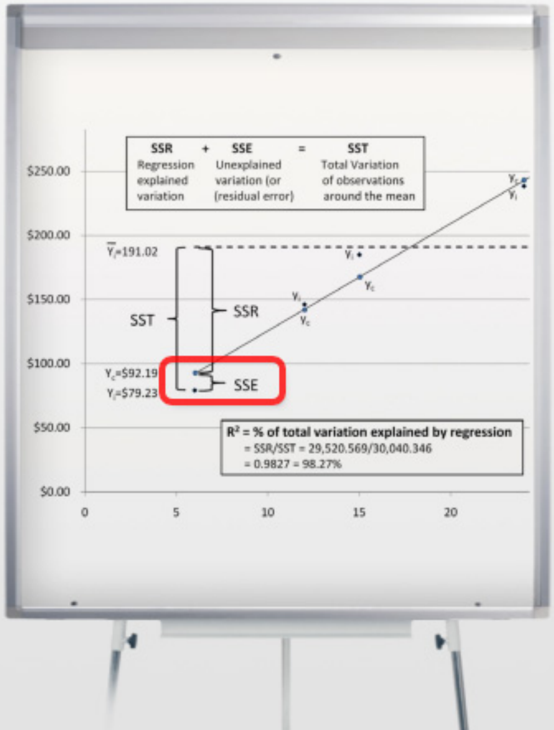
Variation in the Regression Model - Unexplained Variation

Unexplained Variation. The sum of squares error (SSE) is a measure of the variation of Y that is not explained by the regression equations. SSE is the sum of the squared differences between the observed values of Y and the calculated value of Y. This is the random variation of the observations around the regression line.

$SSE = \Sigma(Y_i - Y_c)^2$

You can use the following formula to speed SSE calculation:

$SSE = \Sigma Y^2 - A\Sigma Y - B\Sigma XY$



User Instructions: Select Next to continue.

19 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

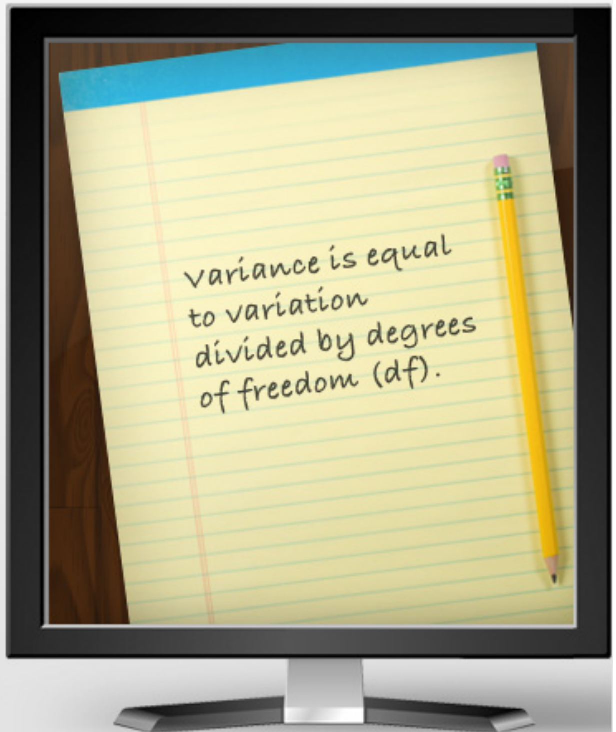
Analysis of Variance - Mean Squares

As previously stated, in order to analyze the variation in the regression model, you have to partition the variation into components that are attributable to different sources.

Thus, you take the total variation and partition it into that which is attributable to the regression equation and the residual error, or unexplained variation.

You then use the sum of squares to form mean squares for the total, explained and unexplained variation, denoted by MST, MSR and MSE, respectively. You calculate the mean squares by dividing the sum of squares by the degrees of freedom.

In regression analysis, degrees of freedom is a statistical concept that is used to adjust for sample bias in estimating the population mean.



D

User Instructions: Select Next to continue.

20 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Analysis of Variance - Mean Squares (cont.)

Mean Square Regression (MSR).

$$MSR = \frac{SSR}{df}$$

For 2-variable linear regression, the value of df for calculating MSR is always one (1). As a result, in 2-variable linear regression, you can simplify the equation for MSR to read:

$$MSR = \frac{SSR}{1} \text{ or } MSR = SSR$$

Mean Square Error (MSE).

$$MSE = \frac{SSE}{df}$$

In 2-variable linear regression, df for calculating MSE is always n - 2. As a result, in simple regression, you can simplify the equation for MSE to read:

$$MSE = \frac{SSE}{n - 2}$$

User Instructions: Select Next to continue.

21 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

ANOVA Table

Analysis of Variance Table. The terms used to analyze variation/variance in the regression model are commonly summarized in an Analysis of Variance (ANOVA) table.

ANOVA Table			
Source	**Sum of Squares**	**df**	**Mean Square****
Regression	SSR	1	MSR
Error	SSE	n-2	MSE
Total	SST	n-1	-
**Mean Square = Sum of Squares/df			

Check SST. Assure that value for SST is equal to SSR plus SSE.

User Instructions: Select Next to continue.

22 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Analysis of Variance - Manufacturing Overhead Examples

Now that you understand how to calculate the variance and variation, and place them in a tabular form in an ANOVA table, let's go through the calculations using the Manufacturing Overhead example.

Since you already calculated these statistics to develop the regression equation to estimate manufacturing overhead, you will begin your calculations with the values in the table to the right.

Statistic	Value
$\sum X$	144
$\sum Y$	846
$\sum XY$	22,647
$\sum X^2$	3,872
$\sum Y^2$	133,296)
$\bar{x}$	24
$\bar{y}$	141
A	5.8272
B	5.6322
n	6

D

User Instructions: Select Next to continue.

23 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Analysis of Variance - MOH Sum of Square Calculations

Step 1. Calculate SST.

$$SST = \Sigma Y^2 - \bar{Y}\Sigma Y$$
$$= 133{,}296 - 141(846)$$
$$= 133{,}296 - 119{,}286$$
$$= 14{,}010$$

Step 2. Calculate SSR.

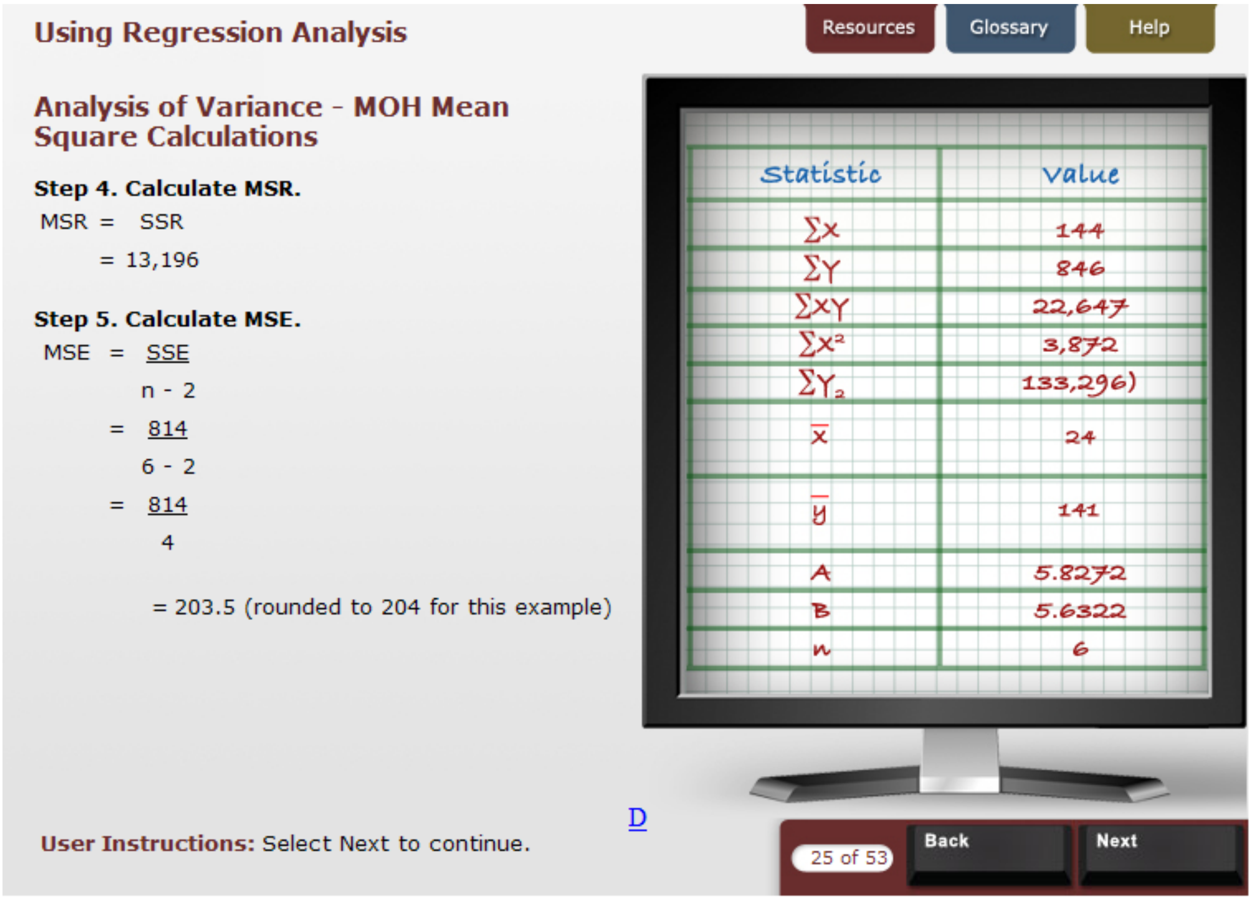
$$SSR = B(\Sigma XY^2 - \bar{X}\Sigma Y)$$
$$= 5.6322[22{,}647 - 24(846)]$$
$$= 5.622[22{,}647 - 20{,}304]$$
$$= 5.6322\ [2{,}343]$$
$$= 13{,}196.24 \text{ (rounded to 13,196 for this example)}$$

Step 3. Calculate SSE.

$$SSE = \Sigma Y^2 - A\Sigma Y - B\Sigma XY$$
$$= 133{,}296 - 5.8272(846) - 5.6322(22{,}647)$$
$$= 133{,}296 - 4929.81 - 127{,}552.43$$
$$= 813.76 \text{ (rounded to 814 for this example)}$$

User Instructions: Select Next to continue.

24 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
Analysis of Variance - MOH Mean Square Calculations
Step 4. Calculate MSR.
MSR = SSR
= 13,196
Step 5. Calculate MSE.
MSE = SSE / n - 2
= 814 / 6 - 2
= 814 / 4
= 203.5 (rounded to 204 for this example)
Statistic Value
ΣX 144
ΣY 846
ΣXY 22,647
ΣX^2 3,872
ΣY_2 133,296)
$\bar{x}$ 24
$\bar{y}$ 141
A 5.8272
B 5.6322
n 6
D
User Instructions: Select Next to continue.
25 of 53
Back
Next

Using Regression Analysis

Resources Glossary Help

Analysis of Variance - MOH ANOVA table

Step 6. Combine the calculated values into an ANOVA table.

ANOVA Table			
Source	**Sum of Squares**	**df**	**Mean Square****
Regression	13,196	1	13,196
Error	814	4	204
Total	14,010	5	-
**Mean Square = Sum of Squares/df			

Step 7. Check SST. Assure that value for SST is equal to SSR plus SSE.

SST = SSR + SSE
14,010 = 13,196 + 814
14,010 = 14,010

User Instructions: Select Next to continue.

26 of 53 Back Next

Using Regression Analysis

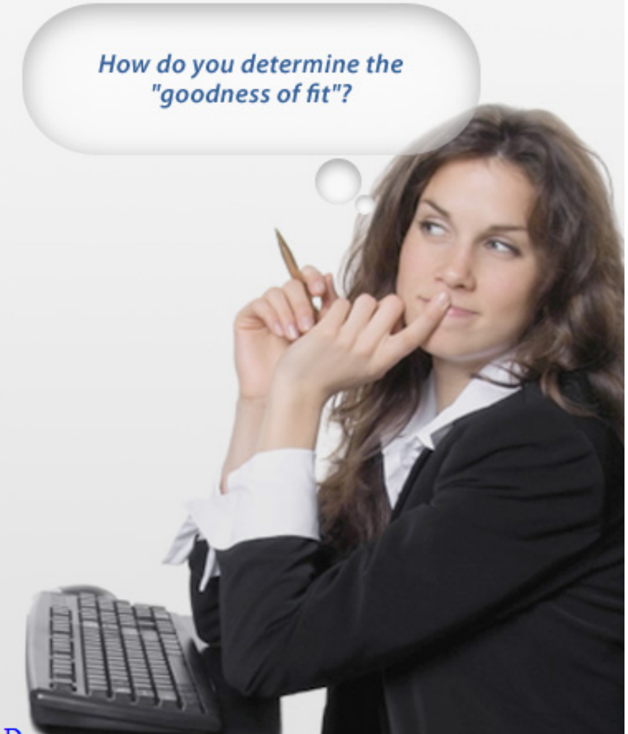
Resources Glossary Help

Measure of Goodness of Fit Statistics

How well does the equation fit the data used in developing the equation? Three statistics are commonly used to determine the "goodness of fit" of the regression equation:

- Coefficient of determination (r-squared)
- Standard error of the estimate (SEE)
- T-test for significance of the regression equation (t-test)

Fit statistics including the r-squared, SEE, and t-test are discussed in the CPRG, Volume 2, Chapter 5, Section 5.4, Goodness of Fit Statistics.



D

User Instructions: Select Next to continue.

27 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

The Coefficient of Determination

Calculating the Coefficient of Determination. Most computer software designed to fit a line using regression analysis will also provide the coefficient of determination for that line. The coefficient of determination (r^2) measures the strength of the association between independent and dependent variables (X and Y).

An r^2 of zero indicates that there is no relationship between X and Y. An r^2 of one indicates that there is a perfect relationship between X and Y. As r^2 gets closer to 1, the better the regression line fits the data set.

In fact, r^2 is the ratio of explained variation (SSR) to total variation (SST). An r^2 of .90 indicates that 90 percent of the variation in Y has been explained by its relationship with X; that is, 90 percent of the variation in Y has been explained by the regression line.

$$r^2 = \frac{SSR}{SST}$$

The range of r^2 is between zero and one ---- $0<r^2<1$

User Instructions: Select Next to continue.

28 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

The Coefficient of Determination - Mfg OH Example

For the manufacturing overhead example:

$$r^2 = \frac{13,196}{14,010} = .94$$

This means that approximately 94% of the variation in manufacturing overhead (Y) can be explained by its relationship with manufacturing direct labor hours (X).

D

User Instructions: Select Next to continue.

29 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

The Standard Error of the Estimate

Standard Error of the Estimate. The standard error of the estimate (SEE) is a measure of the accuracy of the estimating (regression) equation.

The SEE indicates the variability of the observed (actual) points around the regression line (predicted points). That is, it measures the extent to which the observed values (Yi) differ from their calculated values (Y_c).

The SEE is interpreted in a way similar to the way in which the standard deviation is interpreted. That is, given a value for X, we would generally expect the following intervals (based on the Empirical Rule):

Y_c = 1 SEE contains approximately 68 percent of the total observations (Yi)

Y_c = 2 SEE contains approximately 95 percent of the total observations (Yi)

Y_c = 3 SEE contains approximately 99 percent of the total observations (Yi)

User Instructions: Select Next to continue.

30 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

Standard Error of the Estimate - Mfg OH Example

The SEE is equal to the square root of the MSE.

$SEE = \sqrt{MSE}$

For the manufacturing overhead example:

$$
\begin{aligned}
SEE &= \sqrt{204} \\
&= 14.28
\end{aligned}
$$

User Instructions: Select Next to continue.

31 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

T-test for the Significance of the Regression Equation

Steps for Conducting the T-test for the Significance of the Regression Equation.

The regression line is derived from a sample. Because of sampling error, it is possible to get a regression relationship with a rather high r^2 (e.g. greater than 80 percent) when there is no real relationship between X and Y. That is, when there is no statistical significance.

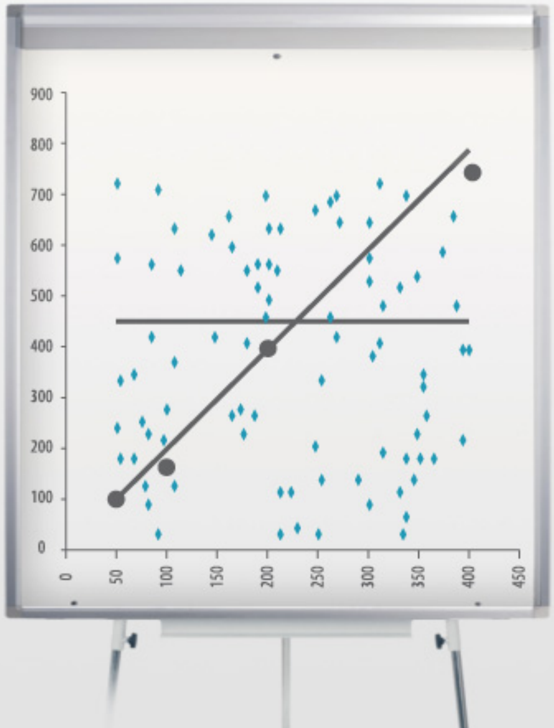
This phenomenon will occur only when you have very small sample data sets. You can test the significance of the regression equation by applying the T-test. Applying the T-test is a 4-step process:

Step 1. Determine the significance level (α).

Step 2. Calculate T.

Step 3. Determine the table value of t.

Step 4. Compare T to the t Table value.



D

User Instructions: Select Next to continue.

32 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

T-test for the Significance of the Regression Equation (cont.)

Step 1. Determine the significance level (α).

α = 1 - confidence level

The selection of the significance level is a management decision; that is, management decides the level of risk associated with an estimate which it will accept. In the absence of any other guidance, use a significance level of .10.

Step 2. Calculate T. Use the values of MSR and MSE from the ANOVA table:

$$T = \sqrt{\frac{MSR}{MSE}}$$

User Instructions: Select Next to continue.

33 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

T-test for the Significance of the Regression Equation (cont.)

Step 3. Determine the table value of t. From a t Table, select the t value for the appropriate degrees of freedom (df). In 2-variable linear regression:

df = n - 2

Step 4. Compare T to the t Table value. Decision rules:

If T > t, use the regression equation for prediction purposes. It is likely that the relationship is significant.

If T < t, do not use the regression equation for prediction purposes. It is likely that the relationship is not significant.

If T = t, a highly unlikely situation, you are theoretically indifferent and may elect to use or not use the regression equation for prediction purposes.

User Instructions: Select Next to continue.

34 of 53 Back Next

Using Regression Analysis

Resources Glossary Help

T-test for Mfg OH Example

Step 1. Determine the significance level (α).
Assume that we have been told to use α = .05.

Step 2. Calculate T.

$$T = \sqrt{\frac{MSR}{MSE}}$$

$$= \sqrt{\frac{13{,}196}{204}}$$

$$= \sqrt{64.69}$$

$$= 8.043$$

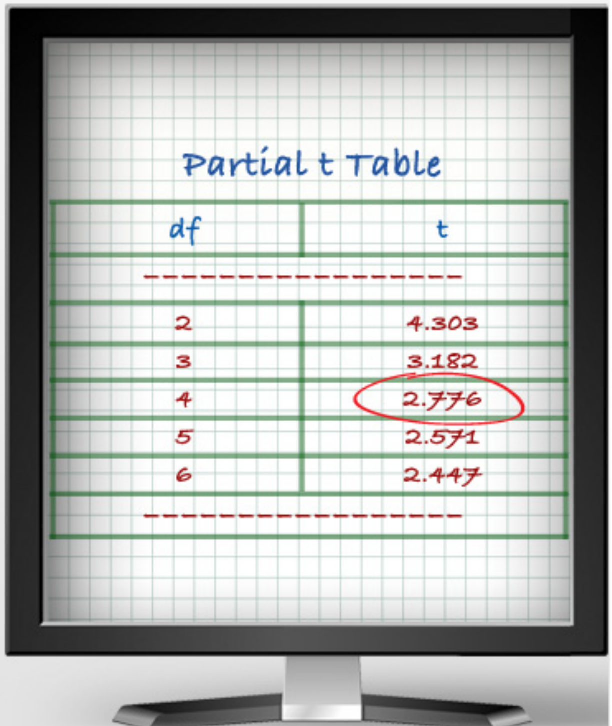
Step 3. Determine the table value of t. The partial table below is an excerpt of a t table.

$$df = n - 2$$

$$= 6 - 2$$

$$= 4$$

Reading from the table, the appropriate value is 2.776.



Partial t Table

df	t
---	---
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
---	---

D

User Instructions: Select Next to continue.

35 of 53 Back Next

Using Regression Analysis

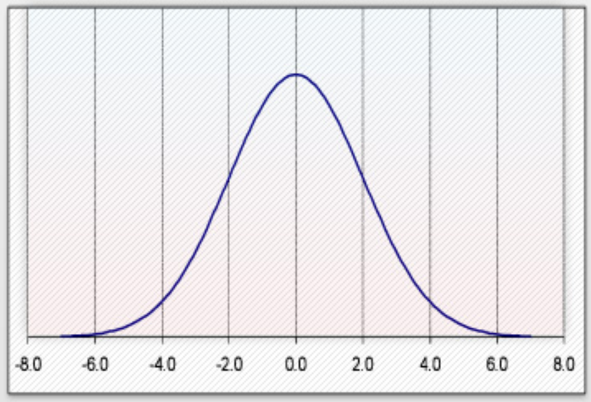
Resources Glossary Help

T-test for Mfg OH Example (cont.)

Step 4. Compare T to the t Table value. Since T (8.043) > t (2.776), use the regression equation for prediction purposes. It is likely that the relationship is significant.

Note: There is not normally a conflict in the decision indicated by the T-test and the magnitude of r^2. If r^2 is high, T is normally > t. A conflict could occur only in a situation where there are very few data points. In those rare instances where there is a conflict, you should accept the decision indicated by the T-test. It is a better indicator than r^2 because it takes into account the sample size (n) through the degrees of freedom (df).

The further the calculated T is away from Zero (0), the more confident we are that the slope is not really zero – the more confident we are that a relationship really exists



D

User Instructions: Select Next to continue.

36 of 53 Back Next

Using Regression Analysis

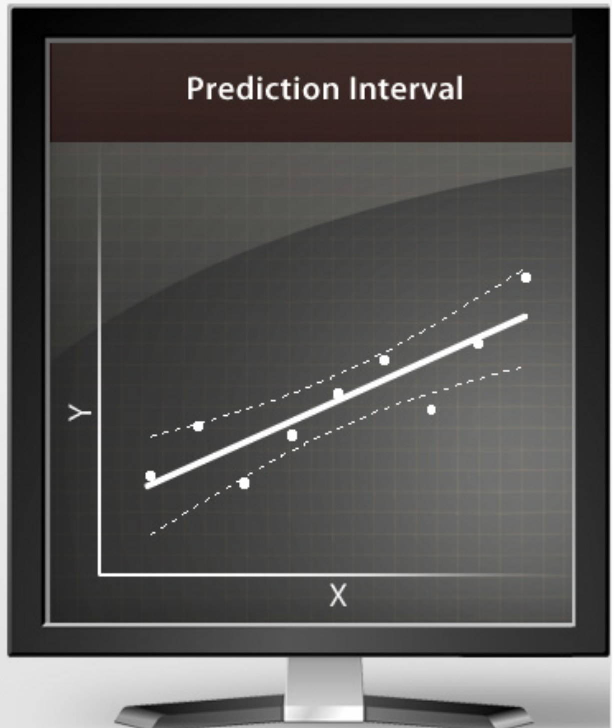
Resources | Glossary | Help

Calculating and Using a Prediction Interval

You can develop a regression equation and use it to calculate a point estimate for Y given any value of X. However, a point estimate alone does not provide enough information for sound negotiations.

You need to be able to establish a range of values which you are confident contains the true value of the cost or price which you are trying to predict. In regression analysis, this range is known as the prediction interval.

The process to construct a prediction interval is discussed in the CPRG, Volume 2, Chapter 5, Section 5.5, Prediction Intervals.



D

User Instructions: Select Next to continue.

37 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

Caculating and Using a Prediction Interval (cont.)

For a regression equation based on a small sample, you should develop a prediction interval, using the following equation:

$$Y_C \pm (SEE)\sqrt{1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum X^2 - n\bar{X}^2}}$$

$$1241034 \pm 2.776\,(14.27)\sqrt{1+\frac{1}{6}+\frac{(21-24)^2}{3{,}872 - 6(24)^2}}$$

$$1241034 \pm 39.6135\sqrt{1+.1667+\frac{(-3)^2}{3{,}872 - 3{,}456}}$$

$$1241034 \pm 39.6135\sqrt{1.1667+\frac{9}{416}}$$

$$1241034 \pm 39.6135\sqrt{1.1667+.0216}$$

$$1241034 \pm 39.6135\sqrt{1.1883}$$

$$1241034 \pm 39.6135\,(1.0901)$$

$$1241034 \pm 43.1827$$

When X = 21 the prediction interval is $80.9207 \leq Y \leq 167.2861$

User Instructions: Select Next to continue.

38 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Caculating and Using a Prediction Interval (cont.)

Constructing a Prediction Interval for the Manufacturing Overhead Example. Assume that you want to construct a 95 percent prediction interval for the manufacturing overhead estimate at 2,100 manufacturing direct labor hours. Earlier in the chapter, you calculated Y_c and the other statistics in the table to the right:

Using the table data, you would calculate the prediction interval as follows: When X = 21 the prediction interval is: $80.9207 < Y < 167.2861$.

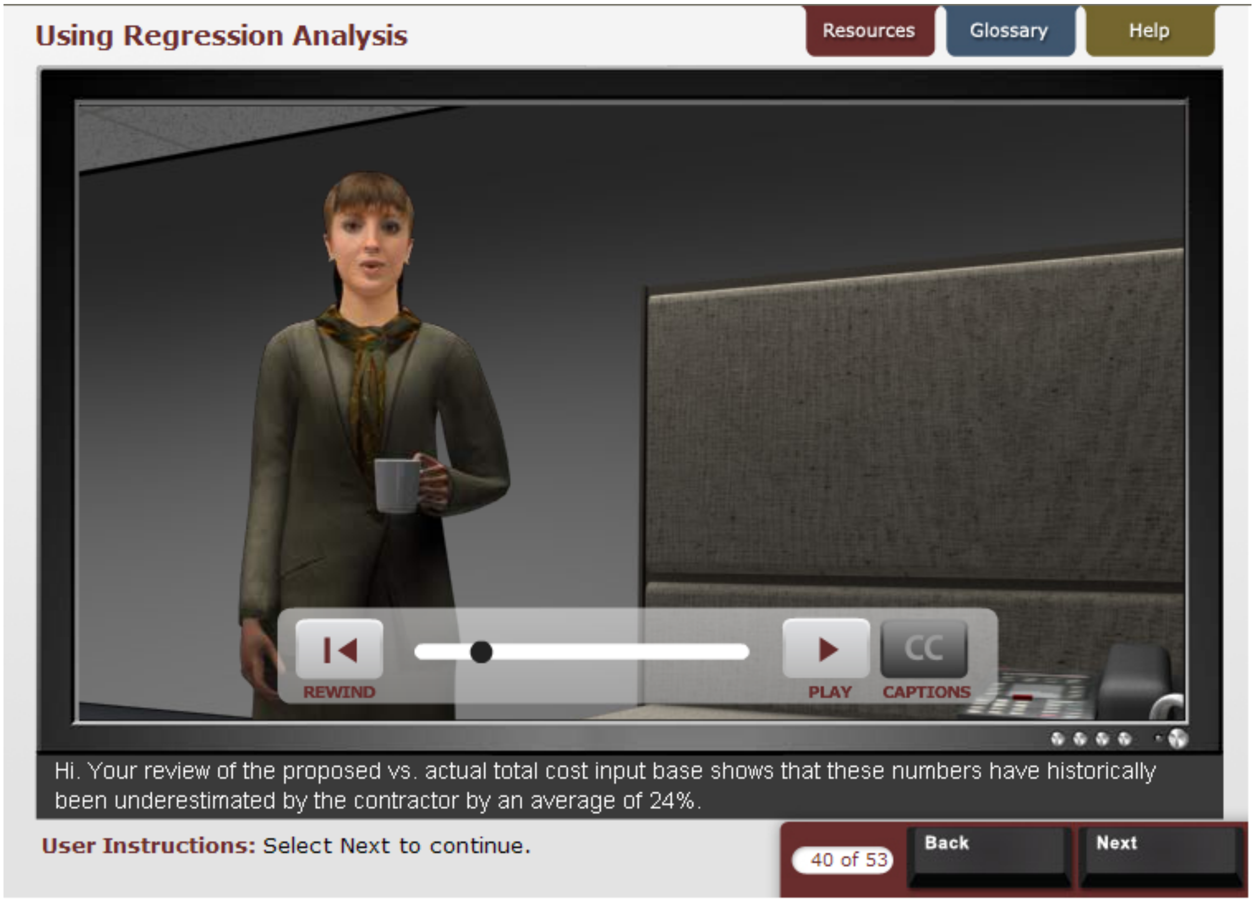
Prediction Statement: You would be 95 percent confident that the actual manufacturing overhead will be between $80,921 and $167,286 at 2,100 manufacturing direct labor hours.

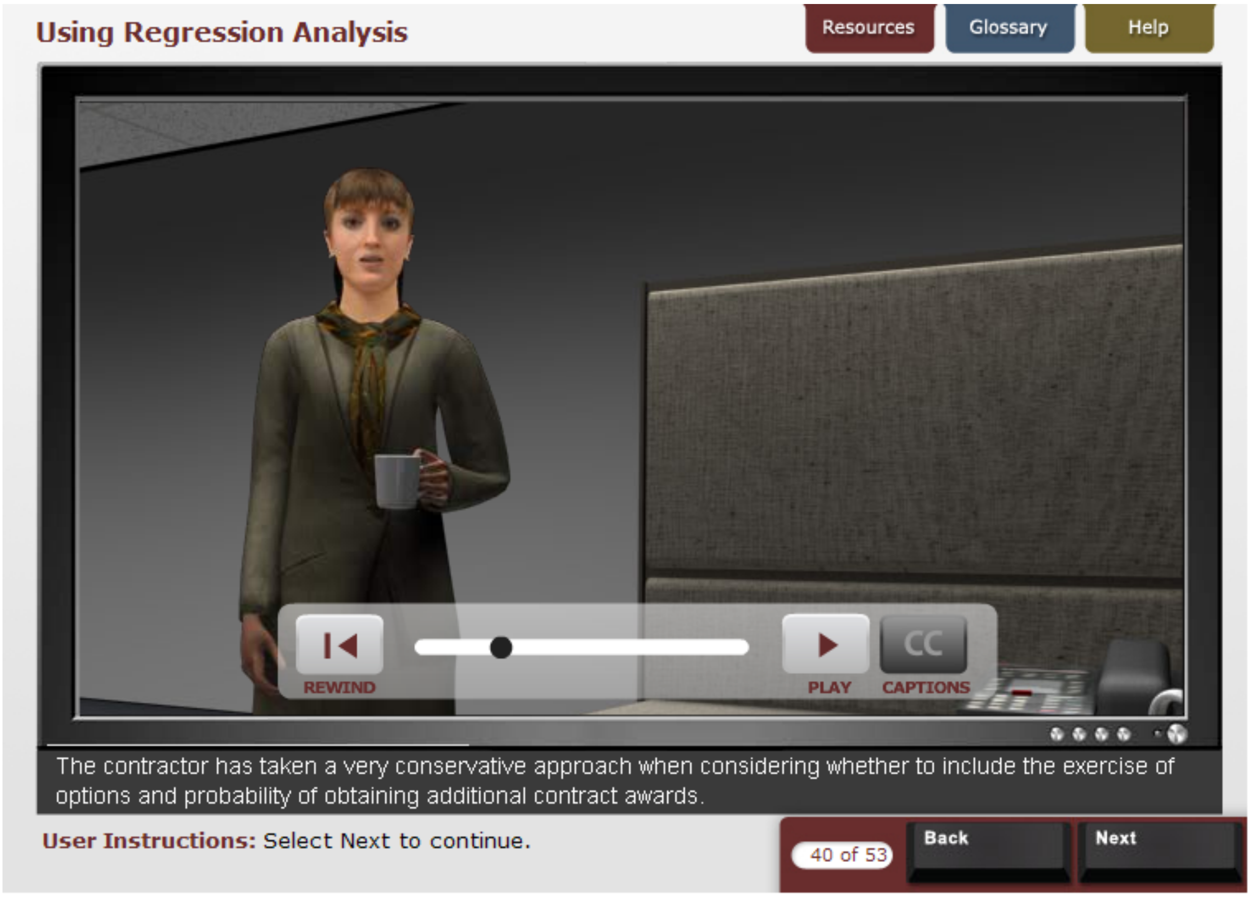
Statistic	Value
Y_c	124.1034
t (use n - 2 df)	2.776
SEE	14.27
$\bar{X}$	24
X^2	3,872

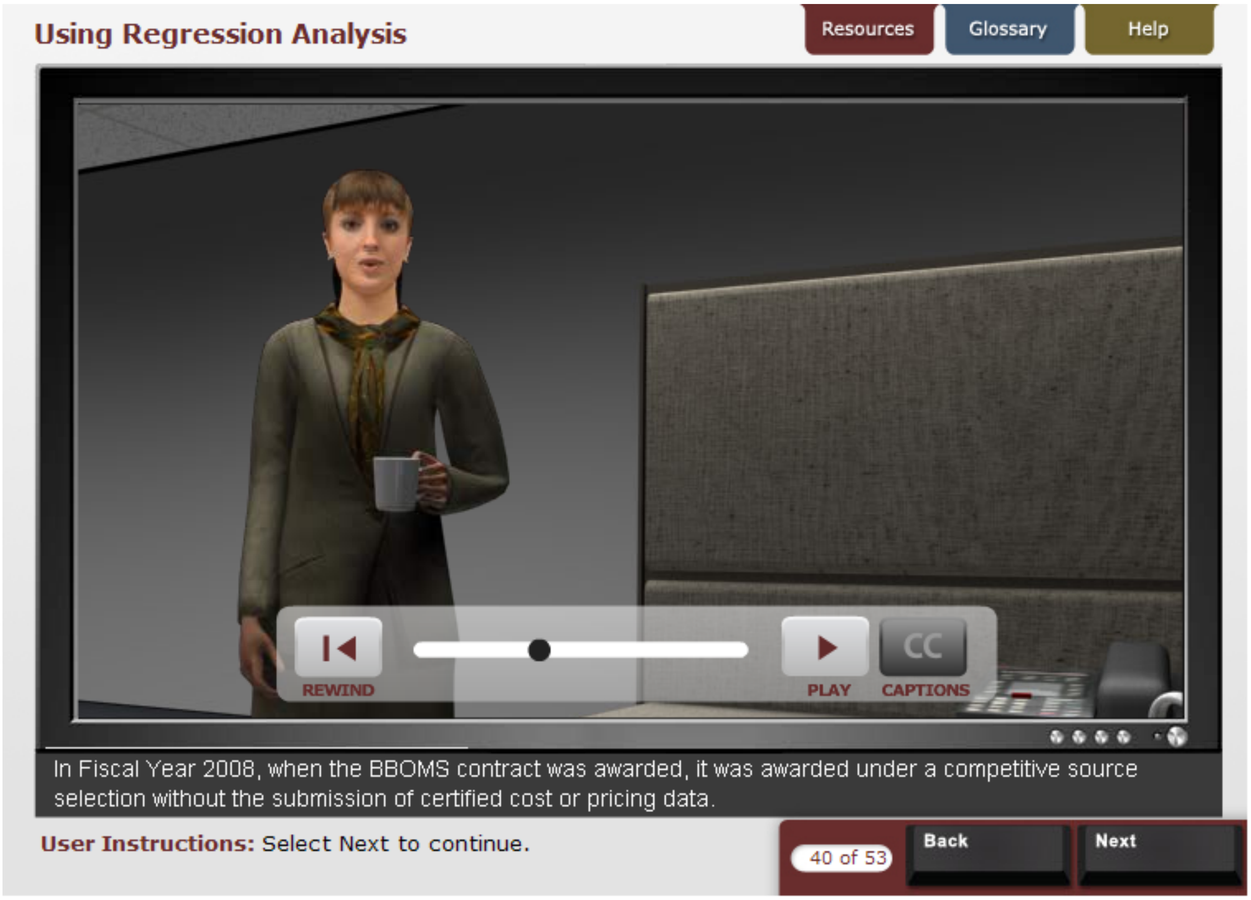
D

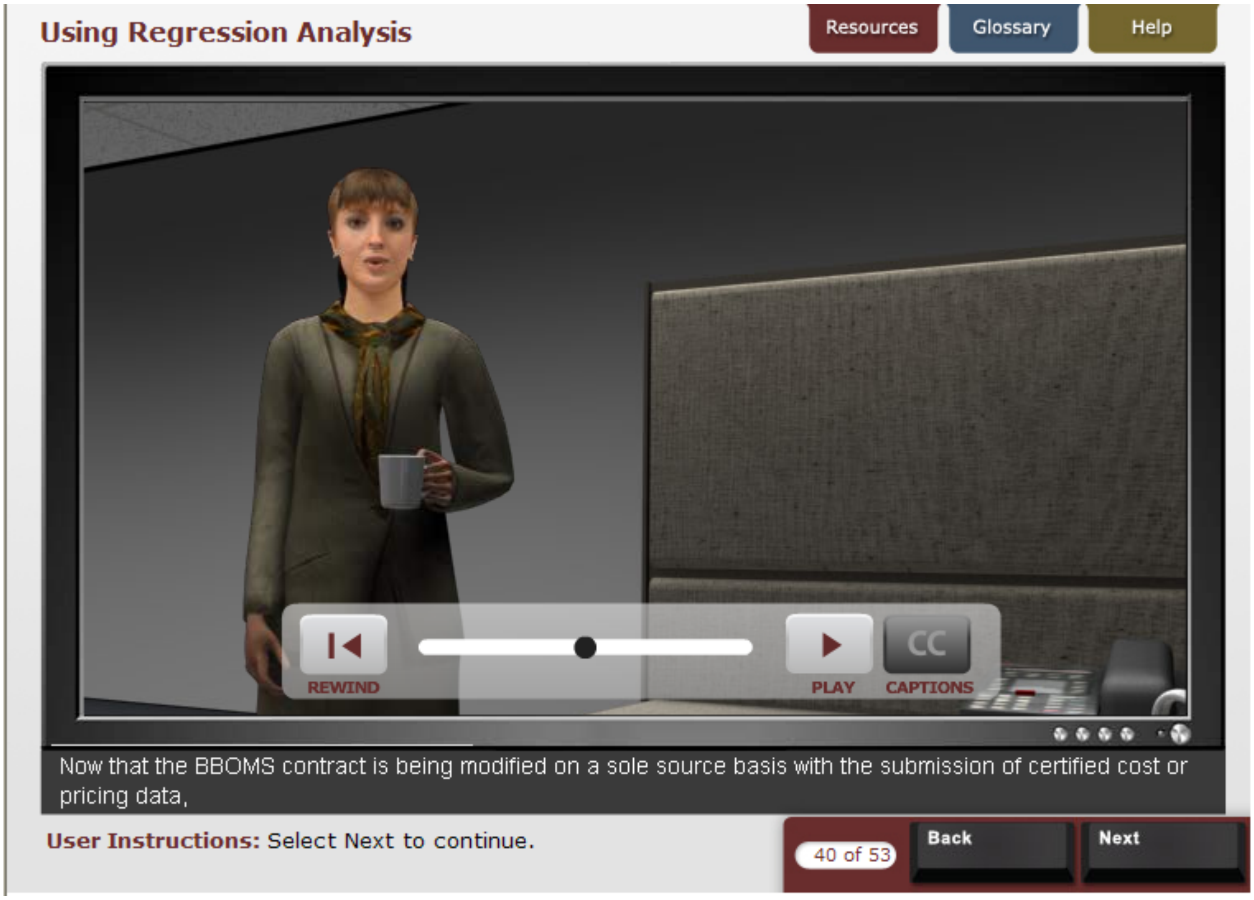
User Instructions: Select Next to continue.

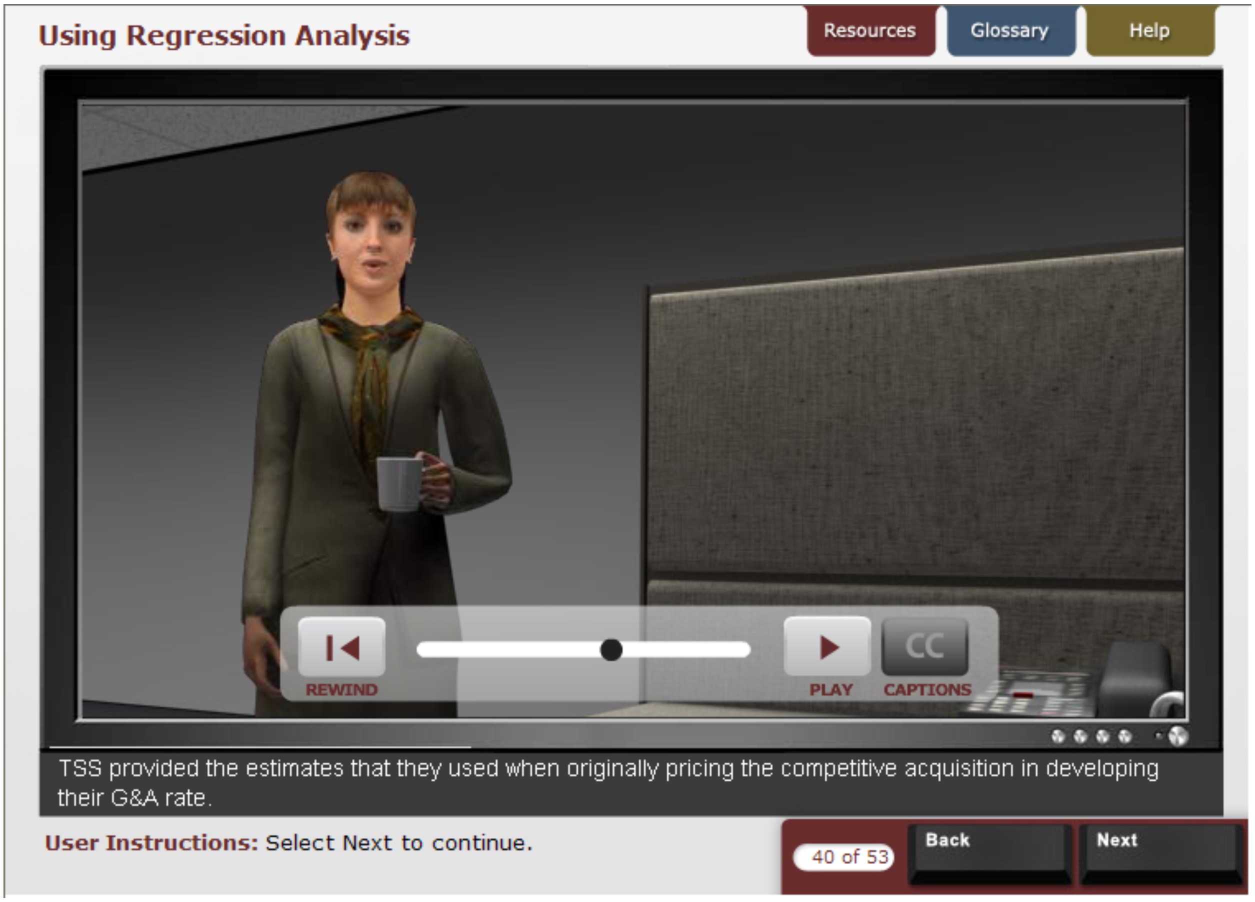
39 of 53 Back Next

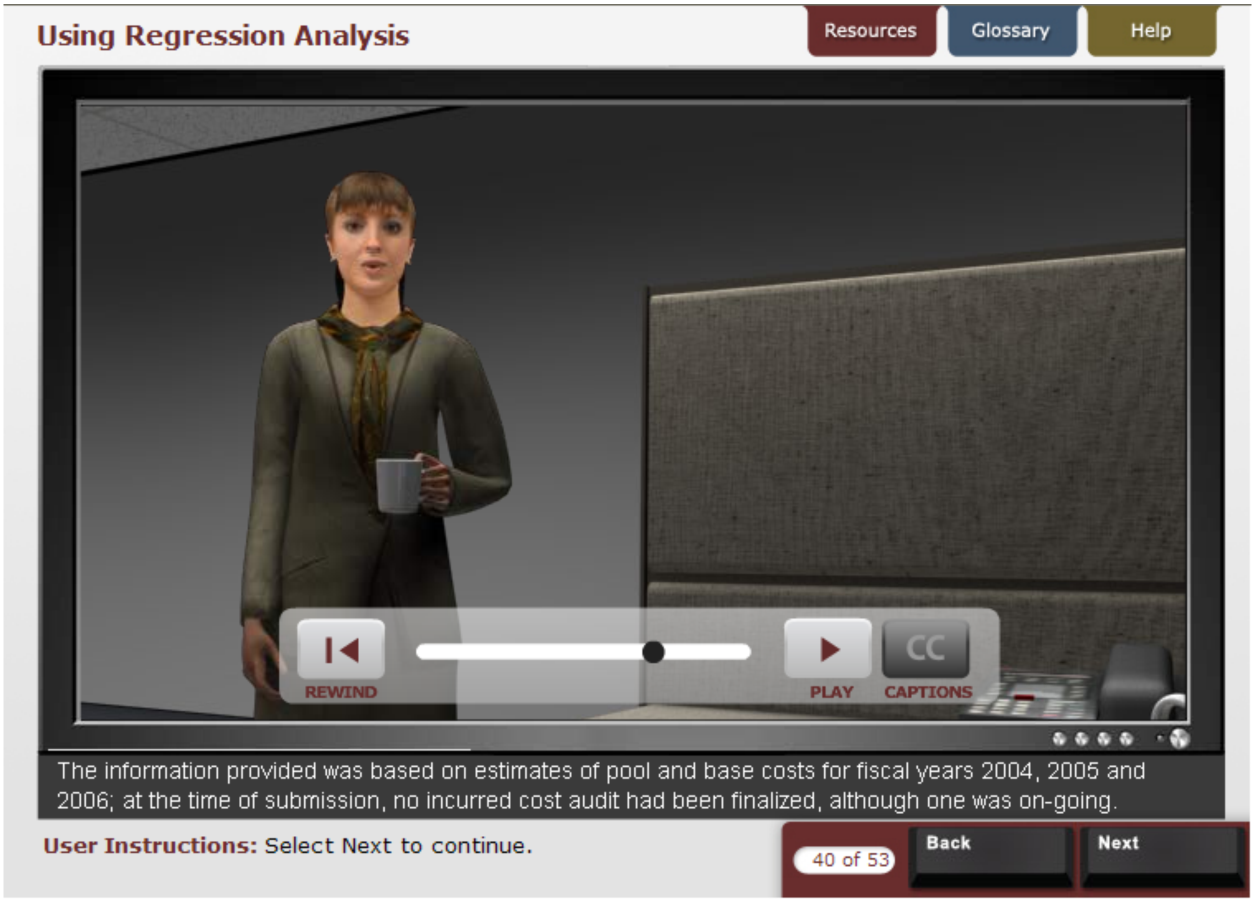

Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CC
CAPTIONS
Hi. Your review of the proposed vs. actual total cost input base shows that these numbers have historically been underestimated by the contractor by an average of 24%.
User Instructions: Select Next to continue.
40 of 53
Back
Next

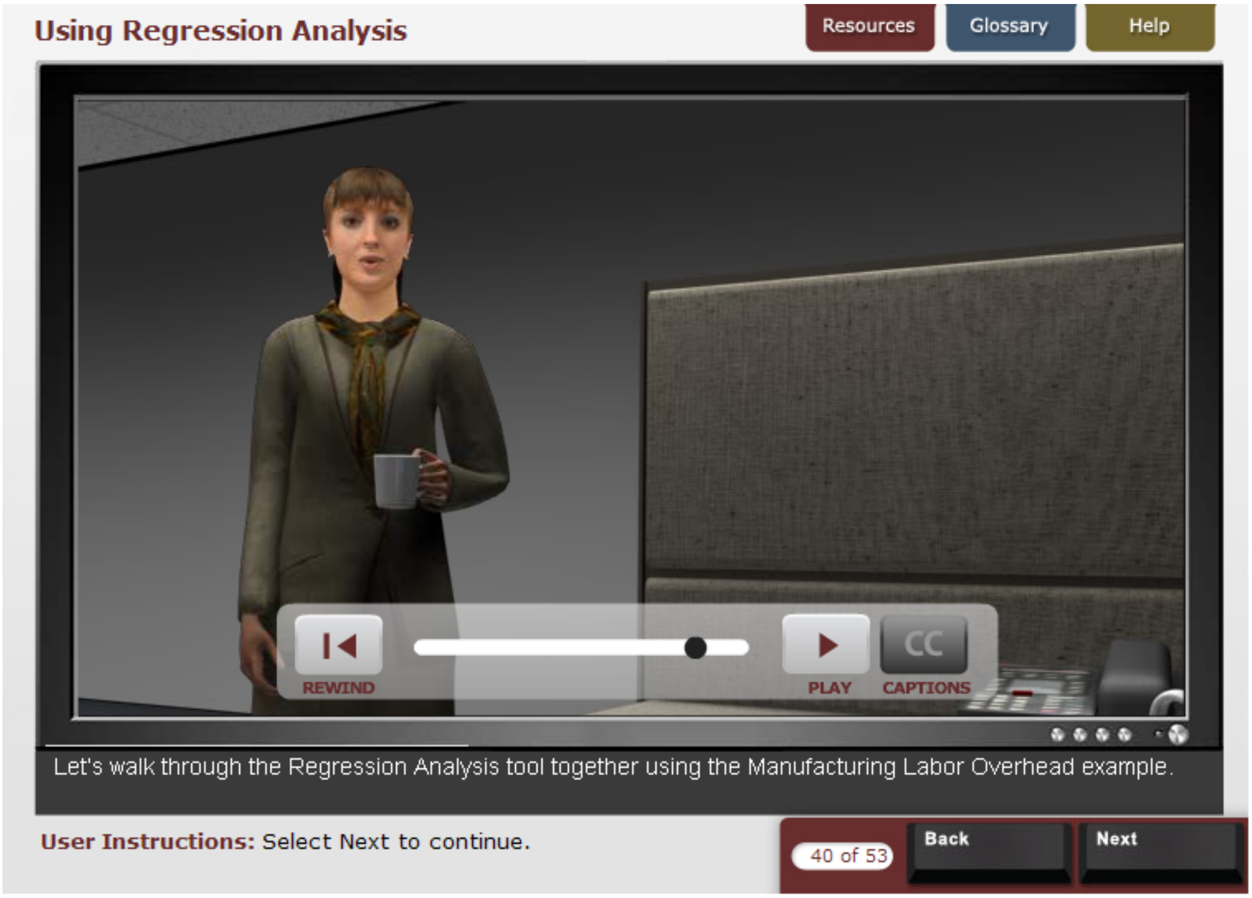

Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
The contractor has taken a very conservative approach when considering whether to include the exercise of options and probability of obtaining additional contract awards.
User Instructions: Select Next to continue.
40 of 53
Back
Next

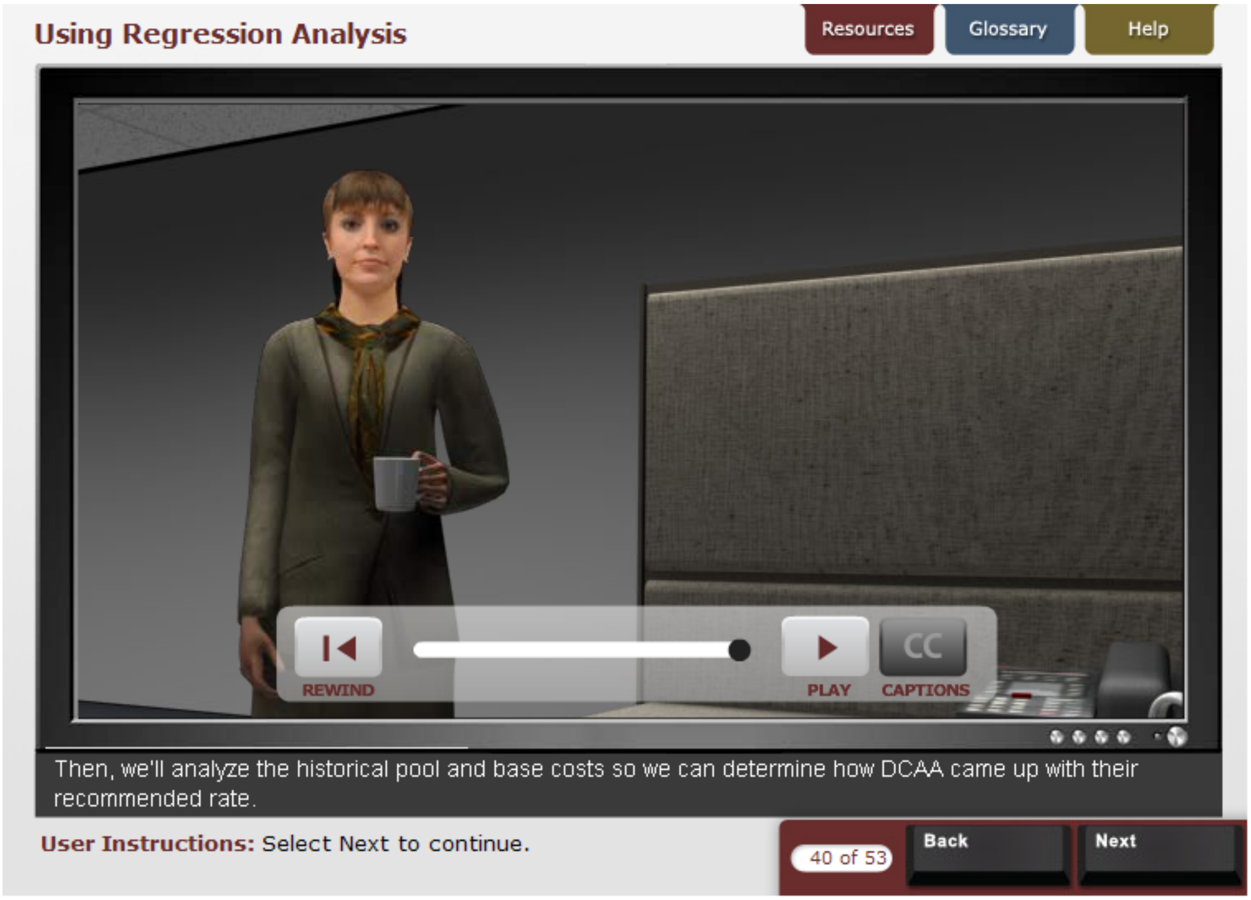

Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CC
CAPTIONS
In Fiscal Year 2008, when the BBOMS contract was awarded, it was awarded under a competitive source selection without the submission of certified cost or pricing data.
User Instructions: Select Next to continue.
40 of 53
Back
Next

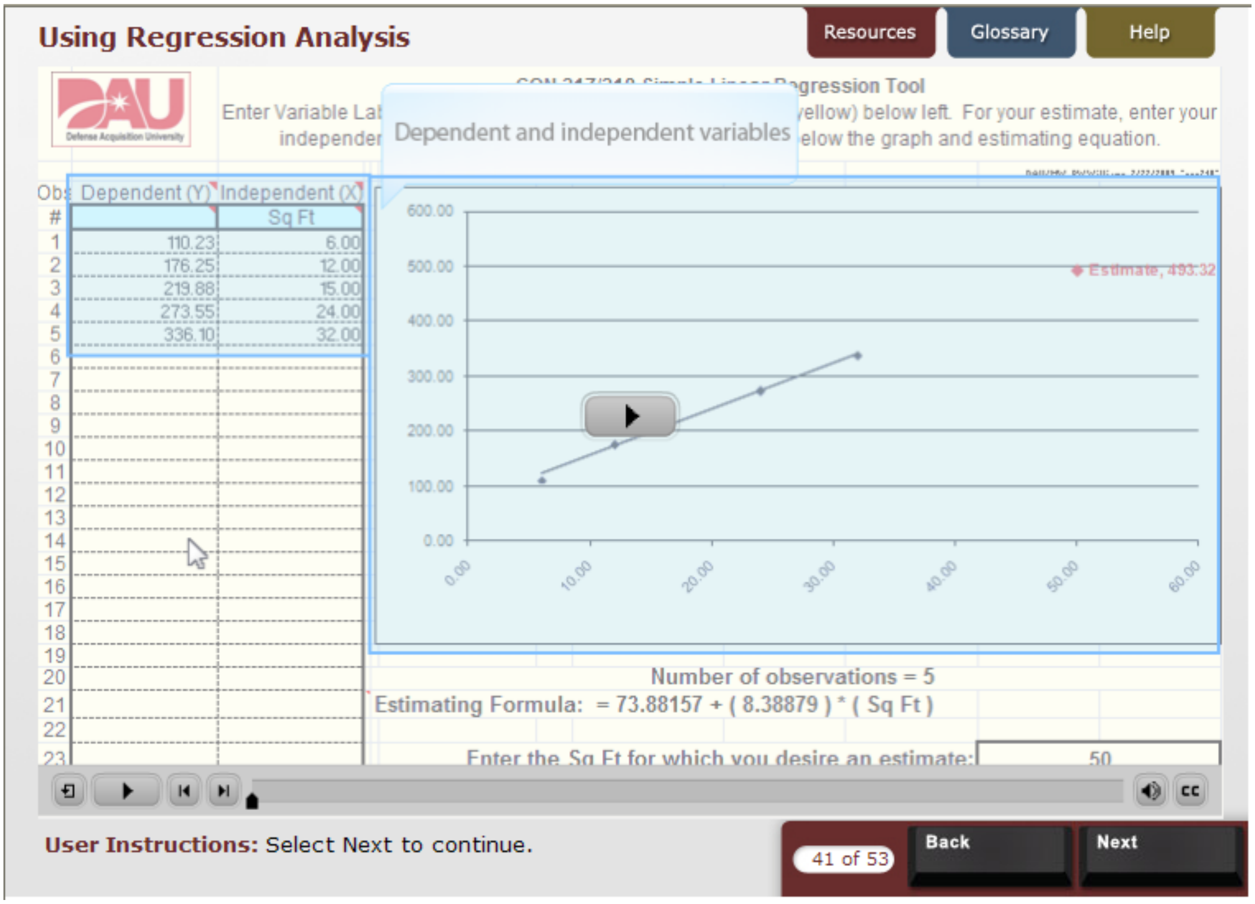

Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Now that the BBOMS contract is being modified on a sole source basis with the submission of certified cost or pricing data,
User Instructions: Select Next to continue.
40 of 53
Back
Next

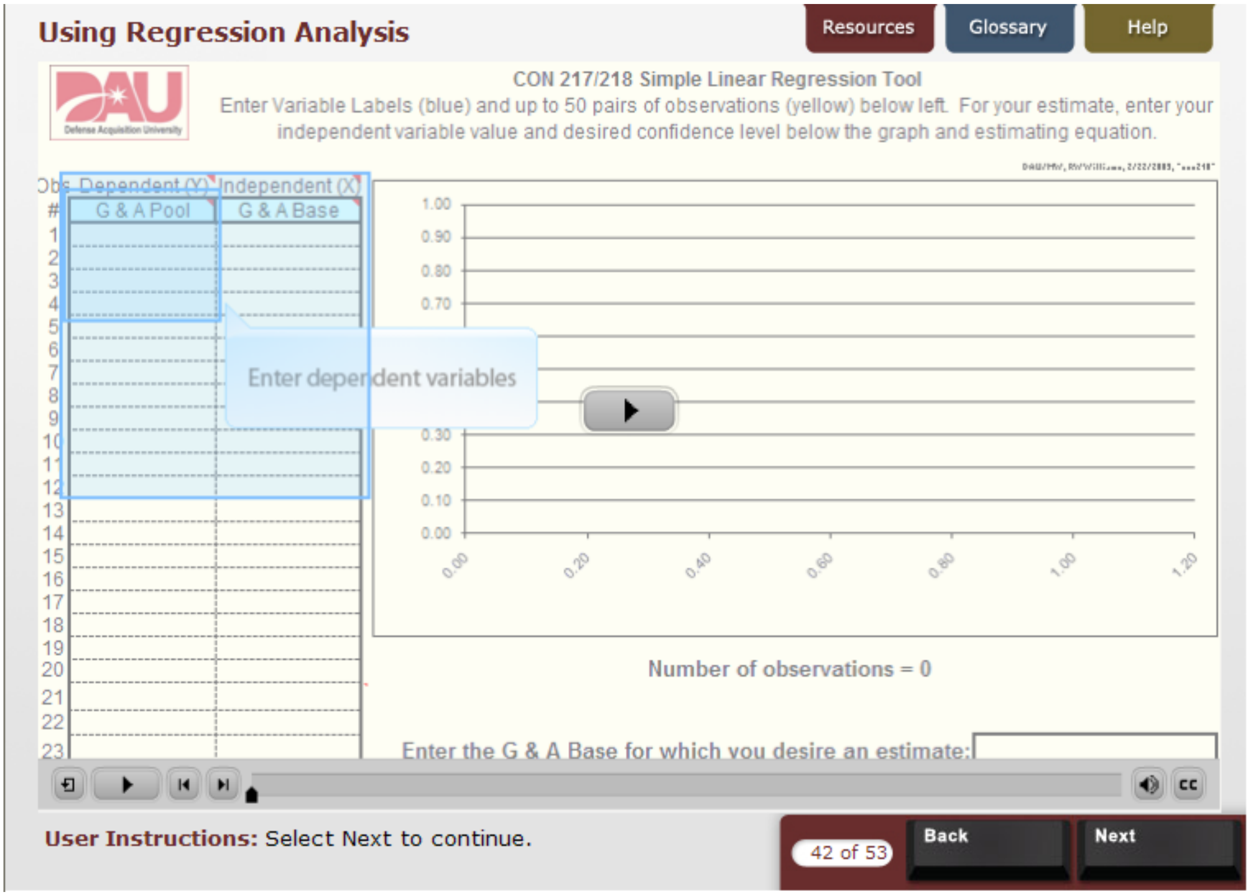

Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
TSS provided the estimates that they used when originally pricing the competitive acquisition in developing their G&A rate.
User Instructions: Select Next to continue.
40 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
The information provided was based on estimates of pool and base costs for fiscal years 2004, 2005 and 2006; at the time of submission, no incurred cost audit had been finalized, although one was on-going.
User Instructions: Select Next to continue.
40 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CAPTIONS
Let's walk through the Regression Analysis tool together using the Manufacturing Labor Overhead example.
User Instructions: Select Next to continue.
40 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
REWIND
PLAY
CC
CAPTIONS
Then, we'll analyze the historical pool and base costs so we can determine how DCAA came up with their recommended rate.
User Instructions: Select Next to continue.
40 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
Dependent and independent variables
Obs #
Dependent (Y)
Independent (X)
Sq Ft
110.23 6.00
176.25 12.00
219.88 15.00
273.55 24.00
336.10 32.00
Estimate, 493.32
Number of observations = 5
Estimating Formula: = 73.88157 + (8.38879) * (Sq Ft)
Enter the Sq Ft for which you desire an estimate: 50
User Instructions: Select Next to continue.
41 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
CON 217/218 Simple Linear Regression Tool
Enter Variable Labels (blue) and up to 50 pairs of observations (yellow) below left. For your estimate, enter your independent variable value and desired confidence level below the graph and estimating equation.
Dependent (Y)
Independent (X)
G & A Pool
G & A Base
Enter dependent variables
Number of observations = 0
Enter the G & A Base for which you desire an estimate:
User Instructions: Select Next to continue.
42 of 53
Back
Next

Resources Glossary Help

Using Regression Analysis

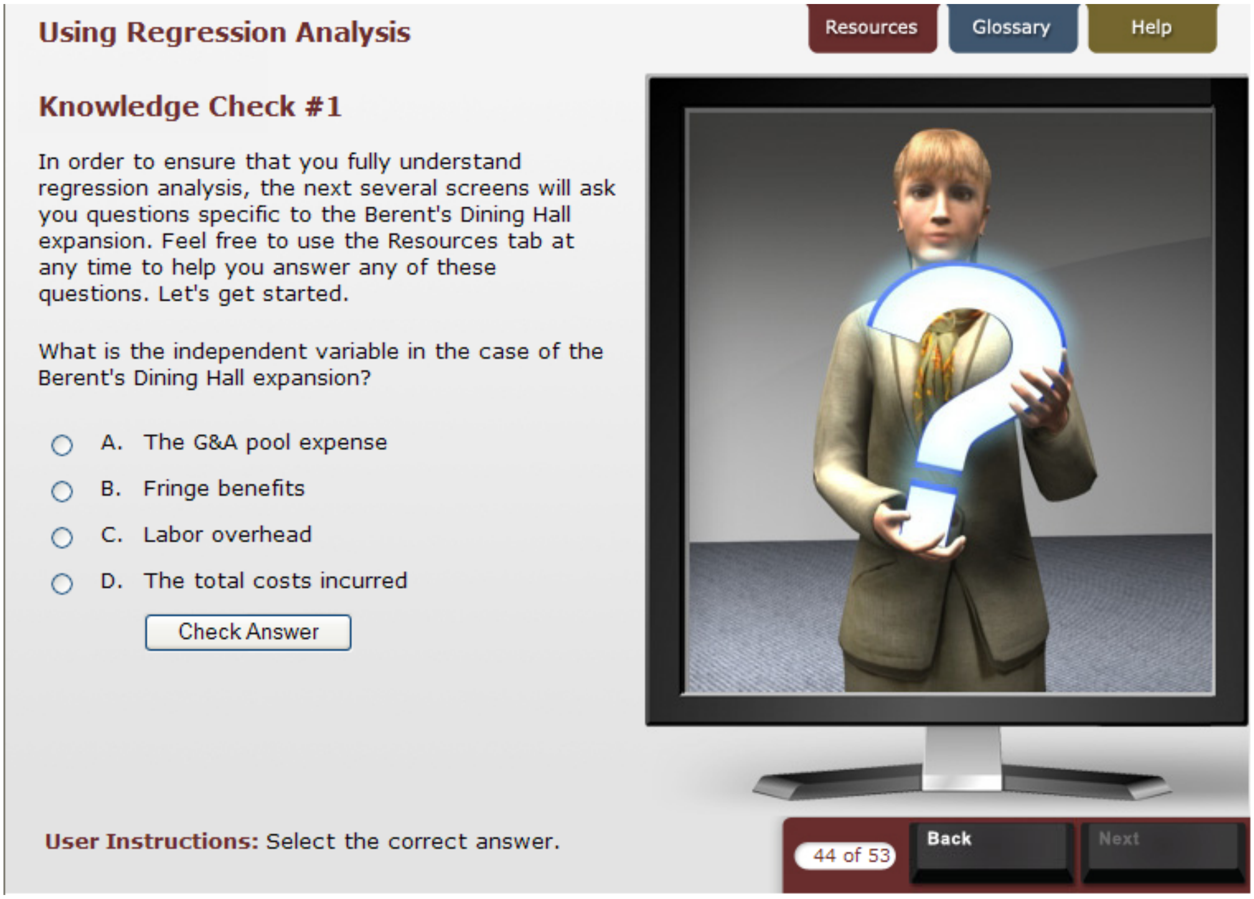
Calculating the Rate

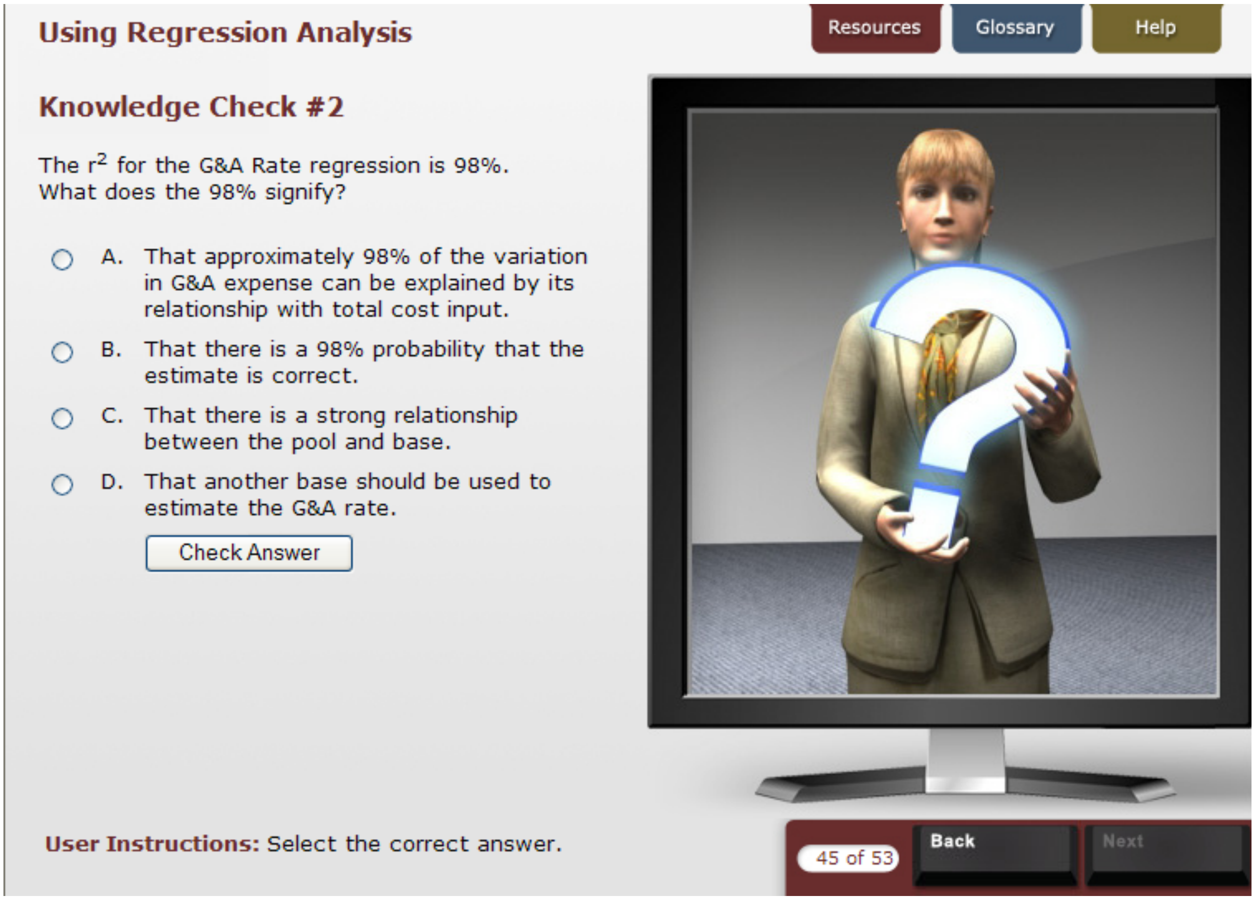
Now that you have used the regression tool to calculate the pool given a base of $64,500,000, you have to calculate the rate. In order to calculate the G&A Rate, you would take the estimated pool costs of $7,204,563.01 and divide that value by the estimated base of $64,500,000. The result gives you a G&A Rate of 11.17%.

- G&A Pool = $5,000,251.85 + .03418 (G&A Base)
- G&A Pool = $5,000,251.85 + .03418 ($64,500,000)
- G&A Pool = $7,204,563.01
- G&A Rate = Pool/Base = $7,204,563.01/$64,500,000 = 11.17%

User Instructions: Select Next to continue.

43 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
Knowledge Check #1
In order to ensure that you fully understand regression analysis, the next several screens will ask you questions specific to the Berent's Dining Hall expansion. Feel free to use the Resources tab at any time to help you answer any of these questions. Let's get started.
What is the independent variable in the case of the Berent's Dining Hall expansion?
A. The G&A pool expense
B. Fringe benefits
C. Labor overhead
D. The total costs incurred
Check Answer
User Instructions: Select the correct answer.
44 of 53
Back
Next


Using Regression Analysis
Resources
Glossary
Help
Knowledge Check #2
The r^2 for the G&A Rate regression is 98%.
What does the 98% signify?
A. That approximately 98% of the variation in G&A expense can be explained by its relationship with total cost input.
B. That there is a 98% probability that the estimate is correct.
C. That there is a strong relationship between the pool and base.
D. That another base should be used to estimate the G&A rate.
Check Answer
User Instructions: Select the correct answer.
45 of 53
Back
Next

Using Regression Analysis

Resources | Glossary | Help

Knowledge Check #3

Should you use this estimating equation to predict the G&A pool and calculate the G&A rate to be used in forward pricing?

- A. Yes. It is highly unlikely that a significant relationship does not exist between X and Y.
- B. Yes, but only if he has base amounts that are strictly within the relevant range.
- C. Maybe. The t-test will assist in making that call. Although rare, conflicts do exist even with a high r-squared due to the small sample size involved.
- D. No, the confidence level is not realistic.

Check Answer

User Instructions: Select the correct answer.

46 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

Knowledge Check #4

You have reviewed the statistical output of the regression tool and conducted the t-test. Should you use the estimating equation to predict the pool costs associated with the base of total cost input?

- A. Yes. T of 10.12084635 > t of 2.91998558 at the 90% confidence level.
- B. Yes. T of 24.56076988 > t of 1.885618083 at the 80% confidence level.
- C. No. T of 10.12084635 < t of 24.56076988 at the 99% confidence level.
- D. Yes. T of 9.9248432 > t of 10.12084635 at the 99% confidence level.

Check Answer

User Instructions: Select the correct answer.

47 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Knowledge Check #5

Using the estimating equation Pool = $5,000,251.85 + (.03418) * (G&A Base) with an estimated base of $64,500,000 yielded a G&A Rate of 11.17%.

What would the rate be if the estimated base was $70,000,000?

- A. $7,392,528
- B. 10.56%
- C. 9.47%
- D. 11.46%

Check Answer

User Instructions: Select the correct answer.

48 of 53 | Back | Next

Using Regression Analysis

Resources Glossary Help

Knowledge Check #6

According to CPRG, Volume 2, Chapter 5, Section 5.7, the G&A rate calculated using a base of $64,500,000 is only a point estimate, which means it is only the most likely outcome. There could be a range of estimates that are correct.

What would be a reasonable range of G&A rates to consider when developing the estimate?

- A. 11.07 - 11.27%
- B. 15.07 - 15.27%
- C. 11% - 13%
- D. $7,138,415 - $7,270,711

Check Answer

User Instructions: Select the correct answer.

49 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

Knowledge Check #7

You have decided to use 11.17% as your G&A rate when calculating the G&A expense prenegotiation objective. What is the G&A expense for the Berent's Dining Hall expansion?

- A. $116,582.38
- B. $199,431.28
- C. $163,827.53
- D. $42,742.66

Check Answer

User Instructions: Select the correct answer.

50 of 53 Back Next

Using Regression Analysis

Resources | Glossary | Help

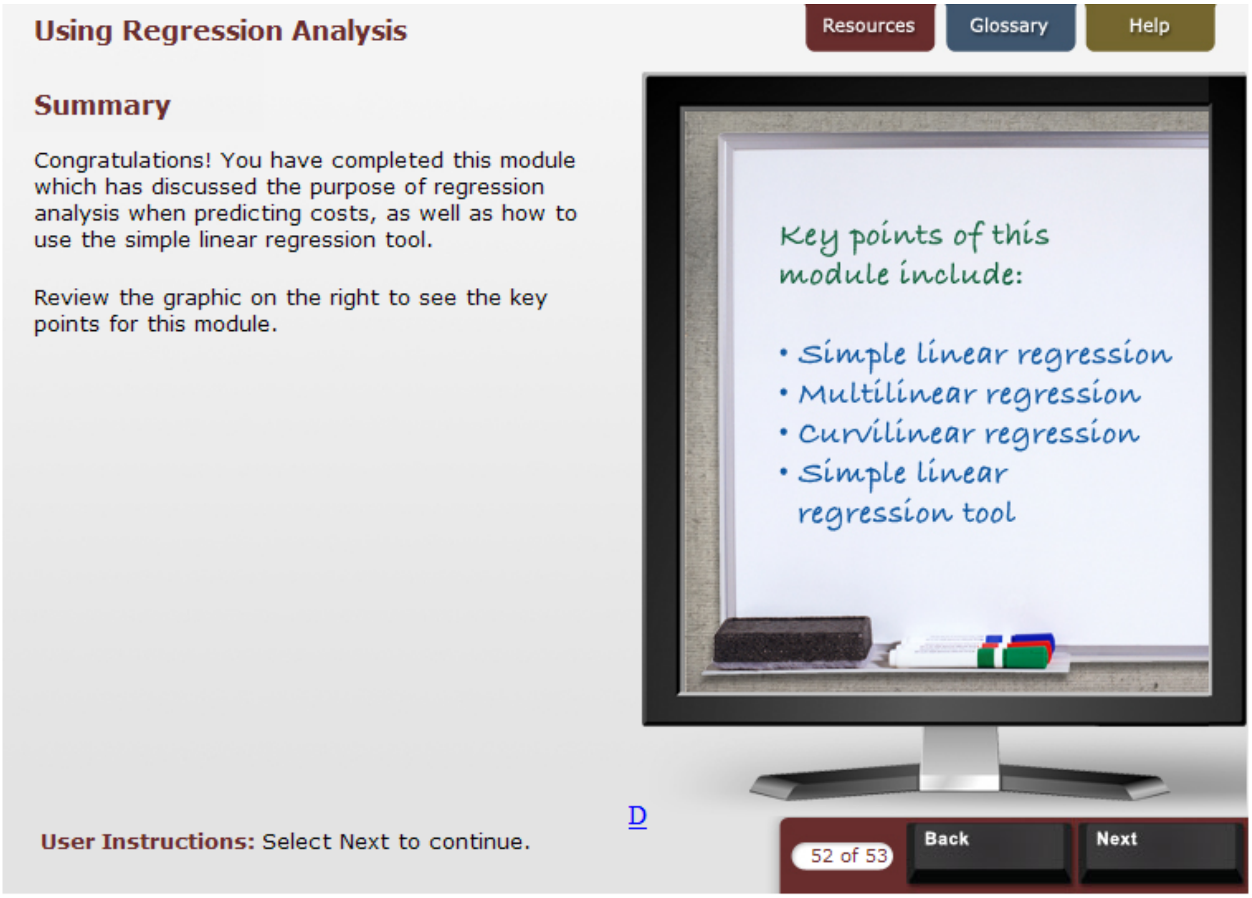
Prenegotiation Objective - G&A Rate

Using regression analysis, you calculated a G&A rate of 11.17%. This value is applied to your prenegotiation objective for TCI in your Berent's Dining Hall spreadsheet.

	B	C	E	F	G	I	J	K
1	**Contractor's Proposal**		**Government Objective**			**Negotiated Amount**		
2	Cost Elements	Proposed Amount	Pre-negotiation Amount	Cost Data Analysis	Rates	Pre-negotiation Amount	Cost Data Analysis	Rates
3	Material and Subcontracts							
4	Purchased Parts & Raw Material	$ 412,805.99	$ 382,655.89					
5								
6	Scrap @ 5% of PP/RM	$ 20,640.30	$ 17,219.52	Scrap Rate	4.50%	$ -	Scrap Rate	
7								
8	Subcontracts	$ 643,834.38	$ 643,834.38					
9								
10	Total Material Costs	$ 1,077,280.67	$ 1,043,709.78			$ -		
11								
12	Direct Labor	$ 144,000.00	$ 143,000.00					
13								
14	Labor Overhead @ 135%	$ 194,400.00	$ 193,050.00	Direct Labor Overhead Rate	135%	$ -	Direct Labor Overhead Rate	
15								
16	Other Direct Costs	$77,665.97	$86,932.36					
17								
18	Subtotal	$ 1,493,346.64	$ 1,466,692.14			$ -		
19								
20								
21	General & Administrative Expense @ 15% of TCI	$ 224,002.00	$ 163,827.53	G&A Expense Rate	11.17%	$ -	G&A Expense Rate	

User Instructions: Select Next to continue.

51 of 53 Back Next


Using Regression Analysis
Resources
Glossary
Help
Summary
Congratulations! You have completed this module which has discussed the purpose of regression analysis when predicting costs, as well as how to use the simple linear regression tool.
Review the graphic on the right to see the key points for this module.
Key points of this module include:
• Simple linear regression
• Multilinear regression
• Curvilinear regression
• Simple linear regression tool
D
User Instructions: Select Next to continue.
52 of 53
Back
Next

Using Regression Analysis

Resources | Glossary | Help

Summary (cont.)

Now that you have completed this module, you should be able to:

- Define regression analysis
- Identify the different regression analysis methods
- Identify contract pricing situations where simple regression analysis should be considered
- Identify the steps for using simple regression analysis

User Instructions: Select the next module from the Table of Contents to continue.

53 of 53 | Back | Next