assignment Requirements for Multiple Regression Program

Personal Software Process for Engineers

# Program 12E Requirements

Write a program to calculate the three-variable multiple-regression estimating parameters (β0, β1, β2, β3).

Make an estimate from user-supplied inputs, and determine the 70% and 90% prediction intervals for the estimate.

Use a linked list to store the data and the Simpson’s integration routine from 5A to calculate the *t* distribution.

Test the program using the data in Table D16 as the historical data (p. 763). For the user inputs, use 185 LOC of new code, 150 LOC of reused code, and 45 LOC of modified code.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Program # | New LOC | Reused LOC | Modified LOC | Hours |
|  | *w* | *x* | *y* | *z* |
| 1 | 345 | 65 | 23 | 31.4 |
| 2 | 168 | 18 | 18 | 14.6 |
| 3 | 94 | 0 | 0 | 6.4 |
| 4 | 187 | 185 | 98 | 28.3 |
| 5 | 621 | 87 | 10 | 42.1 |
| 6 | 255 | 0 | 0 | 15.3 |
| Sum | 1,670 | 355 | 149 | 138.1 |
| Average | 278.33 | 59.17 | 24.83 | 23.02 |
| Estimate | 185 | 150 | 45 | ?? |

Also test the program using the data in Table A32 as the historical data (p. 556). For the user inputs, use 650 LOC of new code, 3,000 LOC of reused code, and 155 LOC of modified code.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Program # | New LOC | Reused LOC | Modified LOC | Development Hours |
|  | *w* | *x* | *y* | *z* |
| 1 | 1,142 | 1,060 | 325 | 201 |
| 2 | 863 | 995 | 98 | 98 |
| 3 | 1,065 | 3,205 | 23 | 162 |
| 4 | 554 | 120 | 0 | 54 |
| 5 | 983 | 2,896 | 120 | 138 |
| 6 | 256 | 485 | 88 | 61 |
| Sum | 4,863 | 8,761 | 654 | 714 |
| Estimate | 650 | 3,000 | 155 | ??? |

Submit a test report with your results in the format in Table D17 (p. 764).

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Parameter | Expected Value | Actual Value |
| Table D16 |  |  |  |
|  | β0 | 0.56645 |  |
|  | β1 | 0.06533 |  |
|  | β2 | 0.008719 |  |
|  | β3 | 0.15105 |  |
|  | Projected hours | 20.76 |  |
|  | UPI (70%) | 26.89 |  |
|  | LPI (70%) | 14.63 |  |
|  | UPI (90%) | 33.67 |  |
|  | LPI (90%) | 7.84 |  |
| Appendix A |  |  |  |
|  | β0 | 6.7013 |  |
|  | β1 | 0.0784 |  |
|  | β2 | 0.0150 |  |
|  | β3 | 0.2461 |  |
|  | Projected hours | 140.9 |  |
|  | UPI (70%) | 179.7 |  |
|  | LPI (70%) | 102.1 |  |
|  | UPI (90%) | 222.74 |  |
|  | LPI (90%) | 59.06 |  |

# Multiple Regression

## Introduction to Multiple Regression

Suppose you have the following data on six projects:

* development hours required
* new, reused, and modified LOC

You wish to estimate the hours for a new project that you judge will have 650 LOC of new code, 3,000 LOC reused code, and 155 LOC of modified code.

How would you estimate the development hours and the prediction interval?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Program # | New LOC | Reused LOC | Modified LOC | Development Hours |
|  | *w* | *x* | *y* | *z* |
| 1 | 1,142 | 1,060 | 325 | 201 |
| 2 | 863 | 995 | 98 | 98 |
| 3 | 1,065 | 3,205 | 23 | 162 |
| 4 | 554 | 120 | 0 | 54 |
| 5 | 983 | 2,896 | 120 | 138 |
| 6 | 256 | 485 | 88 | 61 |
| Sum | 4,863 | 8,761 | 654 | 714 |
| Estimate | 650 | 3,000 | 155 | ??? |

## Calculating Multiple Regression Parameters

Multiple regression provides a way to estimate the effects of multiple variables when you do not have separate data for each. The 13 steps are as follows.

1. Use the following multiple regression formula to calculate the project value.



1. Find the beta parameters by solving the following simultaneous linear equations.







1. When you calculate the values of the terms, you get the following simultaneous linear equations.



1. Diagonalize, using the Gauss method. This successively eliminates one parameter at a time from the equations by successive multiplication and subtraction to give the following.



1. Solve for the beta terms.



1. Determine the prediction interval by solving for the range with the following equation.



1. Calculate the variance as follows.



1. The variance evaluates to the following.



1. The terms under the square root are as follows.



1. The value of the *t* distribution for a 70% prediction interval is found under the 85% column and two degrees of freedom in Table A2 (p. 489). It is 1.386.
2. The square root is evaluated.



1. The final estimate is as follows.

*z* = 6.71+0.0784\*650+0.0150\*3,000+0.2461\*155

= 140.902 hours

1. The prediction interval of 38.846 hours means that the estimate is between 102.1 and 179.7 hours.

# The Gauss Method

A frequently occurring problem in scientific engineering and business applications is to solve a system of *N* equations with *N* unknowns. Gaussian elimination is a commonly used algorithm for solving these equations.

The Gauss method converts this



to this



This technique is called *triangularization.*

What is the value of *X*?

## Gauss Method Example

Solve this system.

The pivot [Row 1, Col 1] 

1. The equation with the largest absolute coefficient in the pivot column [Any row, Col 1] is exchanged with the equation in the pivot row.



1. Reduce the coefficient for the pivot term to 1 by multiplying through by the reciprocal of the coefficient.



We now have



1. Eliminate the pivot column terms in all rows below the pivot row.
2. Multiply the pivot row by the negated coefficient of the row below the pivot term to be eliminated, which is −2 in this example.

  \* = 

Note: Don’t change the pivot row; put this result in a temp store.

1. Add the result of Step 3a to the row being eliminated.
2. Repeat Steps 3a and 3b for each row below the pivot term.

 +  = 

1. Complete the triangularization by repeating Steps 1 to 3 as needed; then solve for the remaining terms using back-substitution.







## The Gauss Algorithm

On a computer, the equations are represented by a matrix of the coefficients.

*X Y*

2 3 8

3 6 15

The algorithm is the same.

Given

|  |  |  |
| --- | --- | --- |
| *X* | *Y* |  |
| 2 | 3 | 8 |
| 3 | 6 | 15 |

|  |  |  |
| --- | --- | --- |
| *X* | *Y* |  |
| 3 | 6 | 15 |
| 2 | 3 | 8 |

|  |  |  |
| --- | --- | --- |
| *X* | *Y* |  |
| 1 | 2 | 5 |
| 2 | 3 | 8 |

|  |  |  |
| --- | --- | --- |
| *X* | *Y* |  |
| 1 | 2 | 5 |
| 0 | -1 | -2 |

|  |  |  |
| --- | --- | --- |
| *X* | *Y* |  |
| 1 | 2 | 5 |
| 0 | 1 | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *X* | *Y* |  |  |
| 1 | 2 | 5 | *X* = 1 |
| 0 | 1 | 2 | *Y* = 2 |

## A Gauss Algorithm

1. I = 1
2. LET P=AK , I = max{|AJ,I| : I <= J <= N} *(Find the pivot value, which is the largest number in the column below the I,I position.)*
3. IF P = 0 THEN EXIT (*If the remainder of this column is zero, a unique solution does not exist.)*
4. IF K > I THEN FOR J = I TO N + 1 SWAP A(I,J) AND A(K,J) *(Get the pivot value into the I,I position.)*
5. FOR J = I + 1 TO N + 1 LET AI,J = AI,J / AI,I *(Get a one in the I,I position.)* SET AI,I = 1.
6. FOR L = I + 1 to N multiply row I by -AL,I and add to row L *(Zero out the column below the I,I position.)*
7. I = I + 1. IF I < N THEN GO TO 2.
8. EXIT to back-substitution

Document Markings

Copyright 2020 Carnegie Mellon University. All rights reserved.  
  
This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.  
  
Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.  
  
NO WARRANTY. THIS MATERIAL IS FURNISHED ON AN “AS-IS” BASIS WITH NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, ANY WARRANTY WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT, OR THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.  
  
[Distribution Statement A] This material has been approved for public release and unlimited distribution. The United States Government has Unlimited Rights in this material as defined by DFARS 252.227-7013.

The text and illustrations in this material are licensed by Carnegie Mellon University under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

The Creative Commons license does not extend to logos, trade marks, or service marks of Carnegie Mellon University.