# Linear Regression Analysis

Write a program to calculate the linear regression size-estimating parameters.

## Requirements:

Write a program to calculate the linear regression size-estimating parameters for two arrays, each of n numbers. Enhance program 1 to work for the new calculations with minimal duplication.

Given a set of historical data for variables *x* and *y*, you want to determine if a likely value *yk* based on a known or estimated new value *xk*. An example would be the relationship between the estimated object LOC in a program and the actual new and changed program LOC.

Conditions:

The historical x and y data must demonstrate a relationship.

There must be sufficient data produce a statistically significant result (at least three items and preferably five or more.)

Determine the beta0 and beta 1 parameters that best represent the relationship between these *x* and *y* data, and then calculate *yk* using the following formula and the available data.

1.



2.



3.



## Testing:

Thoroughly test the program. At a minimum, use this program to calculate the beta parameters for the three provide data sets.

**TEST DATA**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Table D8:** Size Estimating Regression Data | | | | | **Program Number** | **Estimated Object LOC** | **Estimated New and Changed LOC** | **Actual New and Changed LOC** | | 1 | 130 | 163 | 186 | | 2 | 650 | 765 | 699 | | 3 | 99 | 141 | 132 | | 4 | 150 | 166 | 272 | | 5 | 128 | 137 | 291 | | 6 | 302 | 355 | 331 | | 7 | 95 | 136 | 199 | | 8 | 945 | 1206 | 1890 | | 9 | 368 | 433 | 788 | | 10 | 961 | 1130 | 1601 | | Sum | 3828 | 4632 | 6389 | | Average | 382.8 | 463.2 | 638.9 | |

1. Use the data in above for estimated object LOC and actual new and changed LOC. The resulting values should be *beta\_0* = -22.55 and *beta\_1* = 1.7279.
2. Calculate the *beta\_0* and *beta\_1* parameters for the regression fit of estimated new and changed LOC to actual new and changed LOC columns in Table D8. The answer in this case should be *beta\_0* = -23.92 and *beta\_1* = 1.4310.
3. Calculate the *beta\_0* and *beta\_1* parameters for the estimated new and changed LOC and the actual new and changed LOC for the programs 2A, 3A and 4A that you have developed.

**WORKED EXAMPLE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Program Number** | **Estimated Object LOC** | **Estimated New and Changed LOC** | ***XiYi*** | **Xi2** |
| 1 | 1 | 130 | 186 | 24180 |
| 2 | 2 | 650 | 699 | 454350 |
| 3 | 3 | 99 | 132 | 13068 |
| 4 | 4 | 150 | 272 | 40800 |
| 5 | 5 | 128 | 291 | 37248 |
| 6 | 6 | 302 | 331 | 99962 |
| 7 | 7 | 95 | 199 | 18905 |
| 8 | 8 | 945 | 1890 | 1786050 |
| 9 | 9 | 368 | 788 | 289984 |
| 10 | 10 | 961 | 1601 | 1538561 |
| Sum | 3828 | 6389 | 4303108 | 2540284 |
| Average | 382.8 | 638.9 |  |  |



=638.9-1.7279\*382.8=-22.55

Size total= = -22.55 + 1.7279 \* 389 = 644

**Table 4-1. Test Results Format-- Program 4A**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Expected B0 | Expected B1 | Actual B0 | Actual B1 |
| Table D8: Estimated Class vs Actual New and Changed LOC | -22.55 | 1.7279 |  |  |
| Table D8: Estimated New and Changed LOC vs Actual New and Changed LOC | -23.92 | 1.4310 |  |  |
| Program 2A: Estimated New and Changed LOC vs Actual New and Changed LOC |  |  |  |  |
| Program 3A: Estimated New and Changed LOC vs Actual New and Changed LOC |  |  |  |  |
| Program 4A: Estimated New and Changed LOC vs Actual New and Changed LOC |  |  |  |  |