assignment Requirements for *t* Distribution Program

Personal Software Process for Engineers

# Program Requirements

Write a program to calculate the correlation between two series of numbers and calculate the significance of that correlation.

Use the Simpson’s integration routine from Program 5A to calculate the values of the *t* distribution.

Hold the data in a linked list.

Test the program with the actual new and changed LOC in Table D12 as the *x* data and the development hours in Table D12 as the *y* data (p. 759).

|  |  |  |
| --- | --- | --- |
| Item Number | Actual New and Changed LOC | Development Hours |
| *n* | *x* | *y* |
| 1 | 186 | 15.0 |
| 2 | 699 | 69.9 |
| 3 | 132 | 6.5 |
| 4 | 272 | 22.4 |
| 5 | 291 | 28.4 |
| 6 | 331 | 65.9 |
| 7 | 199 | 19.4 |
| 8 | 1890 | 198.7 |
| 9 | 788 | 38.8 |
| 10 | 1601 | 138.2 |
| Totals | 6389 | 603.2 |

Also using Program 9E, determine the correlation and significance between actual N&C and actual development time and estimated N&C and development time for your assignments to date.

Prepare and submit a test report that includes these data and uses the format in Table D13 (p. 759).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Expected Value | | | Actual Value | | |
|  | *r* | *t* | 2\*(1 − *p*) | *r* | *t* | 2\*(1 − *p*) |
| Table D12 | 0.9543 | 9.0335 | 1.80\*10−5 |  |  |  |
| Actual LOC vs. Dev. Time | N/A | NA | NA |  |  |  |
| Estimated LOC vs. Dev. Time | NA | NA | NA |  |  |  |

# Correlation

Correlation is a measure of the degree to which two variables are linearly related. If two variables are highly related, one variable’s value may be used to predict the other.

## Correlation Example

The following example illustrates correlation using height and weight of ten school children.

* Given: The height and weight of ten school children
* Question: To what degree are these data linearly related?

Using linear regression, we examine the dependency of weight on height.

* Calculate β0 and β1 regression parameters.
* Use known heights for the *x* values.
* Use corresponding weights for the *y* values.
* Draw the regression line.

Now we examine the opposite dependency of height on weight.

* Calculate β0 and β1 regression parameters.
* Use known weights for the *x* values.
* Use corresponding heights for the *y* values.
* Draw the regression line.

The size of the gap between the scissors formed by the intersection of the two lines illustrates the quality of the correlation.

If the correlation is perfect, the lines will coincide.

The weaker the correlation, the wider the gap between the two lines.

## Using Correlation

The correlation *rxy* can range from +1 to −1.

* Near +1 implies a strong positive relationship: when *x* increases, so does *y*.
* Near −1 implies a strong negative relationship: when *x* increases, *y* decreases.
* Near 0 implies no relationship.

Correlation is used in the PSP to judge the quality of the linear relation of various historic process data that are used for planning.

For this purpose we use the value of the relation *rxy* squared or *r*2.

|  |  |
| --- | --- |
| If *r*2 is … | the relationship is … |
| .9 ≤ *r*2 | predictive, use it with high confidence |
| .7 ≤ *r*2 *< .9* | strong and can be used for planning |
| .5 ≤ *r*2 < .7 | adequate for planning but use with caution |
| *r*2 *< .5* | not reliable for planning purposes |

## Limitations of Correlation

Correlation doesn’t not imply cause and effect. A strong correlation may be coincidental. For example, from 1840 to 1960, no U.S. president elected in a year ending in 0 survived his presidency. Coincidence or …?

Many coincidental correlations may be found in historical process data. To use a correlation you must understand the cause-and-effect relationship in the process.

## Calculating Correlation

The formula for calculating the correlation coefficient *r* is



where

* *x* and *y* are the two paired sets of data
* *n* is the number of their members

# Significance

### The Significance Test

The significance test determines the likelihood that a strong correlation is random and is therefore of no practical significance.

For example, a data set with two points will always have an *r*2 = 1, but this correlation is not significant.

## Calculating Significance

The procedure for calculating the correlation significance is as follows:

1. Calculate *t* as



where *r* is the correlation.

1. Find the probability *p* by numerically integrating the *t* distribution for *n −* 2 degrees of freedom from *−*∞ to *t*.
2. Calculate the distribution tails, .

A tail area ≤ 0.05 is considered strong evidence that there is a relationship.

A tail area ≥ 0.2 is considered as good likelihood of finding the relationship by chance.

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.