Assignment Kit for Program 3

Personal Software Process (PSP) for Engineers: Part I

The Software Engineering Institute (SEI)
is a federally funded research and development center
sponsored by the U.S. Department of Defense and
operated by Carnegie Mellon University.

Personal Software Process for Engineers: Part I

Assignment Kit for Program 3

Overview

Overview

This assignment kit covers the following topics.

Section	See Page
Prerequisites	2
Program 3 requirements	3
Regression overview	5
Correlation overview	6
Calculating regression and correlation	8
An example	9
Assignment instructions	11
Guidelines and evaluation criteria	18

Prerequisites

Reading

- Chapters 5 and 6

Program 3 requirements

Program 3 requirements

Using PSP1, write a program to

- calculate the linear regression parameters β_0 and β_1 and correlation coefficients $r_{x,y}$ and r^2 for a set of *n* pairs of data,
- given an estimate, x_k calculate an improved prediction, y_k where $y_k = \beta_0 + \beta_1 x_k$
- enhance the linked list developed in program 1 to store the *n* data sets, where each record holds two real numbers

Table 1 contains historical estimated and actual data for 10 programs. For program 11, the developer has estimated a proxy size of 386 LOC.

Thoroughly test the program. At a minimum, run the following four test cases.

- Test 1: Calculate the regression parameters and correlation coefficients between estimated proxy size and actual added and modified size in Table 1. Calculate plan added and modified size given an estimated proxy size of x_k = 386.
- Test 2: Calculate the regression parameters and correlation coefficients between estimated proxy size and actual development time in Table 1. Calculate time estimate given an estimated proxy size of x_k = 386.
- Test 3: Calculate the regression parameters and correlation coefficients between plan added and modified size and actual added and modified size in Table 1. Calculate plan added and modified size given an estimated proxy size of x_k = 386.
- Test 4: Calculate the regression parameters and correlation coefficients between plan added and modified size and actual development time in Table 1. Calculate time estimate given an estimated proxy size of x_k = 386.

Expected results are provided in Table 2.

Program Number	Estimated Proxy Size	Plan Added and Modified size	Actual Added and Modified Size	Actual Development Hours
1	130	163	186	15.0
2	650	765	699	69.9
3	99	141	132	6.5
4	150	166	272	22.4
5	128	137	291	28.4
6	302	355	331	65.9
7	95	136	199	19.4
8	945	1206	1890	198.7
9	368	433	788	38.8
10	961	1130	1601	138.2

Table 1

Continued on next page

Expected results

Test	Expected Values					Actual Values				
	β_0	β_1	$r_{x,y}$	r^2	y_k	β_0	β_1	$r_{x,y}$	r^2	y_k
Test 1	-22.55	1.7279	0.9545	0.9111	644.429					
Test 2	-4.039	0.1681	0.9333	.8711	60.858					
Test 3	-23.92	1.43097	.9631	.9276	528.4294					
Test 4	-4.604	0.140164	.9480	.8988	49.4994					

Table 2

Regression

Overview

Linear regression is a way of optimally fitting a line to a set of data. The linear regression line is the line where the distance from all points to that line is minimized. The equation of a line can be written as

$y = \beta_0 + \beta_1 x$

In Figure 1, the best fit regression line has parameters of β_0 = -4.0389 and β_1 = 0.1681.

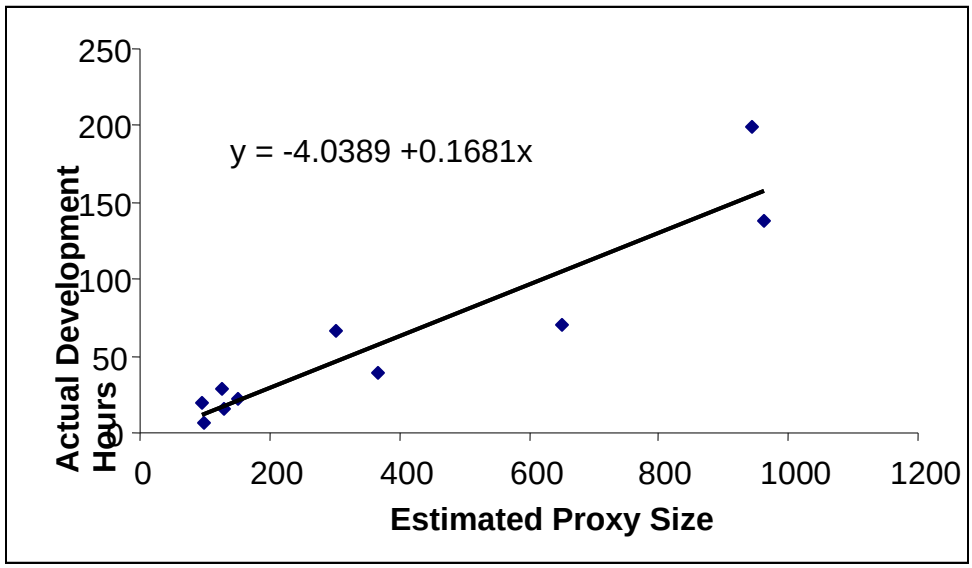

Figure 1

Continued on next page

Using regression in the PSP

Looking at Figure 1, how many hours do you think it would take to develop a program with an estimated proxy size of 500?

Using PROBE method A for time, the estimate would be $TimeEstimate = \beta_0 + \beta_1(500)$ or an estimate of 80.011 hours.

The PSP PROBE method uses regression parameters to make better predictions of size and time based on your historical data.

PROBE methods A and B differ only in the historical data (*x* values) used to calculate the regression parameters. In PROBE method A, **estimated proxy** size are used as the *x* values. In PROBE method B, **plan added and modified** size are used as the *x* values.

PROBE methods for size and time differ only in the historical data (*y* values) used to calculate the regression parameters. To predict improved size estimates, **actual added and modified LOC** are used as the *y* values. To predict time estimates, **actual development times** are used as the *y* values.

Historical Data Used		***x* values**	***y* values**
Size Estimating	PROBE A	Estimated Proxy Size	Actual Added and Modified Size
	PROBE B	Plan Added and Modified Size	Actual Added and Modified Size
Time Estimating	PROBE A	Estimated Proxy Size	Actual Development Time
	PROBE B	Plan Added and Modified Size	Actual Development Time

Correlation

Overview

The correlation calculation determines the relationship between two sets of numerical data.

The correlation $r_{x,y}$ can range from +1 to -1.

- Results near +1 imply a strong positive relationship; when *x* increases, so does *y*.
- Results near -1 imply a strong negative relationship; when *x* increases, *y* decreases.
- Results near 0 imply no relationship.

Using correlation in the PSP

Correlation is used in the PSP to judge the quality of the linear relation in various historical process data that are used for planning. For example, the relationships between estimated proxy size and actual time or plan added and modified size and actual time.

For this purpose, we examine the value of the relation r_{xy} squared, or r^2.

If r^2 is	the relationship is
$.9 \le r^2$	predictive; use it with high confidence
$.7 \le r^2 < .9$	strong and can be used for planning
$.5 \le r^2 < .7$	adequate for planning but use with caution
$r^2 < .5$	not reliable for planning purposes

Limitations of correlation

Correlation doesn't imply cause and effect.

A strong correlation may be coincidental.

> From 1840 to 1960, no U.S. president elected in a year ending in 0 survived his presidency.
> Coincidence or Correlation?

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 regression and correlation

Calculating regression and correlation

The formulas for calculating the regression parameters β_0 and β_1 are

$$\beta_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - \left(n x_{avg} y_{avg}\right)}{\left(\sum_{i=1}^{n} x_i^2\right) - \left(n x_{avg}^2\right)}$$

$$\beta_0 = y_{avg} - \beta_1 x_{avg}$$

The formulas for calculating the correlation coefficient $r_{x,y}$ and r^2 are

$$r_{x,y} = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$

$$r^2 = r * r$$

where

- Σ is the symbol for summation
- *i* is an index to the *n* numbers
- *x* and *y* are the two paired sets of data
- *n* is the number of items in each set *x* and *y*
- x_{avg} is the average of the *x* values
- y_{avg} is the average of the *y* values

An example

An example

In this example, we will calculate the regression parameters (β_0 and β_1 values) and correlation coefficients $r_{x,y}$ and r^2 of the data in the Table 3.

n	***x***	***y***
1	130	186
2	650	699
3	99	132
4	150	272
5	128	291
6	302	331
7	95	199
8	945	1890
9	368	788
10	961	1601

Table 3

$$\beta_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - \left(n x_{avg} y_{avg}\right)}{\left(\sum_{i=1}^{n} x_i^2\right) - \left(n x_{avg}^2\right)}$$

1. In this example there are 10 items in each dataset and therefore we set *n* = 10.
2. We can now solve the summation items in the formulas.

n	*x*	*y*	x^2	$x*y$	y^2
1	130	186	16900	24180	34596
2	650	699	422500	454350	488601
3	99	132	9801	13068	17424
4	150	272	22500	40800	73984
5	128	291	16384	37248	84681
6	302	331	91204	99962	109561
7	95	199	9025	18905	39601
8	945	1890	893025	1786050	3572100
9	368	788	135424	289984	620944
10	961	1601	923521	1538561	2563201
Total	$\sum_{i=1}^{10} x_i = 3828$	$\sum_{i=1}^{10} y_i = 6389$	$\sum_{i=1}^{10} x_i^2 = 2540284$	$\sum_{i=1}^{10} x_i y_i = 4303108$	$\sum_{i=1}^{10} y_i^2 = 7604693$
	$x_{avg} = \frac{3828}{10} = 382.8$	$y_{avg} = \frac{6389}{10} = 638.9$			

Continued on next page

An example, cont.

3. We can then substitute the values into the formulas

$$\beta_1 = \frac{(4303108) - (10 * 382.8 * 638.9)}{(2540284) - (10 * 382.8^2)}$$

$$\beta_1 = \frac{1857399}{1074926} = 1.727932$$

$$r_{x,y} = \frac{10(4303108) - (3828)(6389)}{\sqrt{[10(2540284) - (3828)^2][10(7604693) - (6389)^2]}}$$

$$r_{x,y} = \frac{18573988}{\sqrt{[10749256][35227609]}} \qquad r_{x,y} = \frac{18573988}{194594601}$$

$$r_{x,y} = 0.9545$$

$$r^2 = 0.9111$$

4. We can then substitute the values in the β_0 formula

$$\beta_0 = y_{avg} - \beta_1 x_{avg}$$

$$\beta_0 = 638.9 - 1.727932 * 382.8 = -22.5525$$

5. We now find y_k from the formula $y_k = \beta_0 + \beta_1 x_k$

$$y_k = -22.5525 + 1.727932 * 386 = 644.4294$$

Assignment instructions

Assignment instructions

Before starting program 3, review the top-level PSP1 process script below to ensure that you understand the "big picture" before you begin. Also, ensure that you have all of the required inputs before you begin the planning phase.

PSP1 Process Script

Purpose	To guide the development of module-level programs
Entry Criteria	- Problem description - PSP1 Project Plan Summary form - ***Size Estimating template*** - ***Historical size and time data (estimated and actual)*** - Time and Defect Recording logs - Defect Type, Coding, and Size Counting standards - Stopwatch (optional)

Step	**Activities**	**Description**
1	Planning	- Produce or obtain a requirements statement. - ***Use the PROBE method to*** estimate the added and modified size of this program. - ***Complete the Size Estimating template.*** - ***Use the PROBE method to*** estimate the required development time. - Enter the plan data in the Project Plan Summary form. - Complete the Time Recording log.
2	Development	- Design the program. - Implement the design. - Compile the program, and fix and log all defects found. - Test the program, and fix and log all defects found. - Complete the Time Recording log.
3	Postmortem	Complete the Project Plan Summary form with actual time, defect, and size data.

Exit Criteria	- A thoroughly tested program - Completed Project Plan Summary form with estimated and actual data - ***Completed Size Estimating template*** - ***Completed Test Report template*** - Completed PIP forms - Completed Time and Defect Recording logs

Continued on next page

Planning phase Plan program 3 following the PSP1 planning phase and the PROBE estimating scripts.

PSP1 Planning Script

Purpose	To guide the PSP planning process
Entry Criteria	- Problem description - PSP1 Project Plan Summary form - ***Size Estimating template*** - ***Historical size and time data (estimated and actual)*** - Time Recording log

Step	Activities	Description
1	Program Requirements	- Produce or obtain a requirements statement for the program. - Ensure that the requirements statement is clear and unambiguous. - Resolve any questions.
2	Size Estimate	- ***Produce a program conceptual design.*** - ***Use the PROBE method to*** estimate the added and modified size of this program. - ***Complete the Size Estimating template and Project Plan Summary form.***
3	Resource Estimate	- ***Use the PROBE method to*** estimate the time required to develop this program. - Using the To Date % from the most recently developed program as a guide, distribute the development time over the planned project phases. (Note: This step is completed by the SEI student workbook.)

Exit Criteria	- Documented requirements statement - ***Program conceptual design*** - ***Completed Size Estimating template*** - Completed Project Plan Summary form with estimated program size and development time data - Completed Time Recording log

Verify that you have met all of the exit criteria for the planning phase, **then have an instructor review your plan**. After your plan has been reviewed, proceed to the development phase.

Continued on next page

Assignment instructions, Continued

Use the PROBE method to create size and resource estimates.

Purpose	To guide the size and time estimating process using the PROBE method
Entry Criteria	- Requirements statement - Size Estimating template and instructions - Size per item data for part types - Time Recording log - Historical size and time data
General	- This script assumes that you are using added and modified size data as the size-accounting types for making size and time estimates. - If you choose some other size-accounting types, replace every "added and modified" in this script with the size-accounting types of your choice.

Step	Activities	Description
1	Conceptual Design	Review the requirements and produce a conceptual design.
2	Parts Additions	Follow the Size Estimating Template instructions to estimate the parts additions and the new reusable parts sizes.
3	Base Parts and Reused Parts	- For the base program, estimate the size of the base, deleted, modified, and added code. - Measure and/or estimate the size of the parts to be reused.
4	Size Estimating Procedure	- If you have sufficient estimated proxy size and actual added and modified size data (three or more points that correlate), use procedure 4A. - If you do not have sufficient estimated data but have sufficient plan added and modified and actual added and modified size data (three or more points that correlate), use procedure 4B. - If you have insufficient data or they do not correlate, use procedure 4C. - If you have no historical data, use procedure 4D.
4A	Size Estimating Procedure 4A	- Using the linear-regression method, calculate the β_0 and β_1 parameters from the estimated proxy size and actual added and modified size data. - If the absolute value of β_0 is not near 0 (less than about 25% of the expected size of the new program), or β_1 is not near 1.0 (between about 0.5 and 2.0), use procedure 4B.
4B	Size Estimating Procedure 4B	- Using the linear-regression method, calculate the β_0 and β_1 parameters from the plan added and modified size and actual added and modified size data. - If the absolute value of β_0 is not near 0 (less than about 25% of the expected size of the new program), or β_1 is not near 1.0 (between about 0.5 and 2.0), use procedure 4C.
4C	Size Estimating Procedure 4C	If you have any data on plan added and modified size and actual added and modified size, set β_0 = 0 and β_1 = (actual total added and modified size to date/plan total added and modified size to date).
4D	Size Estimating Procedure 4D	If you have no historical data, use your judgment to estimate added and modified size.

(continued)

Continued on next page

PROBE Estimating Script (Continued)

Step	Activities	Description
5	Time Estimating Procedure	- If you have sufficient estimated proxy size and actual development time data (three or more points that correlate), use procedure 5A. - If you do not have sufficient estimated size data but have sufficient plan added and modified size and actual development time data (three or more points that correlate), use procedure 5B. - If you have insufficient data or they do not correlate, use procedure 5C. - If you have no historical data, use procedure 5D.
5A	Time Estimating Procedure 5A	- Using the linear-regression method, calculate the β_0 and β_1 parameters from the estimated proxy size and actual total development time data. - If β_0 is not near 0 (substantially smaller than the expected development time for the new program), or β_1 is not within 50% of 1/(historical productivity), use procedure 5B.
5B	Time Estimating Procedure 5B	- Using the linear-regression method, calculate the β_0 and β_1 regression parameters from the plan added and modified size and actual total development time data. - If β_0 is not near 0 (substantially smaller than the expected development time for the new program), or β_1 is not within 50% of 1/(historical productivity), use procedure 5C.
5C	Time Estimating Procedure 5C	- If you have data on estimated – added and modified size and actual development time, set β_0 = 0 and β_1 = (actual total development time to date/estimated – total added and modified size to date). - If you have data on plan – added and modified size and actual development time, set β_0 = 0 and β_1 = (actual total development time to date/plan total added and modified size to date). - If you only have actual time and size data, set β_0 = 0 and β_1 = (actual total development time to date/actual total added and modified size to date).
5D	Time Estimating Procedure 5D	If you have no historical data, use your judgment to estimate the development time from the estimated added and modified size.
6	Time and Size Prediction Intervals	- If you used regression method A or B, calculate the 70% prediction intervals for the time and size estimates. - If you did not use the regression method or do not know how to calculate the prediction interval, calculate the minimum and maximum development time estimate limits from your historical maximum and minimum productivity for the programs written to date.

Exit Criteria	- Completed estimated and actual entries for all pertinent size categories - Completed PROBE Calculation Worksheet with size and time entries - Plan and actual values entered on the Project Plan Summary

Continued on next page

Development phase

Develop the program following the PSP1 development phase script.

PSP1 Development Script

Purpose	To guide the development of small programs
Entry Criteria	- Requirements statement - Project Plan Summary form with estimated program size and development time - Time and Defect Recording logs - Defect Type standard and Coding standard

Step	Activities	Description
1	Design	- Review the requirements and produce a design to meet them. - Record in the Defect Recording log any requirements defects found. - Record time in the Time Recording log.
2	Code	- Implement the design following the Coding standard. - Record in the Defect Recording log any requirements or design defects found. - Record time in the Time Recording log.
3	Compile	- Compile the program until there are no compile errors. - Fix all defects found. - Record defects in the Defect Recording log. - Record time in the Time Recording log.
4	Test	- Test until all tests run without error. - Fix all defects found. - Record defects in the Defect Recording log. - Record time in the Time Recording log. - ***Complete a Test Report template on the tests conducted and the results obtained.***

Exit Criteria	- A thoroughly tested program that conforms to the Coding standard - ***Completed Test Report template*** - Completed Time and Defect Recording logs

Verify that you have met all of the exit criteria for the development phase, then proceed to the postmortem phase.

Continued on next page

Postmortem phase

Conduct the postmortem following the PSP1 postmortem script.

PSP1 Postmortem Script

Purpose	To guide the PSP postmortem process
Entry Criteria	- Problem description and requirements statement - Project Plan Summary form with program size and development time data - ***Completed Test Report template*** - Completed Time and Defect Recording logs - A tested and running program that conforms to the coding and size counting standards

Step	Activities	Description
1	Defect Recording	- Review the Project Plan Summary to verify that all of the defects found in each phase were recorded. - Using your best recollection, record any omitted defects.
2	Defect Data Consistency	- Check that the data on every defect in the Defect Recording log are accurate and complete. - Verify that the numbers of defects injected and removed per phase are reasonable and correct. - Using your best recollection, correct any missing or incorrect defect data.
3	Size	- Count the size of the completed program. - ***Determine the size of the base, deleted, modified, base additions, reused, new reusable code, and added parts.*** - ***Enter these data in the Size Estimating template.*** - ***Determine the total program size*** - ***Enter this data in the Project Plan Summary form.***
4	Time	- Review the completed Time Recording log for errors or omissions. - Using your best recollection, correct any missing or incomplete time data.

Exit Criteria	- A thoroughly tested program that conforms to the coding and size counting standards - ***Completed Test Report template*** - Completed Project Plan Summary form - Completed PIP forms describing process problems, improvement suggestions, and lessons learned - Completed Time and Defect Recording logs

Verify that you have met all of the exit criteria for the PSP1 postmortem phase, then submit your assignment.

Continued on next page

Submitting your assignment

When you've completed the postmortem phase, submit your assignment package, source code, and test results to the instructor.

The order for the assignment package is
- PSP1 Project Plan Summary form
- Test Report template
- PIP form
- Size Estimating template
- PROBE Calculation worksheet
- Time Recording log
- Defect Recording log
- source program listing
- test results

Guidelines and evaluation criteria for program 3

Evaluation criteria

Your process report must be
- complete
- legible
- in the specified order

Your process data must be
- accurate
- precise
- self-consistent

Suggestions

Remember, you should complete this assignment today.

Keep your programs simple. You will learn as much from developing small programs as from large ones.

If you are not sure about something, ask your instructor for clarification.

Software is not a solo business, so you do not have to work alone.
- You must, however, produce your own estimates, designs, code, and completed forms and reports.
- You may have others review your work, and you may change it as a result.
- You should note any help you receive from others in your process report. Log the review time that you and your associates spend, and log the defects found or any changes made.

