Programa 8

Proceso Personal para el Desarrollo de Software

Este material fue realizado en base a material del curso “Personal Software Process for Engineers: PartI”, dictado por The Software Engineering Institute (SEI)

Proceso Personal para el Desarrollo de Software

Instructivo para el Programa 8

Descripción

Descripción

El presente instructivo cubre los siguientes temas.

Programa 8 Requerimientos

Requerimientos Programa 8

Usando PSP0.1, escriba un programa para calcular los parámetros de estimación (β_0, β_1, β_2, β_3) para la regresión múltiple con tres variables.

Realizar una estimación de horas para un nuevo desarrollo a partir de las entradas recibidas y determinar el intervalo de predicción del 70% para la estimación realizada.

Probar a fondo el programa.
Pruebe el programa utilizando los datos en la Tabla 1 como datos históricos. Como entrada para estimar el tiempo del nuevo programa a realizar considere que se espera que tenga, 185 LOC de código agregado (added), 150 LOC de código reutilizado (reused), y 45 LOC de código modificado (modified).

Los valores esperados se muestran en la tabla 2.

Prog. #	**Added 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

Tabla 1

Test	**Parameter**	**Expected Value**	**Actual Value**
Table 1	β_0	0.56645	
	β_1	0.06533	
	β_2	0.008719	
	β_3	0.15105	
	Projected Hours	20.76	
	UPI (70%)	26.89	
	LPI (70%)	14.63	

Tabla 2

Programa 8 Requerimientos, Continuación

Requerimientos Programa 8

Además, pruebe el programa utilizando los datos en la Tabla 3, como datos históricos. Como entrada para estimar el tiempo del nuevo programa a realizar considere que se espera que tenga, 650 LOC de código agregado (added), 3000 LOC de código reutilizado (reused), y 155 LOC de código modificado (modified).

Los valores esperados se muestran en la Tabla 4.

Prog#	Added 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

Tabla 3

Test	Parameter	Expected Value	Actual Value
Table 3	β_0	6.7013	
	β_1	0.0784	
	β_2	0.0150	
	β_3	0.2461	
	Projected Hours	140.9	
	UPI (70%)	179.7	
	LPI (70%)	102.1	

Tabla 4

Regresión Múltiple

Descripción

Supón que tienes los siguientes datos de seis proyectos:

- horas de desarrollo requeridas
- LOC agregadas (added), reutilizadas (reused), y modificadas (modified)

Supón que ahora quieres estimar las horas para un nuevo proyecto que piensas que va a tener 650 LOC de código added, 3,000 LOC de código reused, y 155 LOC de código modified.

Cómo estimarías las horas de desarrollo y el intervalo de predicción?

La respuesta es Regresión Múltiple. La Regresión Múltiple es necesaria cuando se tienen datos combinados de varias categorías diferentes de trabajo y no se pueden separar.

Tu tienes tiempos totales de desarrollo para cada uno de los seis proyectos pero no tiene forma de determinar cuantas horas se dedicaron a desarrollo nuevo, reutilización o modificación. Este es un problema habitual porque en general es impráctico medir en forma independiente el tiempo dedicado en cada categoría de trabajo de desarrollo.

Utilizando Regresión Múltiple

1. Utilice la siguiente formula de regresión múltiple para calcular el valor del proyecto.

$$z_k = \beta_0 + w_k\beta_1 + x_k\beta_2 + y_k\beta_3$$

2. Encuentre los parámetros Beta resolviendo las siguientes ecuaciones lineales simultáneas.

$$\beta_0 n + \beta_1\sum_{i=1}^{n} w_i + \beta_2\sum_{i=1}^{n} x_i + \beta_3\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i$$

$$\beta_0\sum_{i=1}^{n} w_i + \beta_1\sum_{i=1}^{n} w_i^2 + \beta_2\sum_{i=1}^{n} w_i x_i + \beta_3\sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i z_i$$

$$\beta_0\sum_{i=1}^{n} x_i + \beta_1\sum_{i=1}^{n} w_i x_i + \beta_2\sum_{i=1}^{n} x_i^2 + \beta_3\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i z_i$$

$$\beta_0\sum_{i=1}^{n} y_i + \beta_1\sum_{i=1}^{n} w_i y_i + \beta_2\sum_{i=1}^{n} x_i y_i + \beta_3\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} y_i z_i$$

3. Resuelve estas ecuaciones con los métodos algebraicos estándar que se utilizan para ecuaciones lineales simultáneas. El procedimiento de eliminación Gaussiana provee una forma estructurada de hacer esto. El intervalo de predicción en la regresión múltiple se determina con las siguientes ecuaciones, utilizando *dof* = *n*-4 grados de libertad.

$$\sigma^2 = \left(\frac{1}{n-4}\right)\sum_{i=1}^{n}\left(z_i - \beta_0 - \beta_1 w_i - \beta_2 x_i - \beta_3 y_i\right)^2$$

$$Range = t(0.35, dof)\sigma\sqrt{1 + \frac{1}{n} + \frac{(w_k - w_{avg})^2}{\sum_{i=1}^{n}(w_i - w_{avg})^2} + \frac{(x_k - x_{avg})^2}{\sum_{i=1}^{n}(x_i - x_{avg})^2} + \frac{(y_k - y_{avg})^2}{\sum_{i=1}^{n}(y_i - y_{avg})^2}}$$

Ejemplo de regresión múltiple

Ejemplo de regresión múltiple

En este ejemplo vamos a calcular los parámetros de regresión y el rango de predicción utilizando los datos históricos que se muestran en la tabla a continuación. Luego usaremos los parámetros de regresión y el rango de predicción para estimar las horas para un proyecto nuevo que se espera tenga 650 LOC de código added, 3,000 LOC de código reused, and 155 LOC de código modified.

Prog#	Added 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

1. Encuentra los parámetros Beta resolviendo las siguientes ecuaciones lineales simultáneas.

$$\beta_0 n + \beta_1 \sum_{i=1}^{n} w_i + \beta_2 \sum_{i=1}^{n} x_i + \beta_3 \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i$$

$$\beta_0 \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i^2 + \beta_2 \sum_{i=1}^{n} w_i x_i + \beta_3 \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i z_i$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} w_i x_i + \beta_2 \sum_{i=1}^{n} x_i^2 + \beta_3 \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i z_i$$

$$\beta_0 \sum_{i=1}^{n} y_i + \beta_1 \sum_{i=1}^{n} w_i y_i + \beta_2 \sum_{i=1}^{n} x_i y_i + \beta_3 \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} y_i z_i$$

2. Al calcular los valores de los términos se obtienen las siguientes ecuaciones lineales simultáneas.

$$6\beta_0 + 4,863\beta_1 + 8,761\beta_2 + 654\beta_3 = 714$$

$$4,863\beta_0 + 4,521,899\beta_1 + 8,519,938\beta_2 + 620,707\beta_3 = 667,832$$

$$8,761\beta_0 + 8,519,938\beta_1 + 21,022,091\beta_2 + 905,925\beta_3 = 1,265,493$$

$$654\beta_0 + 620,707\beta_1 + 905,925\beta_2 + 137,902\beta_3 = 100,583$$

3. Diagonalizamos utilizando el método de Gauss. De esta forma sucesivamente eliminamos un parámetro por vez de las ecuaciones por las sucesivas multiplicaciones y restas para obtener lo siguiente:

$$6\beta_0 + 4,863\beta_1 + 8,761\beta_2 + 654\beta_3 = 714$$

$$0\beta_0 + 580,437.5\beta_1 + 1,419,148\beta_2 + 90,640\beta_3 = 89,135$$

$$0\beta_0 + 0\beta_1 + 4,759,809\beta_2 - 270,635\beta_3 = 5,002.332$$

$$0\beta_0 + 0\beta_1 + 0\beta_2 + 37,073.93\beta_3 = 9,122.275$$

Ejemplo de regresión múltiple, Continuación

Ejemplo de regresión múltiple, continuación

4. La solución para los términos de Beta es.

$$\beta_0 = 6.7013$$
$$\beta_1 = 0.0784$$
$$\beta_2 = 0.0150$$
$$\beta_3 = 0.2461$$

5. Determinamos el intervalo de predicción resolviendo el rango con la siguiente ecuación.

$$Range = t(0.35, dof)\sigma\sqrt{1+\frac{1}{n}+\frac{(w_k - w_{avg})^2}{\sum_{i=1}^{n}(w_i - w_{avg})^2}+\frac{(x_k - x_{avg})^2}{\sum_{i=1}^{n}(x_i - x_{avg})^2}+\frac{(y_k - y_{avg})^2}{\sum_{i=1}^{n}(y_i - y_{avg})^2}}$$

6. Calculamos la varianza como sigue.

$$\sigma^2 = \left(\frac{1}{n-4}\right)\sum_{i=1}^{n}(z_i - \beta_0 - \beta_1 w_i - \beta_2 x_i - \beta_3 y_i)^2$$

7. La desviación estándar es la siguiente.

$$\sigma = \sqrt{\sigma^2} = \sqrt{513.058} = 22.651$$

8. Los términos bajo la raíz cuadrada son los siguientes.

$$(New_k - New_{avg})^2 = (w_k - w_{avg})^2 = (650 - 810.5)^2 = 25{,}760.25$$
$$(\mathrm{Re}use_k - \mathrm{Re}use_{avg})^2 = (x_k - x_{avg})^2 = (3{,}000 - 1{,}460.17)^2 = 2{,}371{,}086.694$$
$$(Modify_k - Modify_{avg})^2 = (y_k - y_{avg})^2 = (155 - 109)^2 = 2{,}116$$

9. Usamos el programa 6 para encontrar el valor de *x*, donde *p* = 0.35 con *n*- 4 grados de libertad para la distribución t. En este caso, *t*(0.35,*dof*) = 1.386.

10. Evaluamos la raíz cuadrada.

$$Range = 1.386 * 22.651\sqrt{1+(1/6)+\frac{25{,}760.25}{580{,}437.5}+\frac{2{,}371{,}086.694}{8{,}229{,}571}+\frac{2{,}116}{66{,}616}} = 38.$$

12. La estimación final es la siguiente.

z = 6.71+0.0784*650+0.0150*3,000+0.2461*155
= 140.902 horas

13. El intervalo de predicción de 38.846 horas significa que la estimación está entre 102.1 y 179.7 horas.

El método de Gauss

El método de Gauss

Un problema que ocurre frecuentemente en la ingeniería científica y en aplicaciones de negocio es resolver un sistema de *N* ecuaciones con *N* incógnitas.

La eliminación Gaussiana es el algoritmo comúnmente utilizado para resolver estas ecuaciones.

El método de Gauss convierte esto

$$\boxed{\begin{array}{l} 2X + 3Y = 8 \\ 3X + 6Y = 15 \end{array}}$$

en esto

$$\boxed{\begin{array}{l} X + 2Y = 5 \\ Y = 2 \end{array}}$$

Esta técnica se llama *triangularización.*

Cuál es el valor de X?

Método de Gauss, ejemplo

Resolvamos el sistema.

El pivot [Fila 1, Columna 1] $\boxed{\begin{array}{l} 2X + 3Y = 8 \\ 3X + 6Y = 15 \end{array}}$

1. La ecuación con el coeficiente de mayor valor absoluto en la columna del pivot [cualquier fila, columna 1] se intercambia con la ecuación en la fila del pivot.

$$\boxed{\begin{array}{l} 3X + 6Y = 15 \\ 2X + 3Y = 8 \end{array}}$$

2. Se reduce a 1 el coeficiente del término del pivot multiplicándolo por el inverso.

$$\boxed{\begin{array}{l} (1/3 * 3X) + (1/3 * 6Y) = (1/3 * 15) \\ 2X + 3Y = 8 \end{array}}$$

Ahora tenemos

$$\boxed{\begin{array}{l} x + 2Y = 5 \\ 2X + 3Y = 8 \end{array}}$$

El método de Gauss, Continuación

El método de Gauss, continuación

3. Se elimina los términos de la columna del pivot en todas las filas debajo de la fila del pivot.

3a. Multiplicar la fila del pivot por el coeficiente negado de la fila debajo del termino del pivot que será eliminado, en nuestro ejemplo es -2.

$$\boxed{\begin{array}{l} X + 2Y = 5 \\ 2X + 3Y = 8 \end{array}}$$

$$\boxed{-2} * \boxed{X + 2Y = 5} = \boxed{-2X + -4Y = -10}$$

Nota: no cambiar la fila del pivot, poner el resultado en una variable temporal.

3b. Agregar el resultado anterior a la fila que será eliminada.

3. Repetir los pasos 3ª y 3b para cada fila debajo del termino del pivot.

$$\boxed{\begin{array}{l} X + 2Y = 5 \\ 2X + 3Y = 8 \end{array}} + \boxed{-2X + -4Y = -10} = \boxed{\begin{array}{l} X + 2Y = 5 \\ -Y = -2 \end{array}} \quad \boxed{Y = 2}$$

4. Completar la triangularización repitiendo los pasos 1 a 3 tantas veces como sea necesario luego resolver los restantes términos utilizando sustitución.

$$\boxed{\begin{array}{l} x + 2y = 5 \\ Y = 2 \end{array}}$$

$$\boxed{\begin{array}{l} X + 4 = 5 \\ Y = 2 \end{array}}$$

$$\boxed{\begin{array}{l} X = ? \\ Y = 2 \end{array}}$$

El método de Gauss, Continuación

El algoritmo de Gauss

En una computadora, las ecuaciones se representan con una matriz que contiene los coeficientes.

X	Y	
2	3	8
3	6	15

El algoritmo es el mismo.

Dado:

X	Y	
2	3	8
3	6	15

1.

X	Y	
3	6	15
2	3	8

2.

X	Y	
1	2	5
2	3	8

3.

X	Y	
1	2	5
0	-1	-2

X	Y	
1	2	5
0	1	2

4.

X	Y	
1	2	5
0	1	2

X=1
Y=2

El método de Gauss, Continuación

Un algoritmo de Gauss

1. I = 1
2. LET $P=A_K$, I = max{$|A_{J,I}|$: 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 $A_{I,J} = A_{I,J} / A_{I,I}$ *(Get a one in the I,I position.)* SET $A_{I,I}$ = 1.
6. FOR L = I + 1 to N multiply row I by $-A_{L,I}$ 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

Instrucciones

Instrucciones

Antes de comenzar el programa 8, repasar el proceso PSP0.1 para asegurarse de comprenderlo. También, asegurarse de tener todas las entradas requeridas antes de comenzar con la fase de planificación.

Entregables

Cuando complete la etapa de postmortem, arme un archivo .zip para enviar al instructor, conteniendo lo siguiente.

- El archivo .mdb con sus datos del ejercicio.
- Código fuente del programa.
- Impresión de pantalla de las pruebas realizadas mostrando el resultado de las mismas.
- Captura de pantalla de la salida del contador de LOC aplicada al ejercicio 8.
- Archivo de texto o documento que contenga la descripción de que clases/módulos/unidades del código construido contienen las distintas categorías (Added, Modified, Delete, etc). Utilizando el contador de LOC que muestra el tamaño de dichas unidades de código.

Ejemplo:

36 LOC Base -> Clase Base.java y muestre el tamaño de la clase utilizando el contador de LOCS aplicado a la clase.

3 LOC Deleted -> En la clase Base.java

10 LOC Modified -> En la clase Base.java

15 LOC Added -> En la clase Base.java

40 LOC Reused -> Clases LibUno.java, LibDos.java y muestre el tamaño de la las mismas utilizando el contador de LOCS aplicado ambas clases.

5 LOC New Reused -> agregadas en NuevaLib.java

Tenga en cuenta que:

.- Deben programar de acuerdo a su estándar de codificación.

.- Deben utilizar el mismo proceso que en el ejercicio anterior, PSP 0.1.
Esto quiere decir que en planificación van a estimar la cantidad de LOC que esperan generar y si parten de algún LOC Base van a indicar cuántas líneas son. Luego, en postmortem, van a usar el contador de LOC que construyeron para contar la cantidad de LOC del ejercicio 8.

Criterios de evaluación para el programa 8

Criterios de Evaluación	Su reporte debe ser • completo	Los datos de su proceso deben ser • exactos • precisos • consistentes

Sugerencias

Recuerde entregar su reporte en fecha.

Mantenga la simplicidad del programa.

Si no está seguro de algo, consulte a su instructor

Debe producir y reportar sus propias estimaciones y datos reales, desarrollar su propio código y llevar adelante su propio juego de pruebas.

