

Programa 3

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 3

Descripción

Descripción

El presente instructivo cubre los siguientes temas

Sección	Página
Requerimientos del Programa 3	3
Regresión Lineal	5
Correlación	6
Calculo de Regresión Lineal y Correlación	7
Ejemplo	8
Instrucciones	10
Criterios de Evaluación	12

Programa 3 Requerimientos

Requerimientos Programa 3

Utilizando PSP0.1, escribir un programa que

- Calcule los parámetros de regresión lineal β_0 y β_1 y los coeficientes de correlación $r_{x,y}$ y r^2 para un conjunto de n pares de datos,
- Dada una estimación, x_k calcular la predicción, y_k donde
$$y_k = \beta_0 + \beta_1 x_k$$
- Extienda la lista encadenada desarrollada en el programa 1, para almacenar conjuntos de pares de reales.

La tabla 1 contiene datos históricos estimados y reales de 10 programas. Para el programa 11, el desarrollador ha estimado por proxy 386 LOC.

Probar a fondo el programa. Como mínimo, correr los cuatro casos siguientes de la prueba.

- Test 1: Calcular los parámetros de regresión y los coeficientes de correlación entre *estimated proxy size* y *actual added and modified size* en la tabla 1. Calcular *plan added and modified size* dada la estimación por proxy $x_k = 386$.
- Test 2: Calcular los parámetros de regresión y los coeficientes de correlación entre *estimated proxy size* y *actual development time* en la tabla 1. Calcular tiempo estimado dada la estimación por proxy $x_k = 386$.
- Test 3: Calcular los parámetros de regresión y los coeficientes de correlación entre *plan added and modified size* y *actual added and modified size* en la tabla 1. Calcular *plan added and modified size* dada la estimación por proxy $x_k = 386$.
- Test 4: Calcular los parámetros de regresión y los coeficientes de correlación entre *plan added and modified size* y *actual development time* en la tabla 1. Calcular tiempo estimado dada la estimación por proxy $x_k = 386$.

Los resultados esperados se muestran en la tabla 2.

Número de Programa	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

Resultados esperados

Test	Valores esperados					Valores Obtenidos				
	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

Regresión Lineal

Descripción

La regresión lineal es una manera óptima de aproximar una línea a un conjunto de datos. La línea de regresión lineal es la línea donde la distancia de todos los puntos a esa línea se reduce al mínimo. La ecuación de una línea se puede escribir como

$$y = \beta_0 + \beta_1 x$$

En la figura 1, la mayor línea de regresión tiene parámetros $\beta_0 = -4.0389$ y $\beta_1 = 0.1681$.

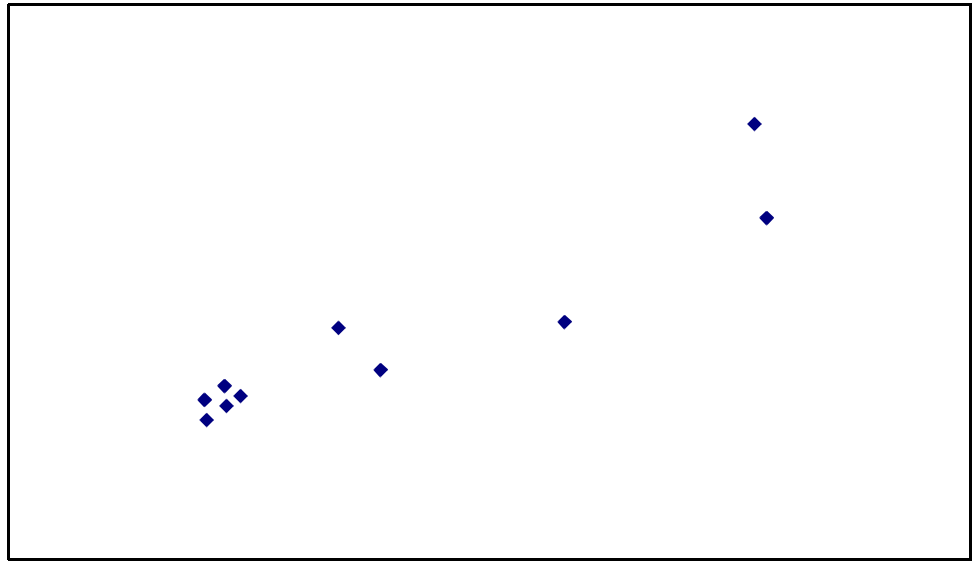


Figura 1

Correlación

Descripción

La correlación determina la relación entre dos conjuntos de datos numéricos.

Toma valores entre +1 y -1.

- Resultados cercanos a +1 implican una fuerte relación positiva. Cuando x crece también crece y.
- Resultados cercanos a -1 implican una fuerte relación negativa; Cuando x decrece también decrece y
- Resultados cercanos a 0 implican no relación.

Limitaciones de la correlación

La correlación no implica causa y efecto

Una fuerte correlación puede ser coincidencia

Desde 1840 hasta 1960, en U.S. ningún presidente electo en un año terminado en 0 sobrevivió su presidencia. Coincidencia o Correlación?

Muchas correlaciones por coincidencia se pueden encontrar en datos históricos de procesos.

Para utilizar una correlación, se debe entender la relación de causar-y-efecto en el proceso.

Cálculo de regresión y Correlación

Cálculo de regresión y correlación

Las fórmulas para calcular los parámetros de regresión β_0 y β_1 son

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

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

Las fórmulas para calcular los coeficientes de correlación $r_{x,y}$ y r^2 son

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

$$r^2 = r * r$$

donde

- i identifica los elementos de los conjuntos x e y
- x e y son conjuntos de datos
- n es el número de ítems de los conjuntos x e y
- x_{avg} es la media de los valores x
- y_{avg} es la media de los valores y

Ejemplo

Ejemplo

En este ejemplo, se calcularán los parámetros de regresión (β_0 y β_1) y los coeficientes de correlación $r_{x,y}$ y r^2 a partir de los datos en la tabla 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

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

1. Hay 10 items en cada conjunto de datos por lo tanto $n = 10$.
2. Se calculan las sumatorias que participan de las fórmulas.

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	$\sum_{i=1}^{10} x_i = 3828$	$\sum_{i=1}^{10} y_i = 6389$	$\sum_{i=1}^{10} x_i^2 = 2540284$	$\sum_{i=1}^{10} x_i y_i = 4303108$	$\sum_{i=1}^{10} y_i^2 = 7604693$
	$x_{avg} = \frac{3828}{10} = 382.8$	$y_{avg} = \frac{6389}{10} = 638.9$			

Ejemplo (cont.)

3. Se sustituyen los valores en las fórmulas

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

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

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

$$r_{x,y} = \frac{18573988}{\sqrt{[10749256][35227609]}} \quad r_{x,y} = \frac{18573988}{19459460.1}$$

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

$$r^2 = 0.9111$$

4. Sustituimos β_0 en la fórmula

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

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

5. Despejamos y_k de la fórmula $y_k = \beta_0 + \beta_1 x_k$

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

Instrucciones

Instrucciones

Antes de comenzar el programa 3, 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 la mismas.
- Captura de pantalla de la salida del contador de LOC aplicada al ejercicio 3.
- 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 ejercicio3

.- Sobre las pruebas que pide la letra:

La idea es que en cada test se toman dos de las columnas de la tabla 1 y se ve su correlación. Y luego obtenidos los parámetros de correlación, dado un x particular se calcula el valor esperado de y según esa correlación.

En el test 1: Se pide tomar como entrada las columnas 2 y 4. Y luego tomar el valor determinado $x=386$ y determinar el y esperado para ese x

En el test2: Se pide tomar como entrada las columnas 2 y 5. Y luego tomar el valor determinado $x=386$ y determinar el y esperado para ese x

En el test3: Se pide tomar como entrada las columnas 3 y 4. Y luego tomar el valor determinado $x=386$ y determinar el y esperado para ese x

En el test4: Se pide tomar como entrada las columnas 3 y 5. Y luego tomar el valor determinado $x=386$ y determinar el y esperado para ese x

Criterios de evaluación para el programa 3

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 esta 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.
