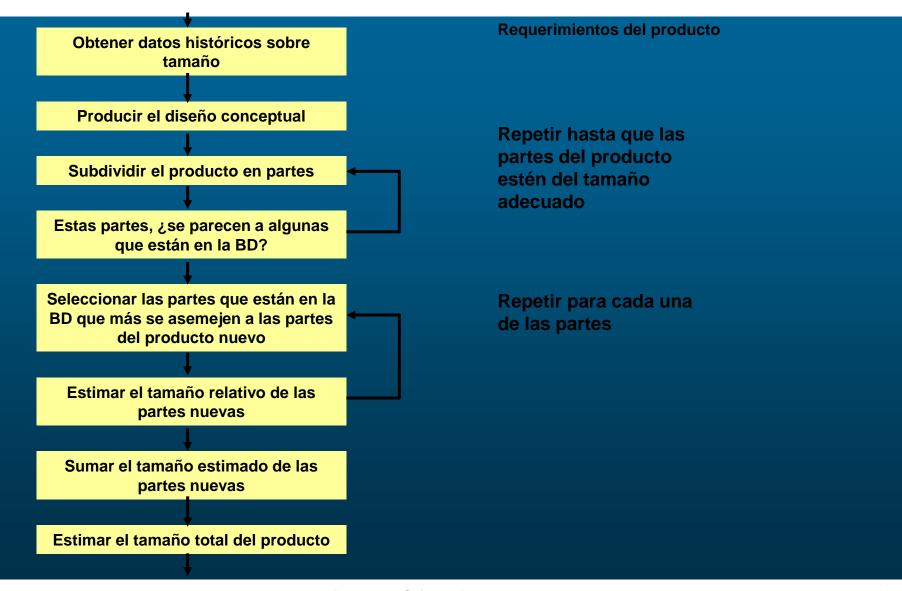
PSP 1

Estimación de tamaño 2a Parte

El proceso de Planeación IV Estimación del Tamaño del Software El Método PROBE

Introducción a la Estimación de Tamaño



El Método de Estimación PROBE



Ejemplo del Método PROBE

Ver ejemplo.

El Diseño Conceptual

Se requiere un diseño conceptual

- + Para relacionar los requerimientos al producto.
- + Para definir los elementos del producto que tendrán las funciones requeridas.
 - + Para estimar el tamaño de lo que será construido.

Para aquellos diseños ya entendidos, el diseño conceptual se puede realizar rápidamente.

Si no entiendes el diseño, no sabes lo suficiente como para hacer una estimación.

En el diseño detallado, no es obligatorio seguir el diseño conceptual.

Identificar los Objetos [1/3]

- + Donde sea posible, seleccionar entidades de la aplicación.
- + Juzgar cuantos métodos tendrá cada objeto.
- + Determinar el tipo de objeto (de datos, de cálculo, de archivo, de control, etc.).
- + Juzgar el tamaño relativo de cada objeto: Muy pequeño (VS), Pequeño (S), Mediano (M), Largo (L) o Muy Largo (VL).

Identificar los Objetos [2/3]

- + Con datos históricos de objetos, determinar el tamaño en LOC por método de cada objeto.
- + Multiplicar el número de métodos para obtener las LOC estimadas del objeto.
- + Juzgar cuales objetos serán añadidos a la biblioteca de reutilización y anotarlos como "Nuevos Reutilizados".

Identificar los Objetos [3/3]

- + Cuando no se encuentre un tipo existente a los objetos, quizá esté compuesto.
- Asegurarse que están en un buen nivel de granularidad.
 - Refinar aquellos que no lo estén.
- + Estar atentos a nuevos tipos de objetos.

Identificación de objetos Ejemplo [1/2]

+ Tres objetos son identificados con el número de métodos, tamaño relativo y LOC:

Objetos Nuevos	Tipo	Métodos	Tamaño	LOC
Matriz	Dato	13	M	115
Sistema Lineal	Cálculo	8	L	197
Lista Ligada	Datos	3	S	49*

Identificación de objetos Ejemplo [2/2]

- + Adicionalmente, se identificaron 2 objetos a reutilizar:
 - Lista ligada 73 LOC.
 - Entrada de datos 96 LOC.
- + A la lista ligada se le añade un nuevo método de 49 LOC.

Estimando el Tamaño del Programa [1/4]

- + El tamaño total del programa consiste de:
- Código nuevo desarrollado (ajustado con los parámetros de regresión)
 - Código reutilizado de la biblioteca
 - Código base de versiones previas, menos borrados
- + El código nuevo desarrollado consiste de:
 - Adiciones a la base (BA)
 - Objetos nuevos (NO)
- Código modificado (M) LOC base que son cambiadas

Estimando el Tamaño del Programa [2/4]

+ Calcular las LOC estimadas nuevas y cambiadas (N) con las LOC estimadas de los objetos (E)

- -E = BA + NO + M
- + Usar la regresión lineal para calcular
 - LOC nuevas y cambiadas
 - Tiempo de desarrollo

$$y_k = \beta_0 + \beta_1 * x_k$$

$$Nuevasycambiadas = \beta_0 + \beta_1 * (BA + NO + M)$$

$$Tiempo = \beta_0 + \beta_1 * (E)$$

Estimando el Tamaño del Programa [3/4]

- + Calcular los parámetros de regresión β_0 y β_1 usando los datos de los programas desarrollados previamente.
- + Para los valore de x usar las LOC estimadas de los objetos (E).
- + Para los valores de y usar:
- LOC actuales nuevas y cambiadas de cada programa finalizado.
- El tiempo total de desarrollo actual de cada programa finalizado.

Estimando el Tamaño del Programa [4/4]

- + El código usado de la biblioteca de reutilización debe ser contado e includido en la estimación total de LOC.
- + El código base consiste de:
 - LOC de la versión previa
 - Código borrado
 - Código modificado
- + Mientras el código base es una forma de reutilización, solamente el código no modificado de la biblioteca de reutilización es contado como LOC reutilizadas.

Completando la Estimación [1/5]

- + La estimación completa consiste de:
- Las LOC estimadas nuevas y cambiadas calculadas con los parámetros de regresión.
- Los intervalos de predicción (UPI y LPI), al 70% y 90% para las LOC nuevas y cambiadas.
- Las LOC totales, considerando nuevas, base, reutilizadas, borradas y modificadas.
- Las LOC reutilizadas nuevas proyectadas a ser añadidas a la biblioteca de reutilización.

Completando la Estimación [2/5]

- + La estimación completa también incluye:
- El tiempo de desarrollo estimado calculado con los parámetros de regresión.
- Los intervalos de predicción superior e inferior al 70% y 90% para el tiempo de desarrollo.

Completando la Estimación Ejemplo [3/5]

Programa Base (B)	695 LOC
Borrado (D)	0 LOC
Modificado (M)	5 LOC
Adiciones a la Base (BA)	0 LOC
Objetos Nuevos: NO = 115 + 197 + 49 =	361 LOC
Programas reutilizados	169 LOC

Completando la Estimación Ejemplo [4/5]

Usar los parámetros de regresión de tamaño para calcular las LOC nuevas y cambiadas (N):

$$Nuevasy cambiadas = \beta_0 + \beta_1 * (BA + NO + M)$$

Código añadido: BA + NO + M = 366 LOC

Nueva y cambiada: N = 62 + 366 * 1.3 = 538 LOC

Total: T = 538 + 695 - 5 + 169 = 1397 LOC

Completando la Estimación **Ejemplo** [5/5]

Usar los parámetros de regresión de tiempo para calcular el tiempo de desarrollo:

$$TiempodeDesarrollo = \beta_0 + \beta_1 * (BA + NO + M)$$

B0 = 108 y B1 = 2.95Tiempo de desarrollo = 108 + 2.95 * 366 = 1186 min

Cálculo de los intervalos de predicción

- + Calcular los rangos de predicción para el tamaño y el tiempo.
- + Calculas los intervalos de predicción superior e inferior (UPU/LPI) para el tamaño.
 - UPI = N + rango = 538 + 235 = 773 LOC
 - LPI = N rango = 538 235 = 303 LOC
- + Calculas los intervalos de predicción superior e inferior (UPU/LPI) para el tiempo.
 - UPI = Tiempo + rango = 1186 + 431 = 1617 min
 - LPI = Tiempo rango = 1186 431 = 775 min

Después del desarrollo

+ Después del desarrollo, medir y registrar el tamaño actual de:

- LOC base, borradas y modificadas
- Adiciones la base
- Objetos nuevos
- Objetos reutilizados
- + Si el objeto fue:
- Planeado pero no desarrollado, su tamaño actual es 0 LOC.
- -Desarrollado pero no planeado, su tamaño planeado es 0 LOC.

Para utilizar PROBE se necesita

- + Datos históricos de los objetos divididos por tipo.
- + Factores de estimación para los tamaños relativos de cada tipo de objeto.
- + Parámetros de regresión para predecir el tamaño y el tiempo basados en datos históricos.
- + Datos históricos estimados y actuales del tamaño y el tiempo de desarrollo para calcular los parámetros de regresión y los intervalos de predicción.

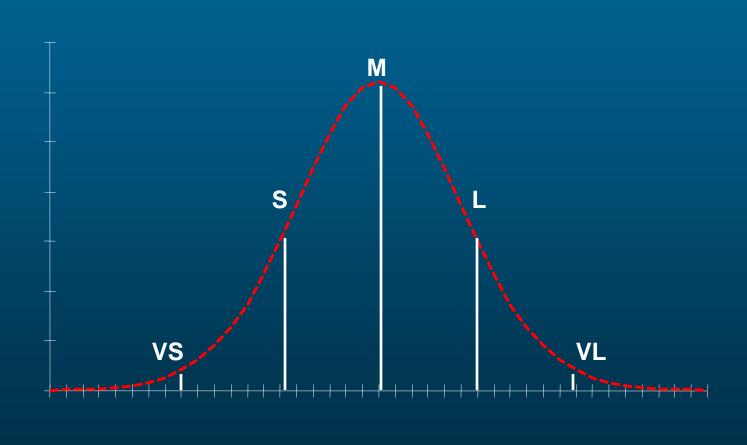
Datos históricos de los objetos

- + El tamaño de los objetos es altamente variable:
 - Depende del lenguaje.
 - Influenciado por el estilo de diseño.
 - Se puede normalizar por varios métodos
- + Seleccionar tipos básicos
 - Lógicos, de control.
 - Entrada/Salida, Archivos, Despliegue
 - Datos, texto, cálculos
 - Manejo de errores, configuración

Factores de Estimación para los Objetos

- + Se buscan los rangos de tamaño para cada tipo que ayude a juzgar los tamaños de los nuevos objetos.
- + Para calcular estos rangos de tamaño
 - Tomar la media.
 - Tomar la desviación estándar.
 - Muy pequeño: VS = media 2 * desviación estándar
 - Pequeño: S = media desviación estándar
 - Medio: M = media
 - Largo: L = media + desviación estándar
 - Muy largo: VL = media + 2 * desviación estándar

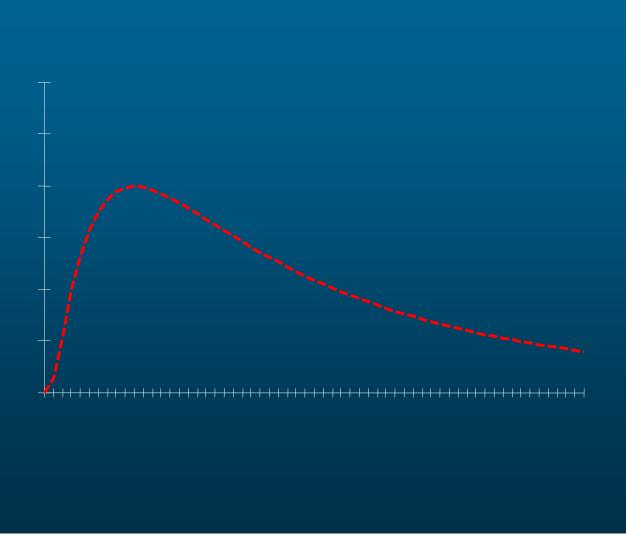
Distribución normal: Rangos de tamaño



Log de la distribución normal

- + Estos rangos de tamaño asumen que los datos del objeto están normalmente distribuidos.
- + Si los datos están distribuidas log(normalmente), tomar el log de los datos antes de realizar los cálculos de los rangos de tamaño.
- + Luego, después de haber calculado los rangos de tamaño, tomar el antilog para obtener los factores en LOC.

Log de la distribución normal



Factores de Estimación [1/3]

- + Se tienen los siguientes datos de un tipo de objeto.
 - 1 objeto, 3 métodos, 39 LOC en total
 - 1 objeto, 5 métodos, 127 LOC en total
 - 1 objeto, 2 métodos, 64 LOC en total
 - 1 objeto, 3 métodos, 28 LOC en total
 - 1 objeto, 1 métodos, 23 LOC en total
 - 1 objeto, 2 métodos, 44 LOC en total
- + Las LOC por método son: 13, 25.4, 32, 9,333, 23, 22

Factores de Estimación [2/3]

- + El logaritmo de estos datos son:
 - 2.565, 3.235, 3.466, 2.234, 3.135, 3.091
 - El promedio es 2.954
 - La desviación estándar es 0.421
- + Los valores de los logaritmos para los rangos de tamaño son:
 - Muy Largo VL: 2.95 + 2*0.42 = 3.79
 - Largo L: 2.95 + 0.42 = 3.37
 - Medio M: 2.95
 - Pequeño S: 2.95 -0.42 = 2.53
 - Muy pequeño VS: 2.95 2*0.42 = 2.11

Factores de Estimación [3/3]

- + Con estos logaritmos de rangos de tamaño, los rangos de LCO son obtenidos tomando los antilogaritmos
 - Muy Largo VL: exp(3.79) = 44.3
 - Largo L: exp(3.37) = 29.1
 - Medio M: exp(2.95) = 19.1
 - Pequeño S: exp(2.53) = 12.6
 - Muy pequeño VS: exp(2.11) = 8.3
- + Repetir estos cálculos para cada tipo de objeto

Ejemplo de Rangos de Tamaño de objetos en C++

Tipo	VS	S	M	L	VL
Cálculo	2.34	5.13	11.25	24.66	54.04
Dato	2.60	4.79	8.84	16.31	30.09
I/O	9.01	12.06	16.15	21.62	28.93
Lógico	7.55	10.98	15.98	23.25	33.83
Configura ción	3.88	5.04	6.56	8.53	11.09
Texto	3.75	8.00	17.07	36.41	77.66

Estimaciones basadas en estadísticas

PROBE usa datos históricos, regresión lineal e intervalos de predicción para producir una estimación de precisión conocida.

La regresión provee el mejor ajuste posible o la mínima variación de una línea en un conjunto de datos.

La varianza de estos datos es usada para determinar los intervalos de predicción.

Bases para la estimación [1/2]

Con PROBE, las estimaciones están basadas en uno de los 4 métodos siguientes:

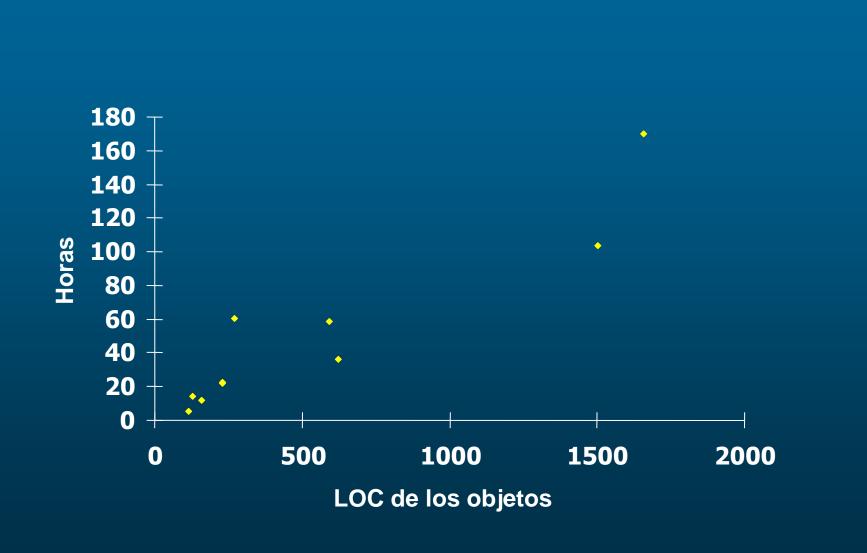
- Método A: regresión con las LOC estimadas de los objetos.
- Método B: regresión con las LOC estimadas nuevas y cambiadas.
 - Método C: el método promedio.
 - Método D: juicio con ingeniería.

Bases para la estimación [2/2]

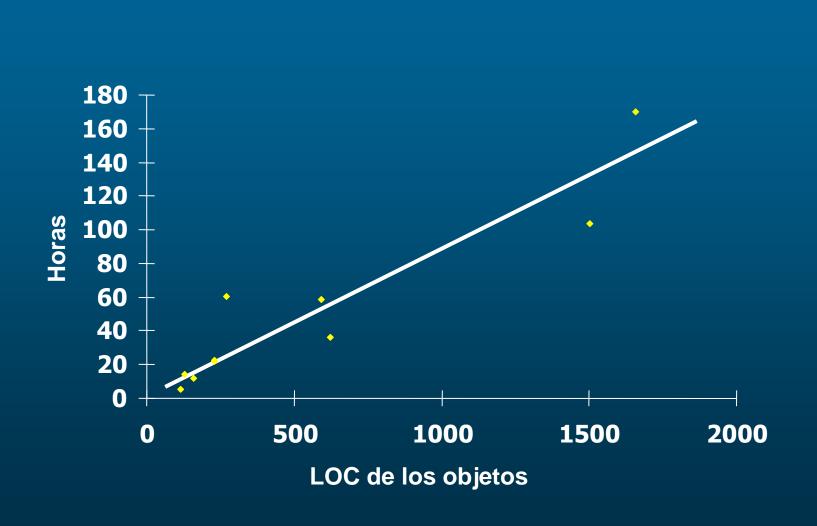
Para usar el método de regresión se necesita:

- Una cantidad razonable de datos históricos.
- Datos que se correlacionan.
- Parámetros razonables.

Datos de Desarrollo



Datos de Desarrollo



Los parámetros de Regresión

Los parámetros de regresión están calculados usando la siguiente fórmula:

$$\beta_{1} = \frac{\sum_{i=1}^{n} x_{i}y_{i} - nx_{avg}y_{avg}}{\sum_{i=1}^{n} x_{i}^{2} - n(x_{avg})^{2}}$$

$$\beta_{0} = y_{avg} - \beta_{1}x_{avg}$$

donde x y y son los datos históricos, n es el número de datos

Método A: Regresión con LOC estimadas

El método A usa la relación entre las LOC estimadas de los objetos y

- Las LOC actuales nuevas y cambiadas.
- Tiempo de desarrollo actual.

Los criterios para usar este método son:

- Tres o más datos que estén correlacionados.
- Parámetros de regresión razonables.

Método B: Regresión con LOC nuevas y cambiadas estimadas

El método B usa la relación entre las LOC nuevas y cambiadas estimadas y

- LOC nuevas y cambiadas actuales.
- Tiempo de desarrollo actual.

Los criterios para usar este método son.

- Tres o más datos que estén correlacionados.
- Parámetros de regresión razonables.

Método C: Método del Promedio

El método C usar un radio para ajustar el tamaño o el tiempo con promedios históricos.

El método del promedio es fácil de usar y requiere de solamente un dato.

Método D: Juicio con Ingeniería

Se usa en ausencia de datos históricos.

Se usa el juicio para estimar las LOC nuevas y cambiadas a partir de las LOC estimadas de los objetos.

Usar el juicio para estimar el tiempo requerido de desarrollo.

Es usado cuando los métodos A, B o C no pueden ser utilizados.

El intervalo de predicción

El intervalo de predicción provee un rango aproximado sobre la estimación.

- Un intervalo de predicción del 90% da el rango con el cual el 90% de las estimaciones caerán probablemente.
 - No es pronóstico, solamente una expectativa.
- Solamente aplica si la estimación se comporta como los datos históricos.

Es calculado con los mismos datos usados para calcular los parámetros de regresión.

El intervalo de predicción

El intervalo de predicción inferior (LPI) y el intervalo de predicción superior (UPI) son calculados con las estimaciones de tamaño y el rango, donde:

- LPI = Estimación Rango
- UPI = Estimación + Rango

Range =
$$t(\alpha/2, n-2)\sigma$$

$$1 + \frac{1}{n} + \frac{(x_k - x_{avg})^2}{\sum_{i=1}^{n} (x_i - x_{avg})^2}$$

El intervalo de predicción

La distribución t es para

- La distribución por los dos lados (alpha/2)
- n-2 grados de libertad

Sigma es la desviación estándar de la regresión lineal de los datos.

$$\sigma = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}$$

El intervalo de predicción La distribución t

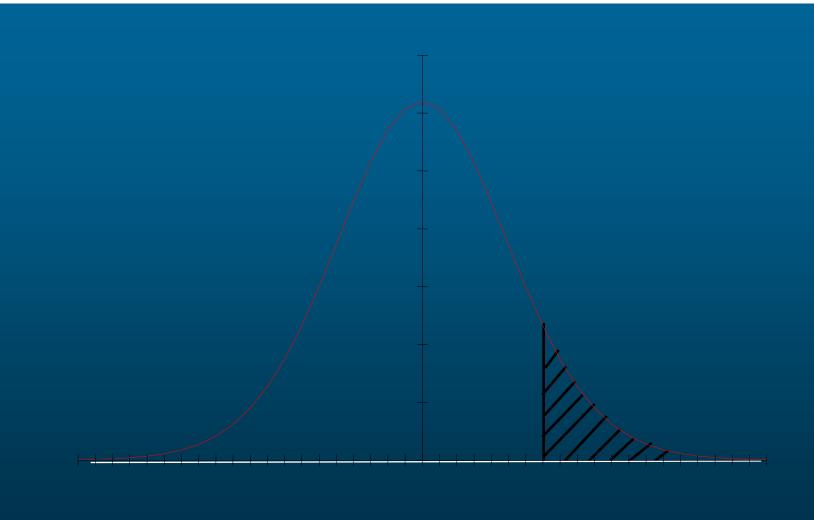
La distribución t

- Es similar a la distribución normal.
- Tiene colas más gordas.
- Es usada para la estimación de parámetros estadísticos con datos limitados.

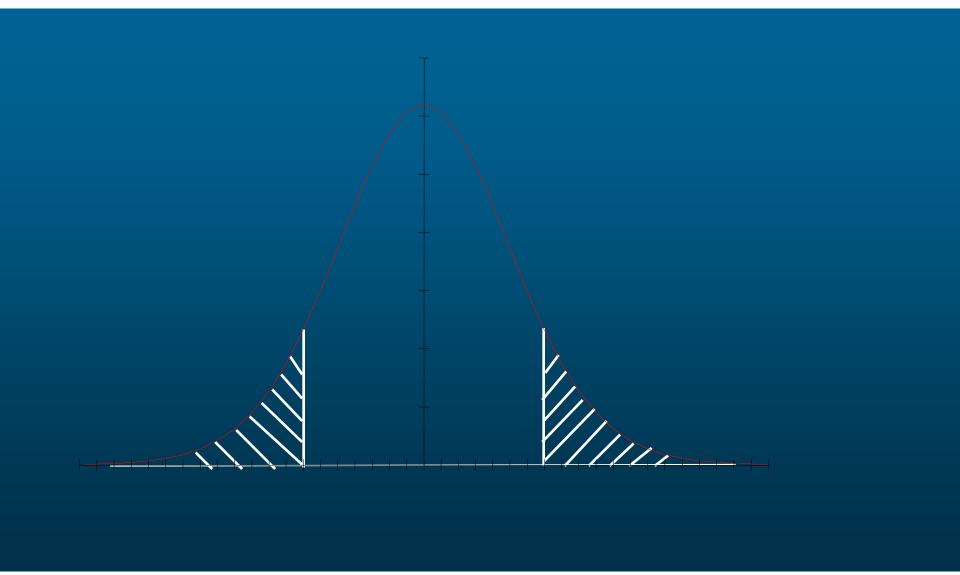
Las tablas de distribución t

- Típicamente dan un rango de probabilidad de un solo lado.
- Nosotros usamos valores de los dos lados en el cálculo de los intervalos de predicción.

El intervalo de predicción Distribución t en un solo lado



El intervalo de predicción Distribución t en los dos lados



El intervalo de predicción Valores de la Distribución t

Las tablas de distribución t

- Dan el valor de probabilidad p desde menos infinito ax
- Para el valor de un solo lado de la cola (el valor de interés), tomar 1 – p
- Para el valor de los dos lados (con las dos colas), tomar 1 - 2*(1-p) = 2p - 1
 - Buscar bajo p = 85% para un intervalo del 70%
 - Buscar bajo p = 95% para un intervalo del 90%

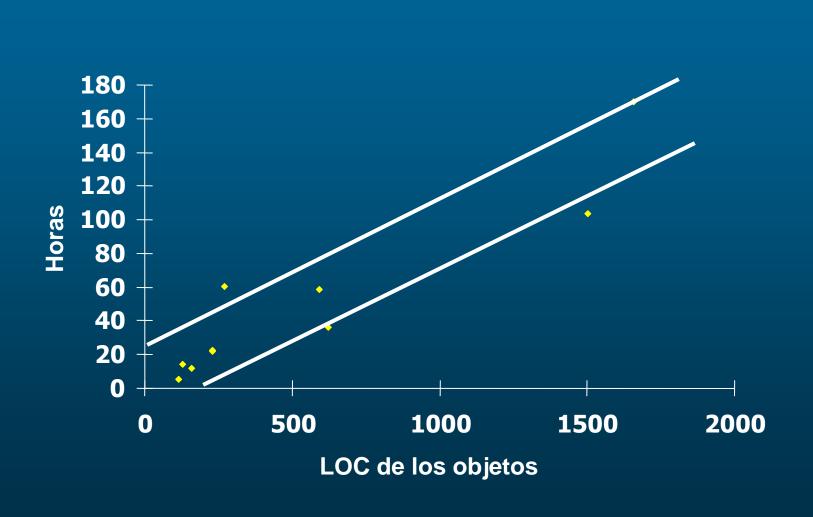
Ejemplo de intervalo de predicción

Calcular el rango de los datos históricos rango = 235 LOC

Intervalo de predicción superior (UPI) UPI = N + rango = 538 + 235 = 773 LOC

Intervalo de predicción inferior (LPI) LPI = N - rango = 538 - 235 = 303 LOC

Datos de Desarrollo



- El tamaño y tiempo proyectados se calculan a partir de estimaciones históricas utilizando PROBE.
- Se selecciona uno de cuatro métodos PROBE posibles, A, B, c, o D.
- El método seleccionado depende de la cantidad de datos históricos que se tengan.
- La mayoría de las herramientas PSP, automatizan los cálculos.

- Si se tienen suficientes estimaciones por proxy y datos reales acerca de piezas agregadas y modificadas (tres o más datos correlacionados), se utiliza el método A.
- Si no se tiene suficientes datos estimados por proxy, pero se tiene suficientes datos planificados y reales acerca de piezas agregadas y modificadas (tres o más datos correlacionados), se utiliza el método B.
- Si se tiene datos escasos o no se correlacionan utilizar el método C.
- Si no se tiene ningún dato histórico, se debe utilizar el método D.

- Método A
- Utilizando regresión lineal, calcular los parámetros B0 y B1 a partir de datos estimados por proxy y datos reales de código agregado y modificado.
- Si el valor absoluto de B0 no es cercano a 0 (menos de un 25% valor esperado del nuevo programa), o 1 no es cercano a 1.0 (entre 0.5 y 2.0), utilizar el método B.

- Método B
- Utilizando regresión lineal, calcular los parámetros B0 y B1 a partir de datos planificados y reales de código agregado y modificado.
- Si el valor absoluto de B0 no es cercano a 0 (menos de un 25% valor esperado del nuevo programa), o B1 no es cercano 1.0 (entre 0.5 y 2.0), utilizar el método C.

- Método C (Promedio o Average)
- Si se tiene algún dato planificado sobre código agregado y modificado,
 - B0 = 0
 - Se toma el acumulado de todos los proyectos de código y tiempo, hasta el momento y se calcula la razón.

$$\beta_{1tiempo} = \sum_{i=1}^{n} (tiempo_i) / \sum_{i=1}^{n} (tamaño_i)$$

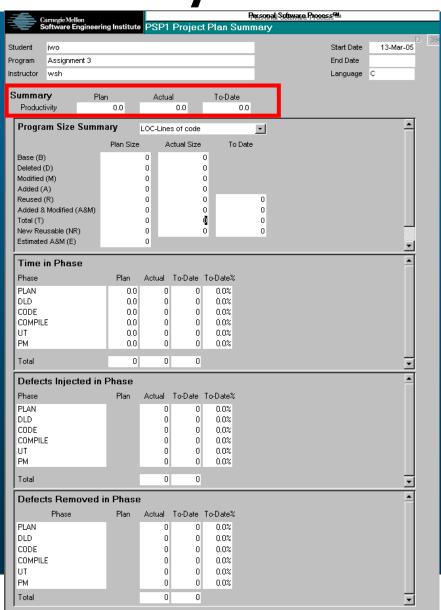
$$\beta_{1tamaño} = \sum_{i=1}^{n} (tamaño_real_i) / \sum_{i=1}^{n} (tamaño_planificado_i)$$

- Método D
- Si no se tienen datos históricos utilizar el juicio profesional para realizar las estimaciones.

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	

PSP1 Resumen del Plan de Proyecto -1

- •El Resumen del Plan de Proyecto incluye una nueva sección de resumen.
- La sección resumen incluye productividad planificada, real, y a la fecha.
- Todos los valores excepto el tamaño real total son calculados



PSP1 Resumen del Plan de Proyecto -2

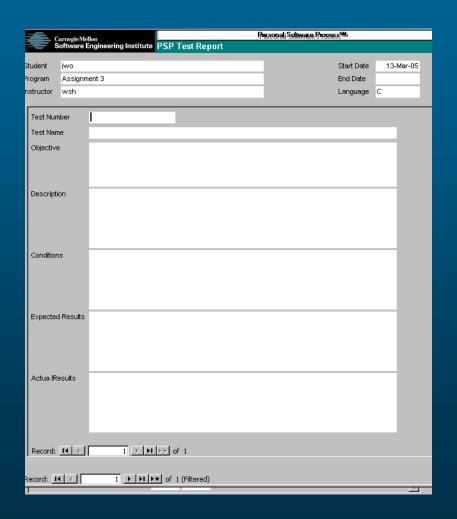
- Productividad planificada, real, y a la fecha
 - -Es el número de unidades de tamaño añadido y modificado por hora.

Productividad = Tamaño A&M / Tiempo Total de Desarrollo * 60

Summary	Plan	Actual	To-Date
Productivity	0.0	0.0	0.0

Plantilla de Reporte de Prueba

- Use este formulario para registrar información de cada prueba.
 - Qué prueba fue ejecutada
 - Cuál información de prueba fue usada
 - Resultados que fueron obtenidos
- Útil para
 - Desarrollo y registro de casos de prueba
 - Realización de pruebas de regresión



Ejercicio 4A Usar PSP1

Escribir un programa para calcular los parámetros de regresión lineal para la estimación del tamaño de un conjunto de n programas de los cuales se tienen los datos históricos de las LOC y las LOC nuevas y cambiadas. Mejorar la lista ligada del programa 1A para que guarde los n registros de datos, donde cada registro contiene un número real.

Probar completamente el programa. Como mínimo, usar este programa para calcular los parámetros β para 3 casos. Para el primero, usar los datos de la siguiente tabla para las LOC estimadas de los objetos y las LOC nuevas y cambiadas (β 0=-22.55 y β 1 = 1.7279). Para el segundo utilizar las LOC nuevas y cambiadas estimadas con las LOC nuevas y cambiadas actuales (β 0=-23.92 y β 1 = 1.4310). Para el tercero utilizar los datos de tus programas 2A, 3A y 4A.

Presentar los resultados en el formato correspondiente.

Ejercicio 4A Datos

Número del programa	LOC estimadas del objeto	LOC nuevas y cambiadas estimadas	LOC nuevas y cambiadas actuales
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
Suma	3828	4632	6389
Promedio	382.8	463.2	638.9

Ejercicio 4A Formato de reporte

Prueba	Resultados esperados		Resultados actuales	
	B ₀	B ₁	B ₀	β ₁
Tabla: Objetos estimados vs. LOC nuevas y cambiadas actuales				
Tabla. LOC nuevas y cambiadas estimadas vs. LOC nuevas y cambiadas actuales				
Programa 2A, 3A, 4A LOC nuevas y cambiadas estimadas vs. LOC nuevas y cambiadas actuales				

Orden de entrega de las formas

Resumen del Plan del Proyecto nivel PSP1.

FRPP del proyecto anterior.

Reporte de pruebas.

Formas de PIPs.

Plantilla de estimación de tamaño.

Forma de Registro de Tiempos.

Forma de Registro de Defectos.

Estándar de codificación

Código Fuente del Programa.

Pantallas de la interfase gráfica.

Pantallas de los resultados.

Tabla de resultados.