

# Personal Software Process<sup>SM</sup>

## Estimación con PROBE II

This material is approved for public release. Distribution is limited by the Software Engineering Institute to attendees.

Sponsored by the U.S. Department of Defense

© 2006 by Carnegie Mellon University



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Tópicos

El intervalo de predicción

Organizando los datos del proxy

Estimando con datos limitados

Estimando la precisión

Estimando consideraciones



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# El Intervalo de Predicción

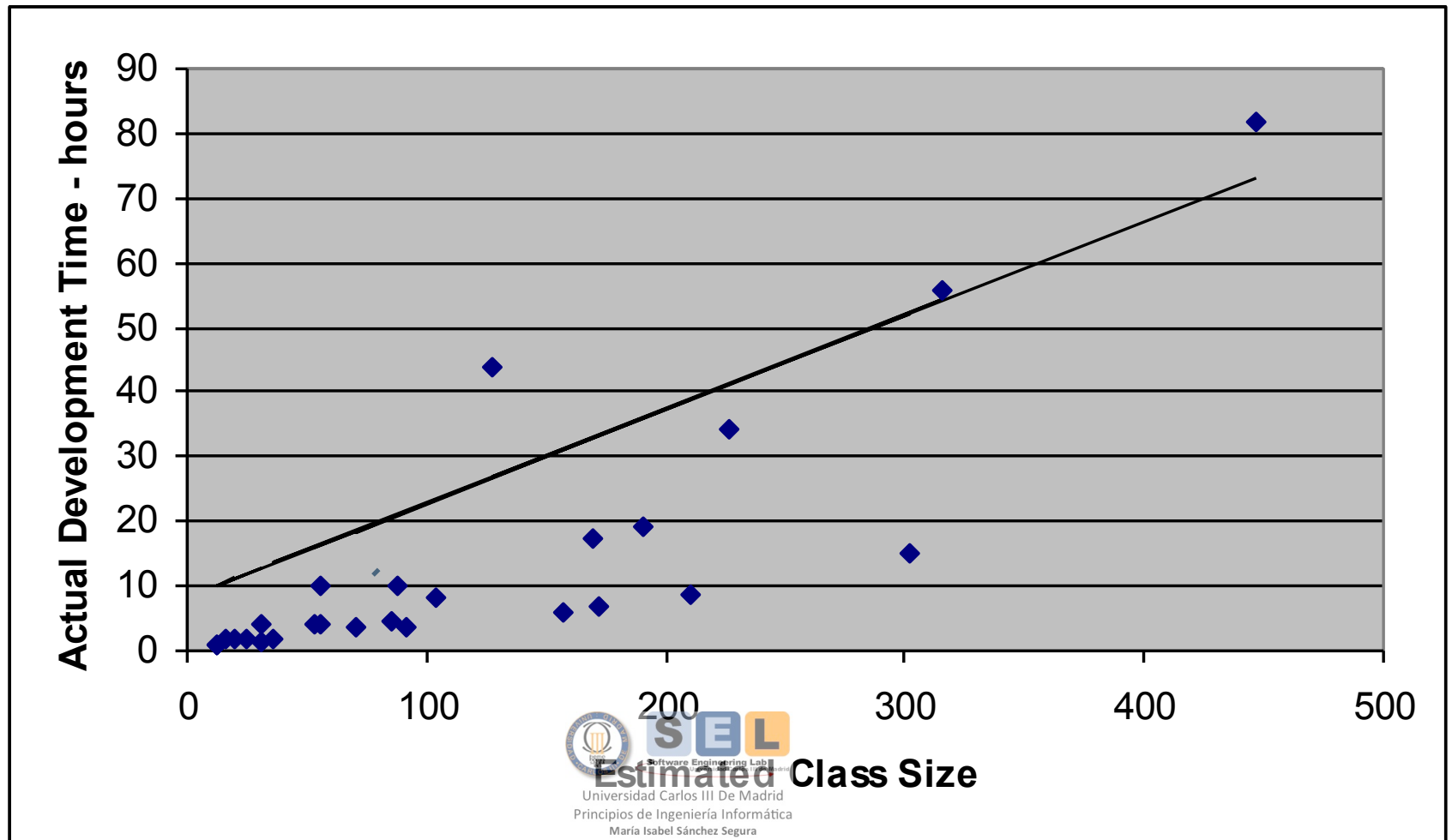
El intervalo de predicción provee un rango probable cercano a la estimación.

- Un intervalo de predicción del 70% da el rango con el cual el tamaño real probablemente sea 70% del tiempo.
- El intervalo de predicción no es un pronóstico, sólo una expectativa
- Aplica sólo si la estimación se comporta similar a los datos históricos
- Es calculado a partir de los mismos datos para calcular los parámetros de regresión.



# Ejemplo de Intervalo de Predicción

27 programas C++



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# El cálculo del Rango

El rango define la probabilidad de error en torno a la proyección en el que el valor real es probable que caiga

Datos muy dispersos tendrá un rango más amplio que los datos agrupados estrechamente

$$\text{Range} = t(p, df) \sigma \sqrt{1 + \frac{1}{n} + \frac{(x_k - x_{avg})^2}{\sum_{i=1}^n (x_i - x_{avg})^2}}$$

Las variables son

$n$  - número de puntos de datos

$\sigma$  - desviación estándar en torno a la regresión lineal

$t(p, df)$  – valor de la distribución  $t$  para la probabilidad  $p$  (70%) y  $df$

$(n-2)$  grados de libertad

$x$  – dato:  $k$  - estimación,  $i$  – un punto de datos, y  $avg$  – promedio del dato

# Cálculo de la Desviación Estándar

La desviación estándar mide la variabilidad de los datos en torno a la regresión lineal.

Datos muy dispersos tendrán una desviación estándar más alta que los datos agrupados cercanamente

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

La desviación estándar  $\sigma$  es la raíz cuadrada de la varianza.



# Calcular el Intervalo de Predicción

Calcular el rango de predicción para el tamaño y el tiempo para el ejemplo de la presentación anterior (transparencias 42 y 43).

Calcular los intervalos de predicción superior (UPI) e inferior (LPI) para el tamaño.

- $UPI = P + \text{Rango} = 538 + 235 = 773 \text{ LOC}$
- $LPI = P - \text{Rango (o 0)} = 538 - 235 = 303 \text{ LOC}$

Calcular los intervalos de predicción UPI y LPI para el tiempo.

- $UPI = \text{Tiempo} + \text{Rango} = 1186 + 431 = 1617 \text{ min.}$
- $LPI = \text{Tiempo} - \text{Rango (o 0)} = 1186 - 431 = 755 \text{ min.}$

# Organizando información del Proxy -1

Para realizar una estimación

- fraccionar el producto planificado en partes
- relacionar estas partes planificadas con las partes que ya se construyeron.
- use el tamaño de las partes construidas previamente para estimar los tamaños de las nuevas partes

Para hacer esto, necesita rangos de tamaño para los tipos de partes que desarrolla a menudo

Por cada tipo de producto, también necesita rangos de tamaño que ayude a discriminar los tamaños de las nuevas partes.



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán



# Organizando información del Proxy -2

Para determinar los rangos de tamaño, empiece con la información de la parte.

Asuma que tiene la siguiente información

- clase A, tres ítems (o métodos), 39 total LOC
- clase B, cinco ítems, 127 total LOC
- clase C, dos ítems, 64 total LOC
- clase D, tres ítems, 28 total LOC
- clase E, un ítem, 12 LOC
- clase F, dos ítems, 21 total LOC

El LOC por ítem es 13, 25.4, 32, 9.333, 12, 10.5.

El objetivo es definir rangos de tamaño que se aproximen a nuestra intuición del tamaño.

# Organizando información del Proxy -3

Para producir los rangos de tamaño, ordene la información como sigue.

El LOC ordenado por ítem: 9.333, 10.5, 12, 13, 25.4, 32.

Organice los datos como sigue.

- Escoja el ítem más pequeño como muy pequeño
- Seleccione el ítem más grande como muy grande:  $VL = 32$ .
- Escoja el ítem del medio como medio:  $M = 12$  o  $13$ .
- Para los rangos grandes y pequeños, seleccione los puntos medios entre  $M$  y  $VS$  y  $My VL$ : 10.9, y 22.25.

Mientras estos pueden ser rangos útiles, no son probablemente estables

Esto es, puntos de datos adicionales probablemente resulten en sustanciales ajustes del rango de tamaño.

# Rangos de Tamaño Intuitivo -1

Juzgando el tamaño, nuestra intuición está generalmente basada en una distribución normal.

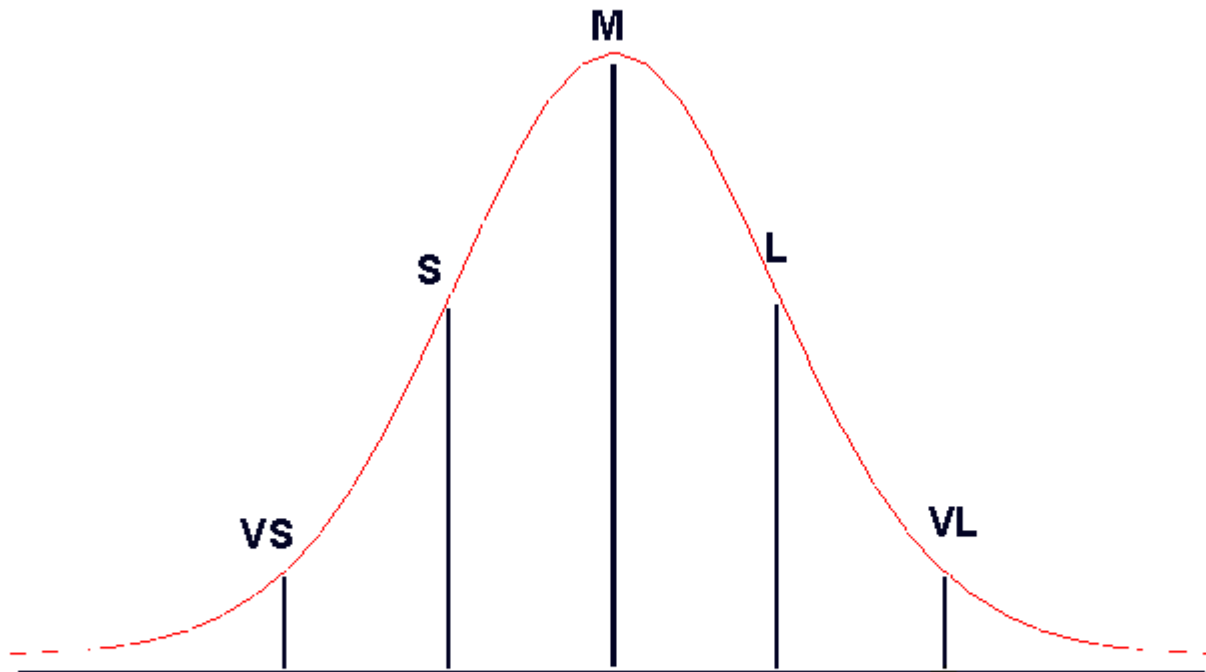
Esto es, pensamos en algo como de tamaño medio si la mayoría de estos ítems se refieren a ese mismo tamaño.

Consideramos algo a ser muy grande si éste es más grande que casi todos los ítems en su categoría.

Cuando los ítems son distribuidos de esa forma, es llamado una distribución normal.

Con datos distribuidos normalmente, los rangos deberían permanecer razonablemente estables con la adición de nuevos puntos de datos

# Rangos de Tamaño Intuitivo -2



Distribución normal



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Rangos de Tamaño Intuitivo -3

Con un gran volumen de datos, podrías calcular la media y desviación estándar de dichos datos

Para los rangos de tamaño

- Medio sería el valor medio
- Grande sería la media más una desviación estándar
- Pequeño sería la media menos una desviación estándar
- Muy grande sería la media más dos desviaciones estándar
- Muy pequeño sería la media menos dos desviaciones estándar

Este método proveería de adecuados rangos de tamaños intuitivos si la información estuviera normalmente distribuida.

# Distribución de la información del tamaño

Datos del tamaño del programa no están normalmente distribuidos

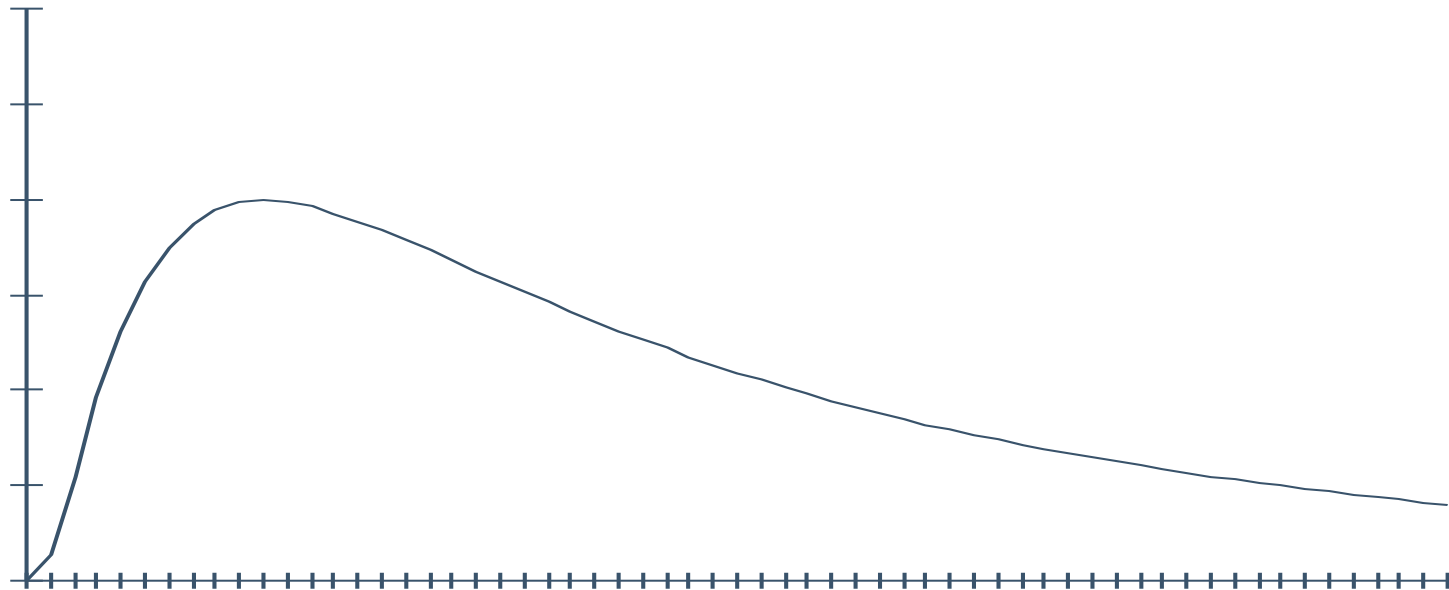
- Muchos valores pequeños
- Unos pocos valores grandes
- Ningún valor negativo

Con los datos de tamaño, la media menos uno o dos desviaciones estándar frecuentemente da valores de tamaño negativo.

La estrategia común para hacer frente a tales distribuciones es tratarlo como una distribución log-normal




# Distribución Log-Normal



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Distribución Log-Normal

□  The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Para normalizar la información del tamaño, hacer lo siguiente:

1. Tomar el logaritmo natural de los datos
2. Determinar la media y desviación estándar de los datos log
3. Calcular los valores promedio, grande, muy grande, pequeño y muy pequeño para los datos log.
4. Tomar el log inverso de los rangos para obtener los valores del tamaño del rango

Este procedimiento generalmente producirá rangos de tamaño útil



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán



# Organizando Datos del Proxy

Esta simple forma para determinar estos rangos de tamaño trabajarán cuando tenga bastante información. En otro caso, puede causar sub estimaciones.

Rangos de estimación comparativa

	<b>VS</b>	<b>S</b>	<b>M</b>	<b>L</b>	<b>VL</b>
<b>Simple</b>	<b>9.33</b>	<b>13.19</b>	<b>17.04</b>	<b>24.52</b>	<b>32.00</b>
<b>Normal</b>	<b>-1.67</b>	<b>7.68</b>	<b>17.04</b>	<b>26.39</b>	<b>35.75</b>
<b>Log-Normal</b>	<b>5.55</b>	<b>9.19</b>	<b>15.22</b>	<b>25.21</b>	<b>41.75</b>



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Estimando con datos limitados -1

Aún luego de usar PSP para muchos proyectos, tendrá que realizar estimaciones con datos limitados cuando

- Se trabaja en un nuevo entorno
- Se usa nuevas herramientas o lenguajes
- Cambio de proceso
- Hacer tareas no familiares

Ya que las estimaciones hechas con información son mas precisas que las conjeturas, use datos siempre que pueda

Use los datos cuidadosamente ya que su uso inapropiado puede conducir a serios errores

# Estimando con datos limitados -2

Dependiendo de la calidad de la información, seleccione uno de los cuatro métodos de estimación PROBE.

## Método

- A **regresión con tamaño del proxy estimado**
- B **regresión con tamaño añadido y modificado del plan**
- C **Método del promedio**
- D **Juicio de la ingeniería**

Para usar el método A o B de regresión, necesita

- una cantidad razonable de información histórica
- información que correlacione
- valores de parámetros  $\beta_0$  y  $\beta_1$  razonable.



# Método A (Regresión): Tamaño del Proxy Estimado

El método A usa la relación entre tamaño del proxy estimado (E) y el real

- Tamaño modificado y añadido
- Tiempo de desarrollo

Los criterios para usar este método son

- Tres o más puntos de información que correlacionen ( $R^2 > 0.5$ )
- Razonables parámetros de regresión
- Conclusión de al menos tres ejercicios con PSP1 o superior



# Método B (Regresión): Tamaño del Plan añadido o modificado

El método B usa la relación entre el tamaño del plan añadido y modificado y

- Tamaño real añadido y modificado
- Tiempo de desarrollo real

Los criterios para usar este método son

- Tres o más puntos de información que correlacionen ( $R^2 > 0.5$ )
- Razonables parámetros de regresión
- Conclusión de al menos tres ejercicios con PSP0.1 o superior



# Método C: Promediando

El método C usa una proporción para ajustar el tamaño o tiempo basado en promedios históricos.

El método del promedio es fácil para usar y requiere solo un punto de datos.

Los promedios asumen que no hay gasto fijado.



# Método D: Juicio de la ingeniería

Use el método D cuando no tenga datos históricos. Use el juicio para

- Proyectar el tamaño añadido y modificado a partir del tamaño de la parte estimada
- Estimar el tiempo de desarrollo

Use el método D cuando no pueda usar los métodos A, B, o C.



# Estimando la precisión

La planificación es una habilidad que debe ser desarrollada:

- PSP ayuda a construir habilidades de planificación
- Aún simples planes son sujetos de error
  - Eventos imprevistos
  - Complicaciones inesperadas
  - Mejores idea de diseño
  - Simplemente errores

La mejor estrategia es planificar en detalle.

- Identificar las tareas reconocidas
- Realizar estimaciones basadas en similares experiencias
- Realizar juicios sobre el resto



# Combinando Estimaciones

Para combinar múltiples estimaciones hechas por un único desarrollador

- Añadir las estimaciones para las partes separadas
- Realizar un cálculo de regresión lineal
- Calcular un conjunto de intervalos de predicción

Múltiples desarrolladores pueden combinar estimaciones hechas independientemente por

- Realizando separadas proyecciones de regresión lineal
- Añadiendo los tamaños y tiempos proyectados
- Añadiendo los cuadrados de rangos individuales y tomar la raíz cuadrada para obtener el intervalo de predicción

# Estimando el Error: Ejemplo

Cuando se estima en partes, el error total será menos que la suma de los errores de las partes.

- Errores tienden a equilibrarse
- No asume ninguna tendencia común

Para un trabajo de 1000-horas con precisión de estimación de  $\pm 50\%$ , el rango de estimación es de 500 a 1500 horas.

Si la estimación es independientemente hecha en 25 partes, cada una con 50% de error, el

- Total sería 1000 horas, como antes
- Rango de estimación sería de 900 a 1100 horas



# Combinando Individuales Errores

Para combinar estimaciones hechas independientemente

- Añadir los valores estimados
- Combinar las varianzas (cuadradas) de los errores

Con 25 estimaciones para un trabajo de 1000-horas

- Cada estimación promedia de 40 horas
- La desviación estándar es 50%, o 20 horas.
- La varianza para cada estimación es 400 horas.
- Las varianzas a añadir hasta 10,000 horas.
- La desviación estándar combinada es la raíz cuadrada de la suma de las varianzas, o 100 horas.
- El rango de estimación es 900 a 1100 horas.



# Ejercicio de Clase -1

Empezar con tres estimaciones.

- A = 45 horas, + o - 10
- B = 18 horas, + o - 5
- C = 85 horas, + o - 25

Cuál es la estimación combinada?



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Ejercicio de Clase -2

Empezar con tres estimaciones.

- A = 45 horas, + o - 10
- B = 18 horas, + o - 5
- C = 85 horas, + o - 25

Cuál es la estimación combinada?

- total =  $45 + 18 + 85 = 148$  horas

Cuál es el rango de la estimación combinada?



# Ejercicio de Clase -3

Empezar con tres estimaciones.

- A = 45 horas, + o - 10
- B = 18 horas, + o - 5
- C = 85 horas, + o - 25

Cuál es la estimación combinada?

- total =  $45 + 18 + 85 = 148$  hours

Cuál es el rango de la estimación combinada?

- varianza =  $100 + 25 + 625 = 750$
- rango = raíz cuadrada de la varianza = 27.4 horas

Cuál es el UPI y LPI combinado?



# Ejercicio de Clase -4

Empezar con tres estimaciones.

- A = 45 horas, + o - 10
- B = 18 horas, + o - 5
- C = 85 horas, + o - 25

Cuál es la estimación combinada?

- total = 45 + 18 + 85 = 148 horas

Cuál es el rango de la estimación combinada?

- varianza = 100 + 25 + 625 = 750
- rango = raíz cuadrada de la varianza = 27.4 horas

Cuál es el UPI y LPI combinado?

- UPI = 148 + 27.4 = 175.4 horas
- LPI = 148 - 27.4 = 120.6 horas

# Usando Múltiples Proxies

Con información del tamaño/hora para varios proxies

- Estimar cada uno como antes
- Combinar el total de estimaciones e intervalos de predicción como se describió

Use regresión múltiple si

- Hay correlación entre el tiempo de desarrollo y cada proxy
- Los proxies no tienen por separado información del tamaño y hora





# Estimando Consideraciones

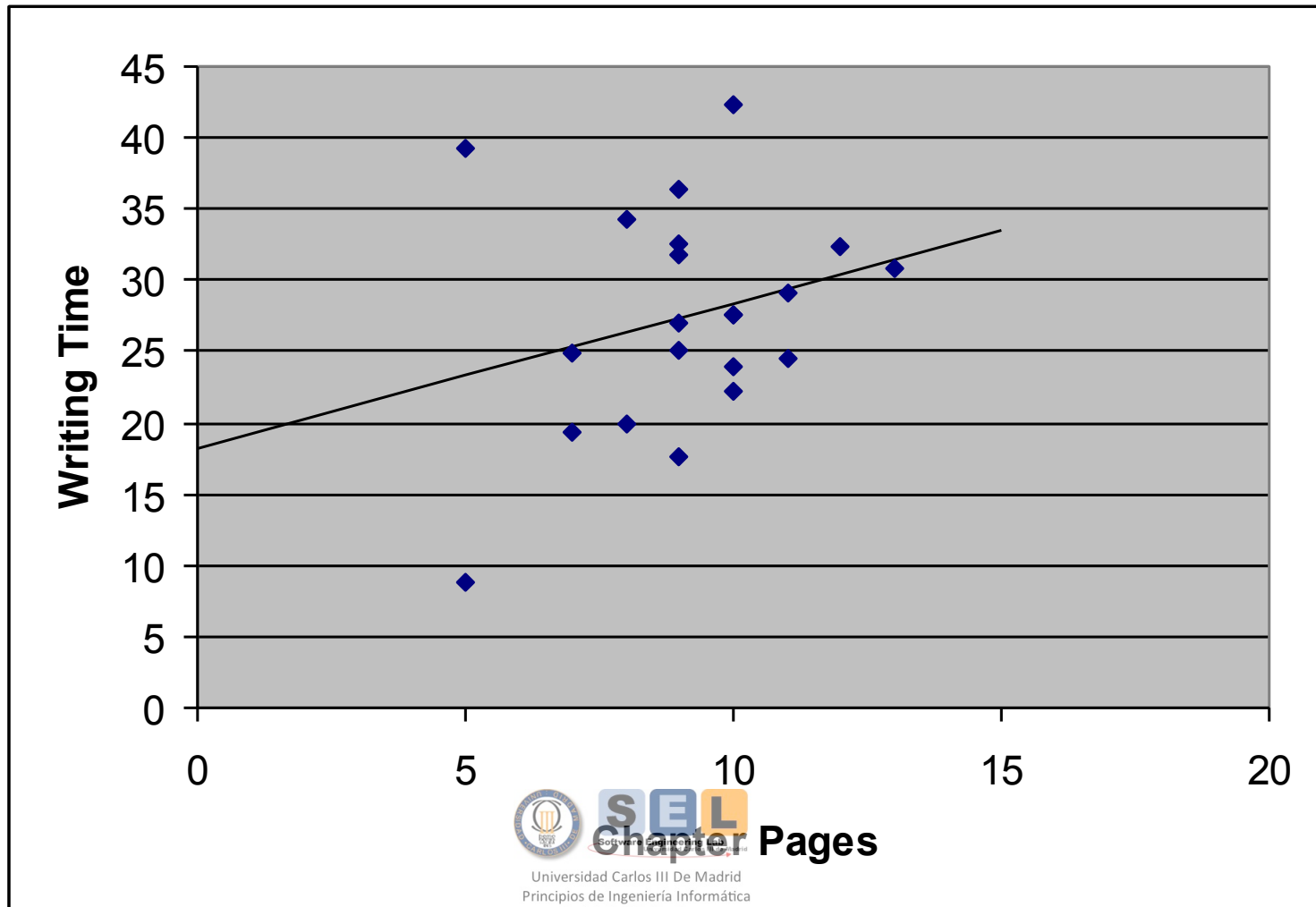
Mientras el método PROBE puede proveer estimaciones precisas, el uso impreciso de datos puede conducir a serios errores.

Un punto extremo puede dar una alta correlación aún cuando los datos restantes estén pobremente correlacionadas.

Similarmente, puntos extremos pueden conducir a erróneos valores  $\beta_0$  y  $\beta_1$ , aún con una alta correlación.



# Correlación con datos agrupados

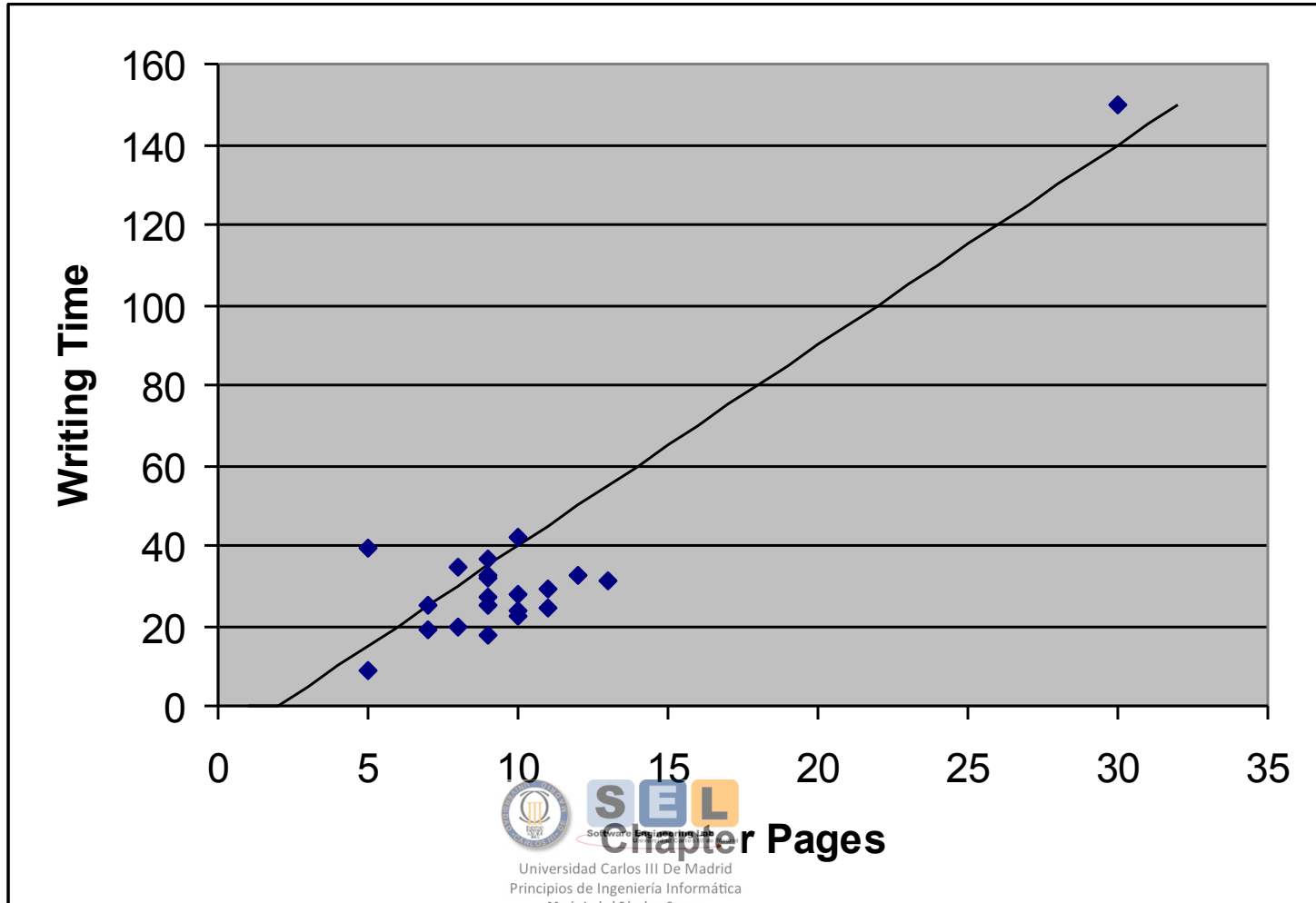


$r = 0.26$



Universidad Carlos III De Madrid  
Principios de Ingeniería Informática  
María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Correlación con un Punto Extremo



$r = 0.91$



Chapter Pages

Universidad Carlos III De Madrid  
Principios de Ingeniería Informática

María Isabel Sánchez Segura  
José Arturo Mora-Soto  
Juan Carlos Alonso Durán

# Conclusiones sobre datos engañosos

Con sólo un punto movido hacia un valor extremo, la correlación para los mismos datos se incrementó de 0.26 a 0.91.

Similarmente, los valores  $\beta_0$  y  $\beta_1$  cambiaron de

- 18.23 a -17,76 para  $\beta_0$
- 1.02 a 5.08 para  $\beta_1$

Con un promedio de productividad de 3.02 y 3.31 horas por página, ambos valores  $\beta_1$  son engañosos.

Con un punto extremo, probablemente no deberías usar regresión



# Mensajes a Recordar

El método PROBE provee una forma estructurada de realizar estimaciones de tamaño y tiempo de desarrollo.

- Usa tu información personal de desarrollo
- Provee un rango conveniente con el cual el tamaño del programa real y el tiempo de desarrollo probablemente encajen

Con un método de estimación estadísticamente conveniente como PROBE, puede calcular el error de estimación probable.

