Medida de Esforço de Desenvolvimento de Software

Unidade V – Estimativas com Base Estatística

Luiz Leão – luizleao@gmail.com
http://www.luizleao.com

Conteúdo Programático

- Utilizar registros estatísticos para gerar estimativas
- Definir o nível de estimativa (linear, exponencial,...)
  - Aprender a fazer interpolação linear ou não
  - Definir os riscos nas estimativas
  - Definir nível de erro

Objetivos

- Usar os registros estatísticos de projetos anteriores para criarmos uma base de estimativas, baseadas em ponto função, que sejam **confiáveis** o suficiente para serem usadas comercialmente.

- Definir o nível de erro de nossas estimativas

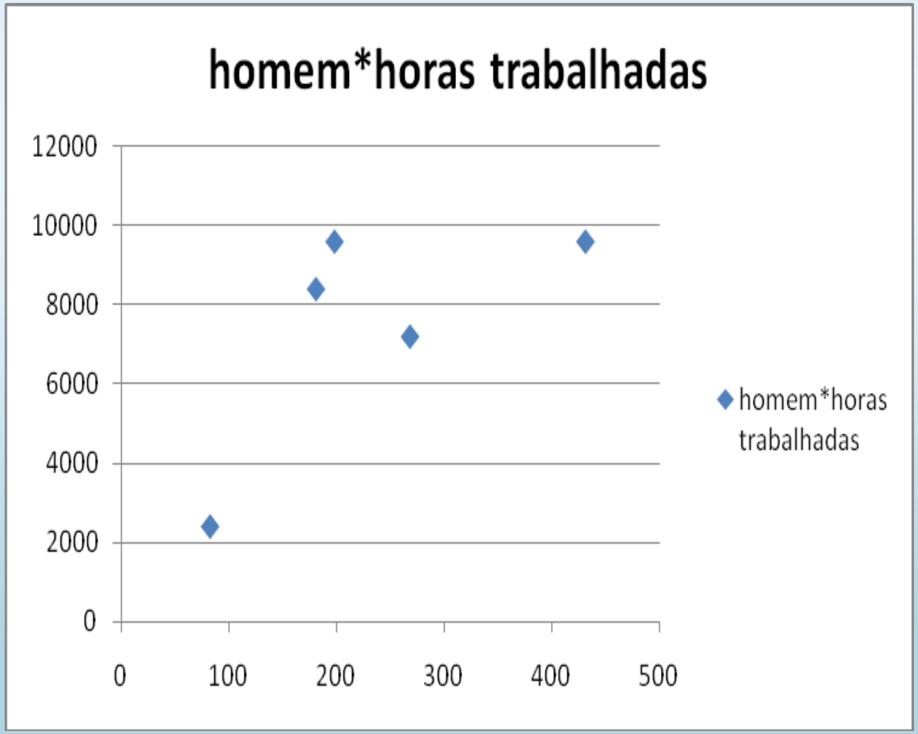
Pontos a serem observados

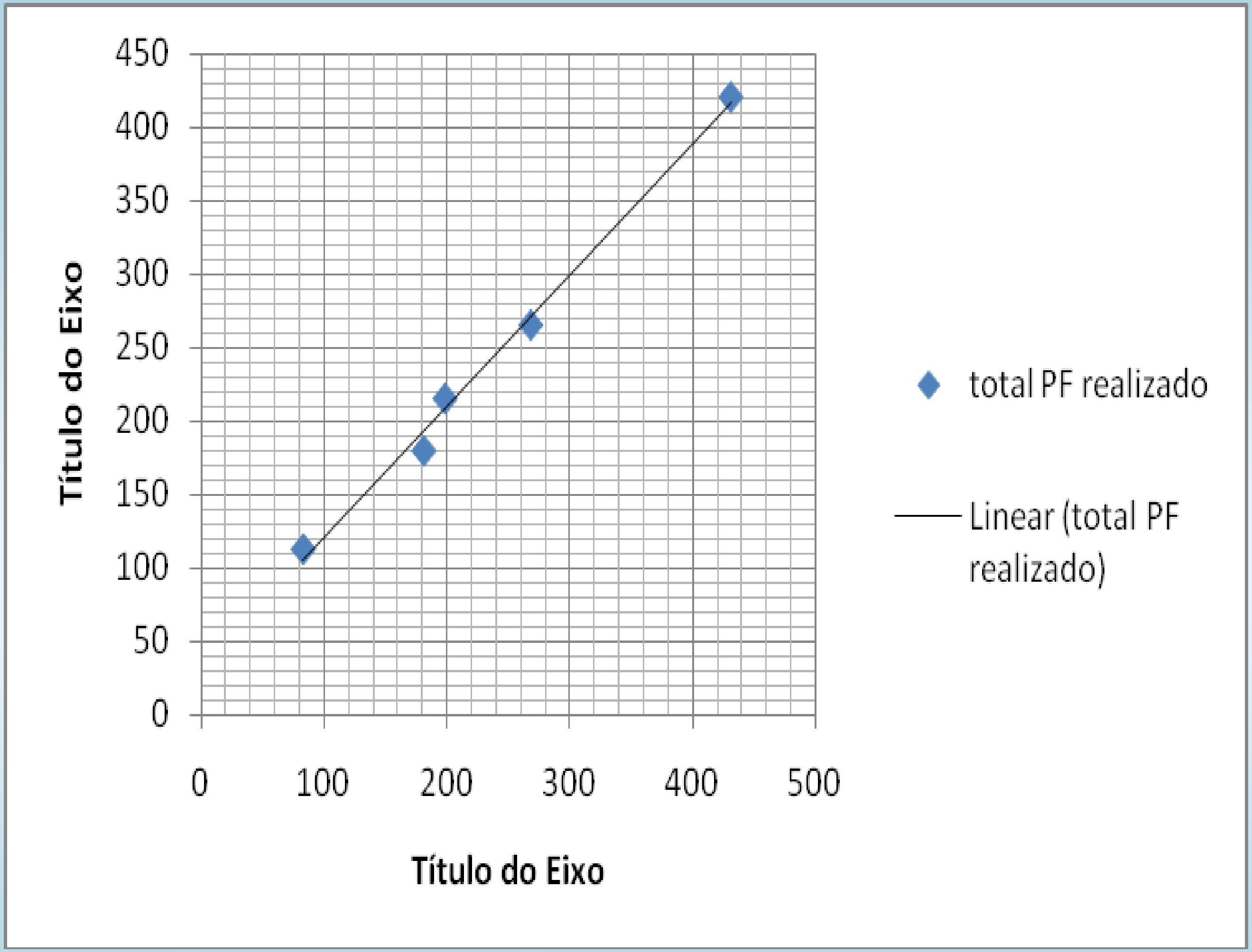
- A importância dos registros feitos durante a execução do projeto e após o fim do projeto.
- Mostrar que estes dados são **pontuais** e **não contínuos**.
- Que existem métodos de interpolação e definição de erro assumido.
- Mostrar as divergências dos dados para estimativas fora do limite dos dados, dependendo da função assumida
- Mostrar como se faz uma interpolação linear, e de outra função
- Determinação do nível de erro assumido

- Ter um conjunto de dados que permita se ter gerencia do processo e do produto é importante, mas precisamos aprender a trabalhar com estes dados.

- Suponha uma empresa com um tempo de vida curto e que entre outros dados registrou-se os apresentados na tabela abaixo:

Projeto	Total PF Estimado	Total PF Realizado	Homem * horas trabalhadas	Prazo (dias uteis)
Modulo de Ponto	82,45	112,2	2400	40
Sistema Estoques	180,90	179,34	8400	70
Gestão de Vendas	198,20	215	9600	160
Controle de Trafego	268,30	265,2	7200	180
Sistema Academico	431,45	420,78	9600	160

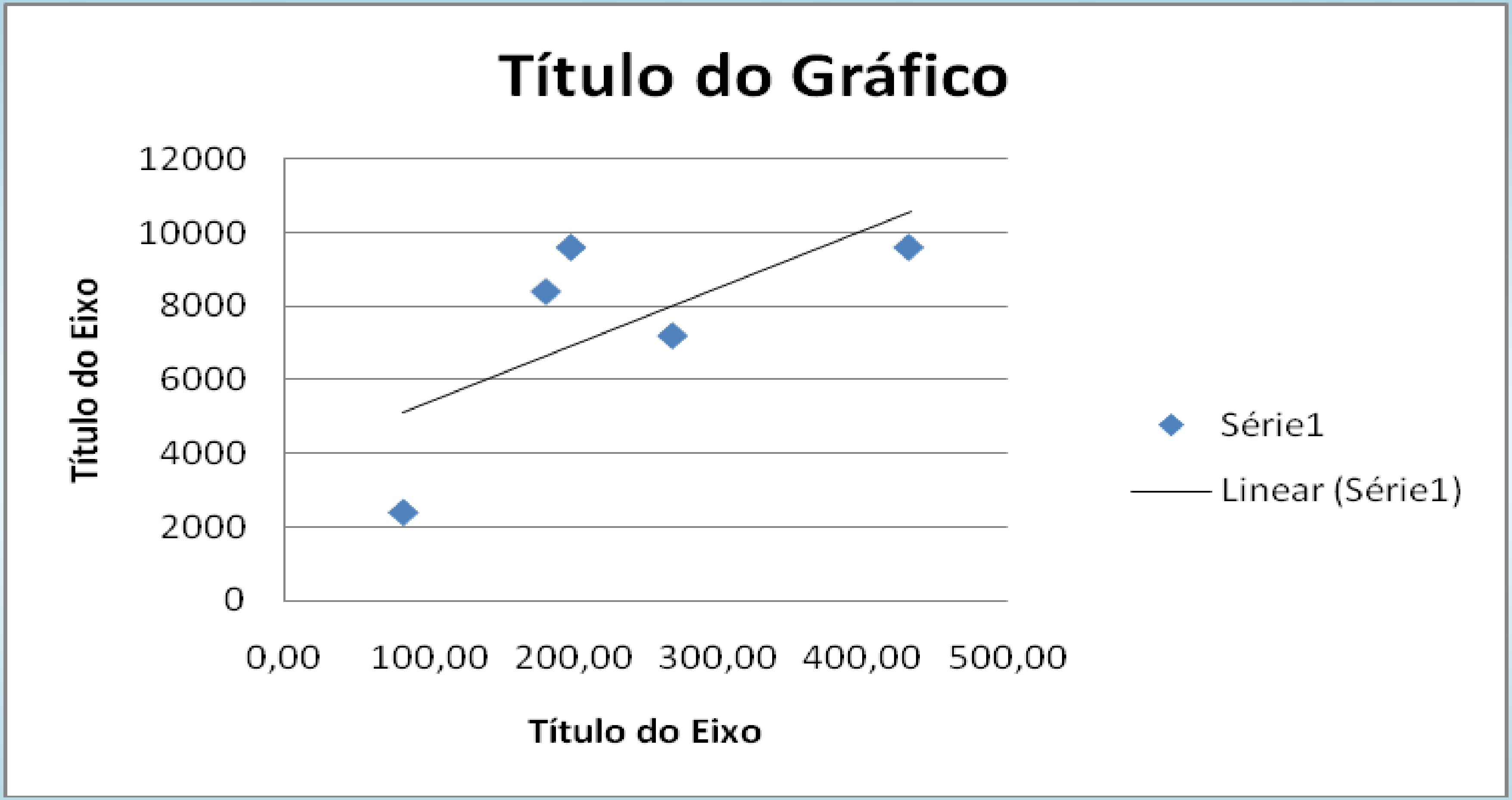




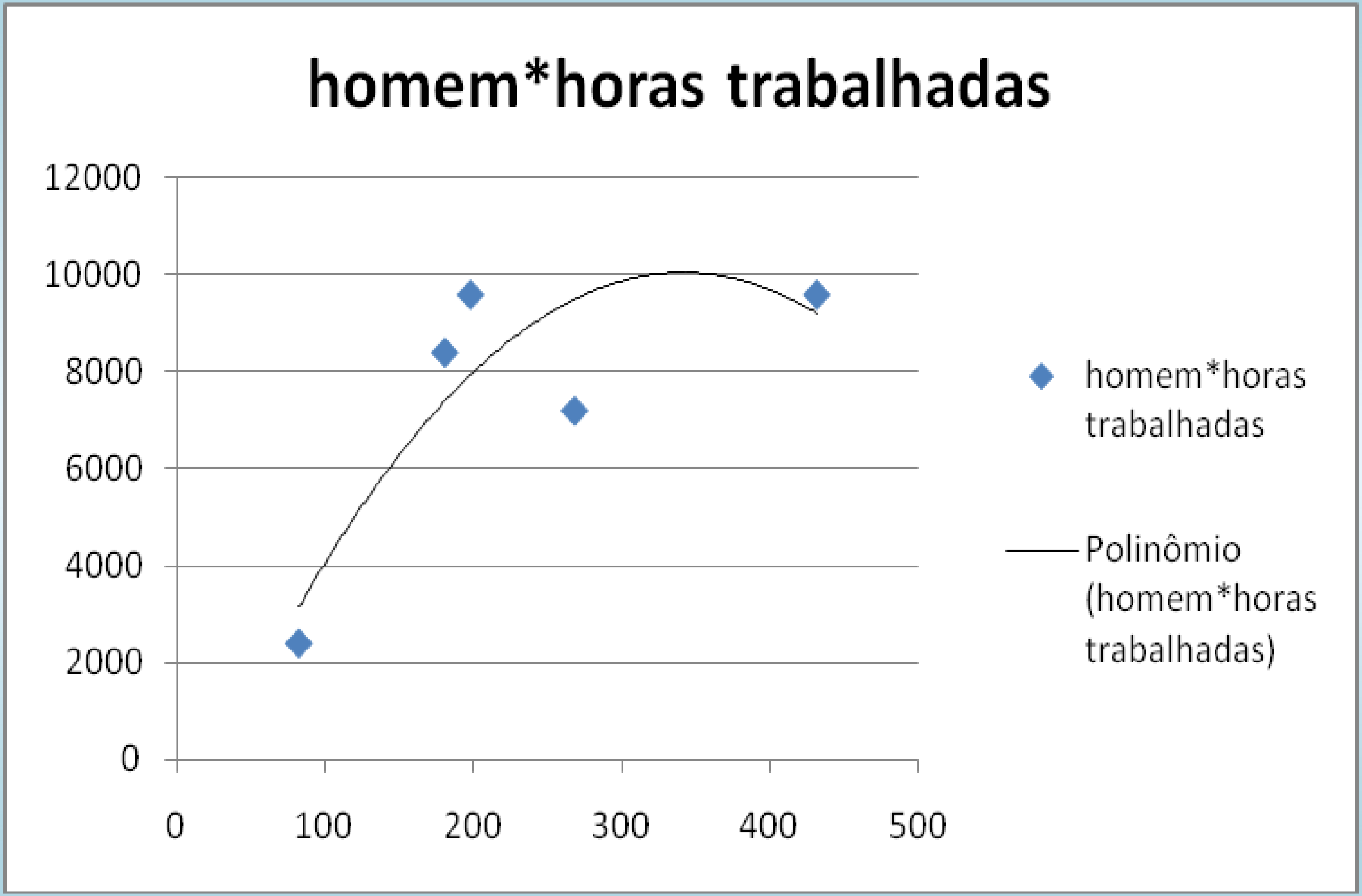
A curva de acompanhamento de estimativas em PF tendencia uma interpolação linear de 45 graus o que mostra que a estimativa está muito próxima da realizada, com pequenas diferenças. (PF estimado X PF realizado)

- Os pontos (ou falta deles) podem nos levar a uma serie de suposições.

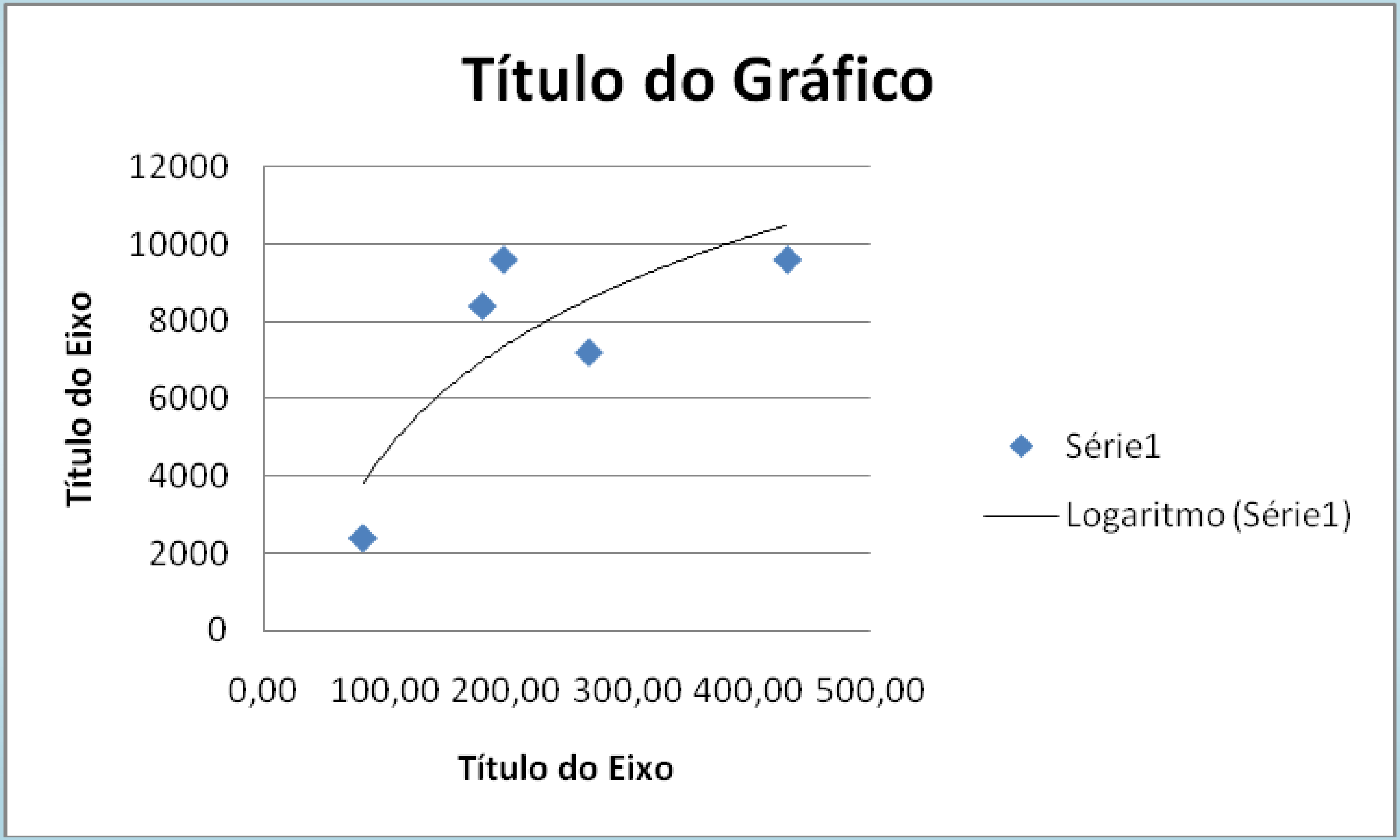
- Ao assumir que o comportamento é próximo de linear, devemos saber que esta simplificação poderá nos custar caro, pois dificilmente temos este tipo de comportamento.

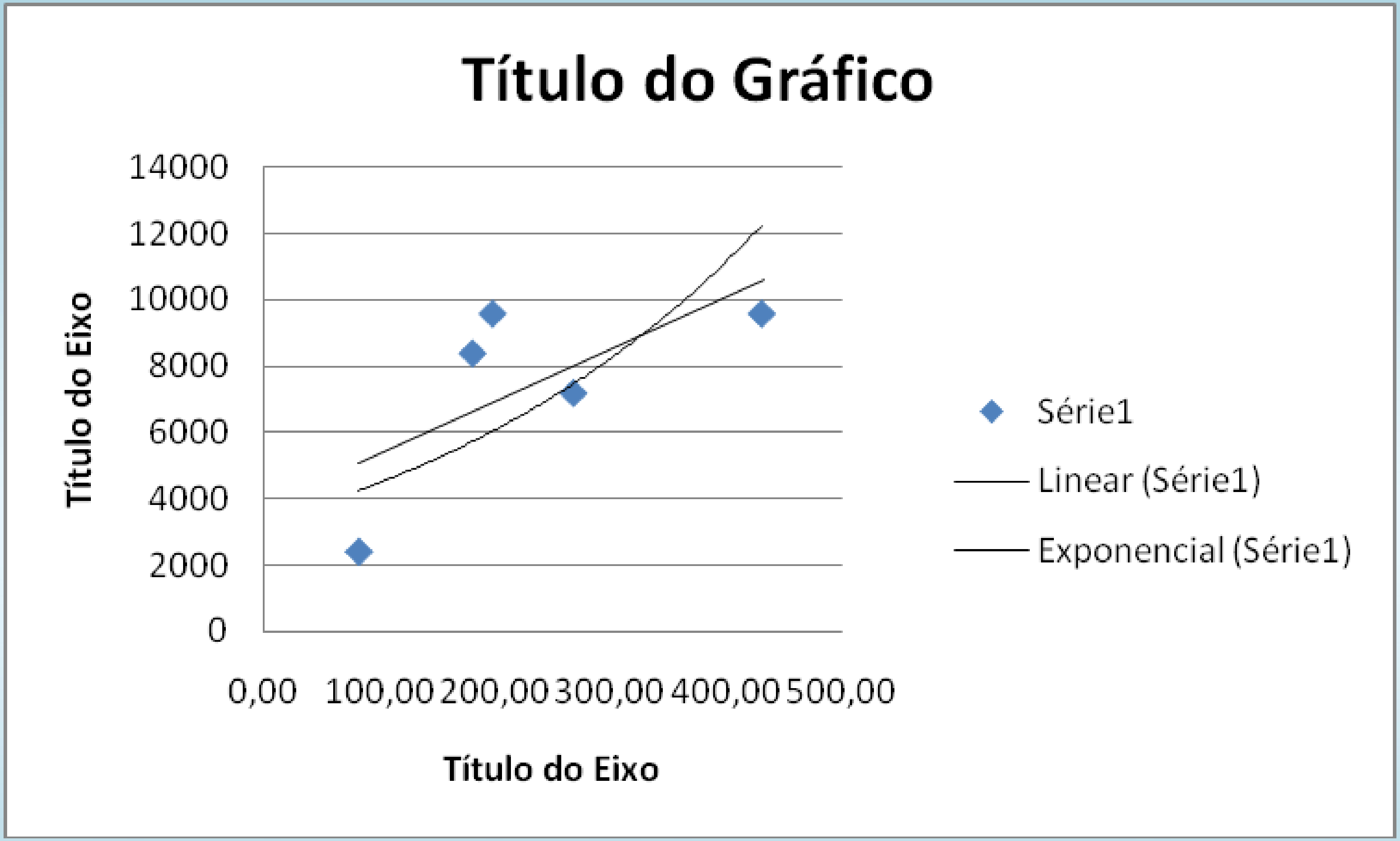


Desejamos corre menos risco podemos considerar que os pontos tem a tendência de uma curva de segundo grau:


homem*horas trabalhadas
12000
10000
8000
6000
4000
2000
0
0
100
200
300
400
500
homem*horas trabalhadas
Polinômio (homem*horas trabalhadas)

Curva Logarítmica





Observe que para um número menor de PF usar a tendência linear nos leva a estimar valores de forma mais alta que da forma exponencial, invertendo-se a tendência a partir do momento que os projetos se tornam significativos

Problema:

- Considere que temos um levantamento de um produto com 230 PF e devemos usar a tabela para definir o esforço necessário.

- Na tabela não temos como obter a informação diretamente, pois nenhum projeto até o momento tem 230 PF.

- Precisamos deduzir um valor a partir dos dados da tabela: chama-se de **interpolação**

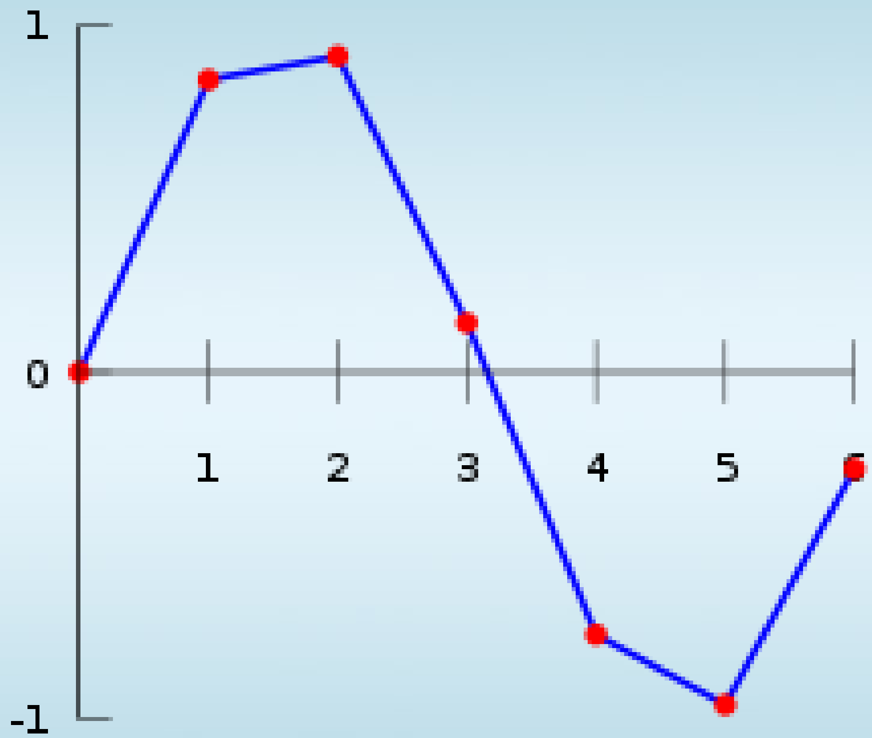
Interpolação

- É o método que permite construir um novo conjunto de dados a partir de um conjunto discreto de dados pontuais previamente conhecidos.

- Em engenharia é comum dispor-se de dados pontuais obtidos a partir de uma amostragem ou de um experimento

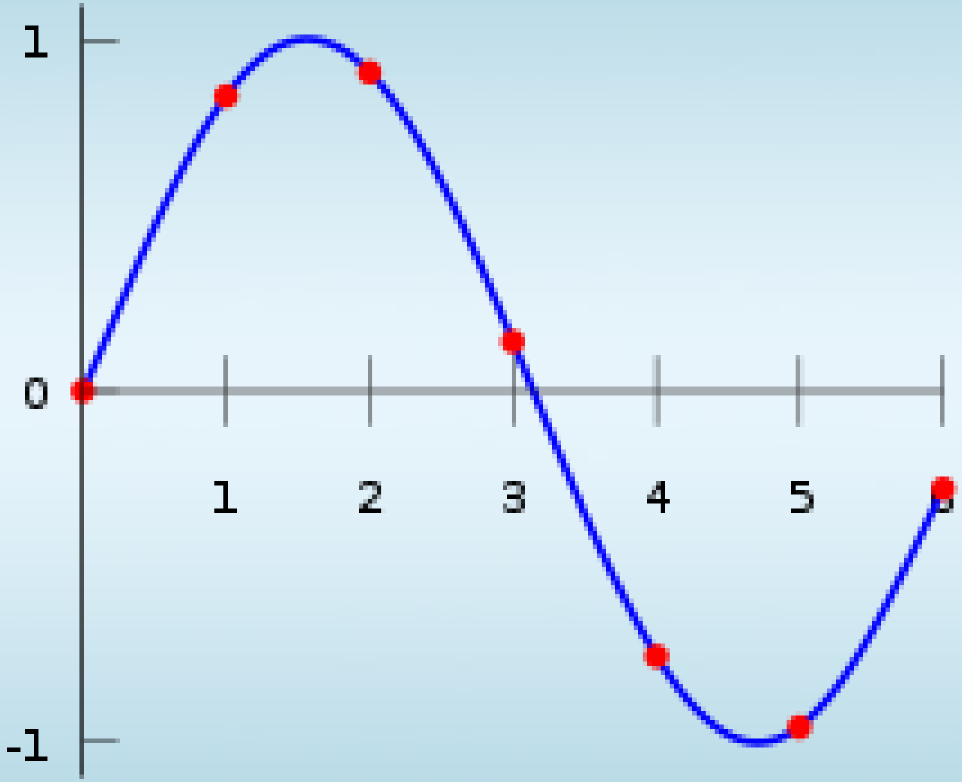
- Tal conjunto de dados pontuais (também denominado *conjunto degenerado*) não possui continuidade, e isto muitas vezes torna demasiado irreal a representação teórica do fenomeno real e observado.

Interpolação



O gráfico acima, mostra que ao se conhecer poucos pontos temos uma série de funções que podem se ajustar:

Interpolação



Exemplo de interpolação polinomial de grau superior a 1.

Interpolação

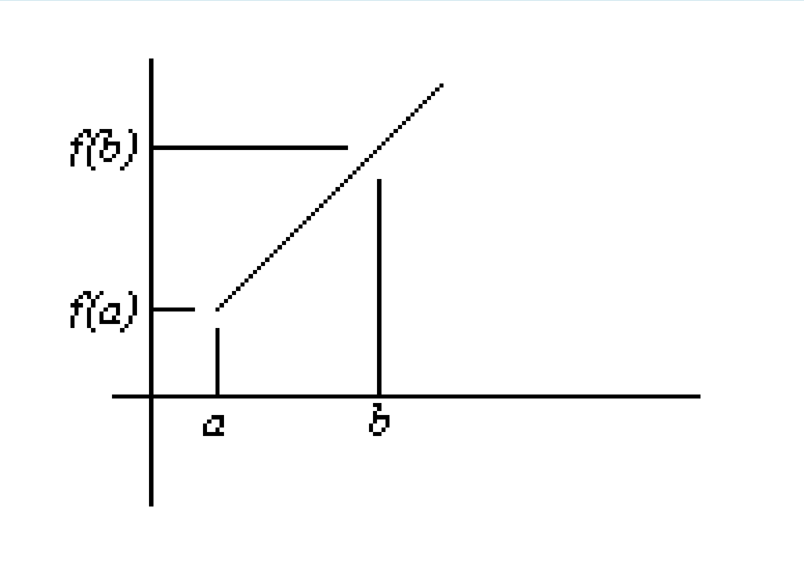
- Consiste basicamente em encontrar uma função que seja a expressão lógica de determinados pontos de uma função desconhecida, ou seja, conhecendo-se (x_1,y_1), (x_2,y_2).....(x_n,y_n) de uma função desconhecida poderemos calcular o valor numérico intermediário da função num ponto não tabelado com certo grau de erro.

Pontos de Amarração:

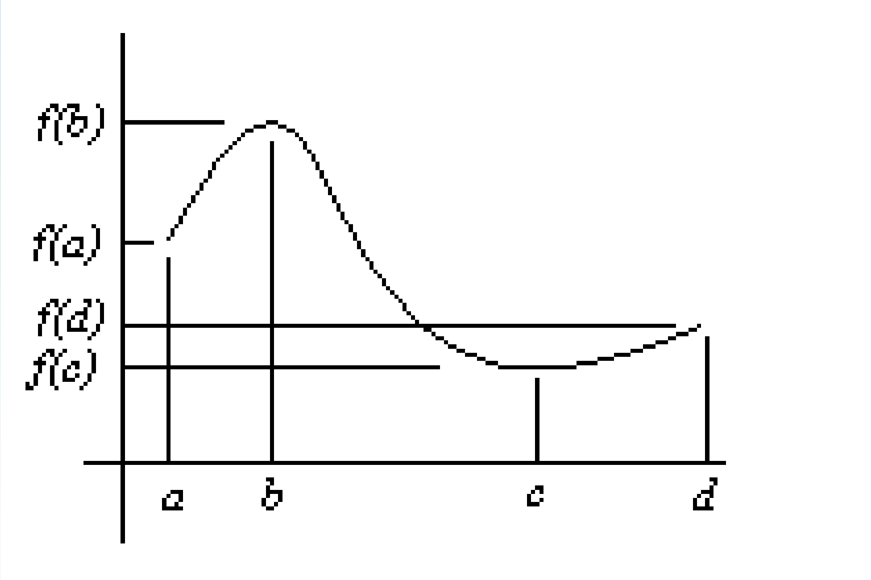
- Os pontos de amarração são os pontos em que a função substituta conterá da função tabela, no qual será construída uma função para um respectivo intervalo.

- Para se fazer escolha de uma infinidade de funções que venham assumir determinados pontos faz-se na verdade a escolha de uma função onde se possa trabalhar com simplicidade, deste modo a função mais simples um polinômio.

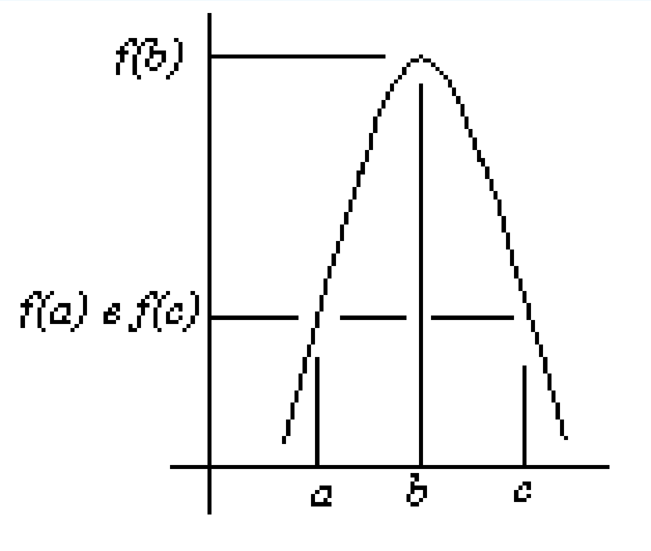
Obs: Nos pontos de amarração f(x) é igual a g(x), g(x) pode ser chamada função substituta.



2 pontos (polinômio de 1º grau)



4 pontos (polinômio de 3º grau)

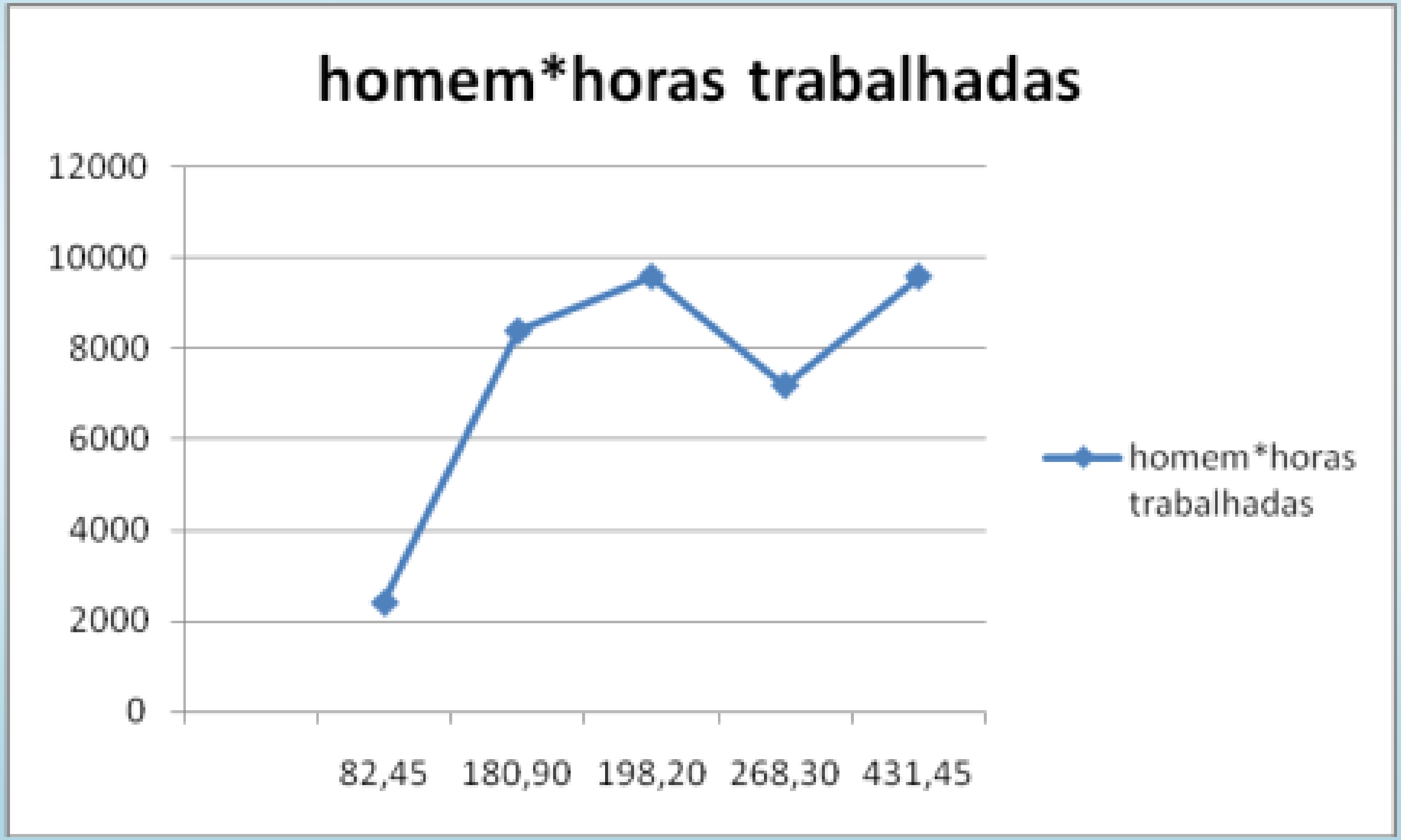


3 pontos (polinômio de 2º grau)

- Para (n+1) pontos, existe um e somente um polinômio de grau não superior a n.

Interpolação Linear

Estamos supondo o comportamento linear entre o dois pares de pontos, E assim, para a tabela inicial tem se o seguinte gráfico



Consultando o gráfico, para o nosso exemplo de 230PF, verificamos que este valor fica entre os pontos conhecidos:

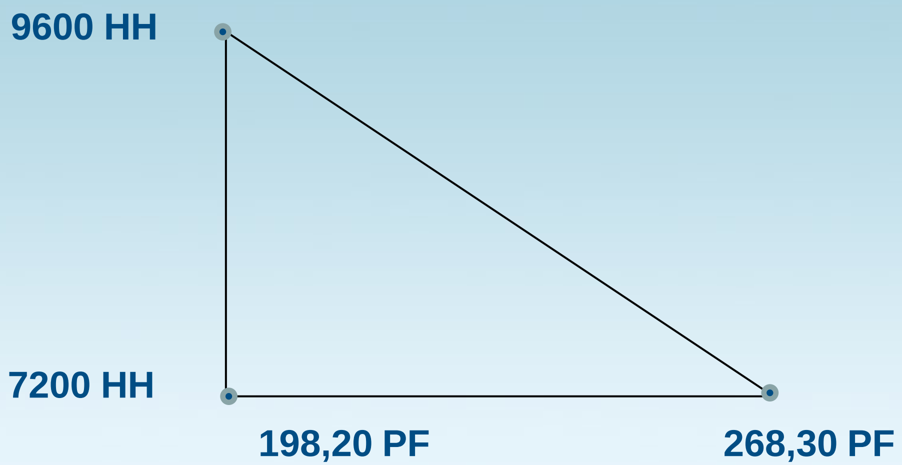
Projeto	Total PF Estimado	Total PF Realizado	Homem*Horas trabalhadas	Prazo (dias uteis)
Modulo de Ponto	82,45	112,2	2400	40
Sistema Estoque	180,90	179,34	8400	70
Gestão de vendas	198,20	215	9600	160
Controle de Tráfego	268,30	265,2	7200	180
Sistema Acadêmico	431,45	420,78	9600	160

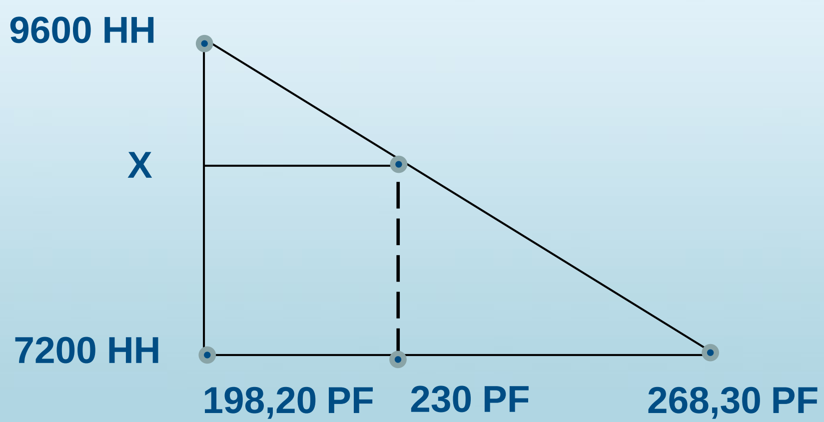
Do projeto gestão de venda:

com 198,20 PF com 9600 homem*hora

E o projeto controle de trafego:

com 268,30 PF e 7200 homem*hora


9600 HH
7200 HH
198,20 PF
268,30 PF


9600 HH
X
7200 HH
198,20 PF
230 PF
268,30 PF

- Considerando os 230 PF e verificando que (268,30 – 230 = 38,30) Por semelhança de triângulos podemos escrever:

$$\frac{X}{2400} = \frac{38{,}30}{70{,}10}$$

- Assim temos X * 70,10 = 2400 * 38,30 donde

  X = 1311,27

- Assim por interpolação linear temos um acréscimo de 1311,27 acima de 7200 nos mostrando um total de: 7200 + 1311,27 = 8511,27 homem*hora

Risco Assumido

- Ao assumir um valor estamos assumindo um risco, que poderá diminuir assumindo-se outro tipo de curva.

- Mas as vezes temos que determinar um valor de estimativa além dos limites do domínio de nossa tabela

- Neste caso podemos estipular a equação da função e usar o novo valor como entrada da função.

- Sempre lembrando que estamos assumindo um risco que pode ser diminuído a medida que vamos aumentando nossa base de dados

Risco Assumido

- Suponha para fins de exemplo que desejamos achar o esforço em Homem*hora para um novo projeto de 600 PF.
- Observando a tabela este valor está acima do maior valor existente.
- Neste caso precisamos determinar a função que melhor atende.
- Vamos supor que vamos trabalhar com uma reta que passa pelos pontos:

268,30 PF, 7200 H*H

431,45 PF 9600 H*H

Identificar a Função

- A equação geral da reta é: y = a*x + b
- Aplicando os pontos temos:

$$7200 = a * 268,30 + b$$

$$9600 = a * 431,45 + b$$

- Com os dois pontos podemos manipular algebricamente para achar a:

$$7200 = a * 268,30 + 9600 – a *431,45$$

$$7200 – 9600 = a * (268,30 - 431,45)$$

$$a = -2400/-163,15 \quad \text{donde} \quad a= 14,71$$

- Da outra equação tiramos o valor de b

$$b = 9600 – 14,71 * 431,45$$

$$\text{donde} \quad b = 3253,37$$

- Assim temos a equação da nossa reta:

$$\mathbf{y = 14,71 * a + 3253,37}$$

Identificar a Função

- Com a equação da nossa reta:

y = 14,71 * a + 3253,37

- Usando a função para acharmos o esforço para 600 pontos temos:

Y = 14,71*600 + 3253,37

- O ESFORÇO ESTIMADO SERÁ:

12079,31 HOMEM * HORA

Erros das Estimativas Interpoladas

- Devemos lembrar que estamos introduzindo um erro devido a função escolhida.

- A partir de uma tabela, em geral, não vamos conseguir obter uma função que modele o fenômeno de maneira exata mas somente uma de uma forma aproximada.

- Surge outro problema:
  - Como escolher o tipo de função que aproxima o fenômeno?
  - Que tipo de interpolação devemos fazer?

Erros das Estimativas Interpoladas

- O fato da aproximação ser razoável (boa) ou não para modelar o fenômeno estudado dependerá da resposta a pergunta.

- Por outro lado, a pista para esta resposta deve estar contida na tabela.

- Os pontos listados na tabela podem mostrar uma tendência que devemos respeitar se desejamos que a função de interpolação represente de forma razoável o fenômeno estudado.

Erros das Estimativas Interpoladas

- Determinados pontos de uma função desconhecida, ou seja, conhecendo-se (x_1 , y_1), (x_2 , y_2).....(x_n , y_n) de uma função desconhecida poderemos calcular o valor numérico intermediário da função num ponto não tabelado com certo grau de erro.

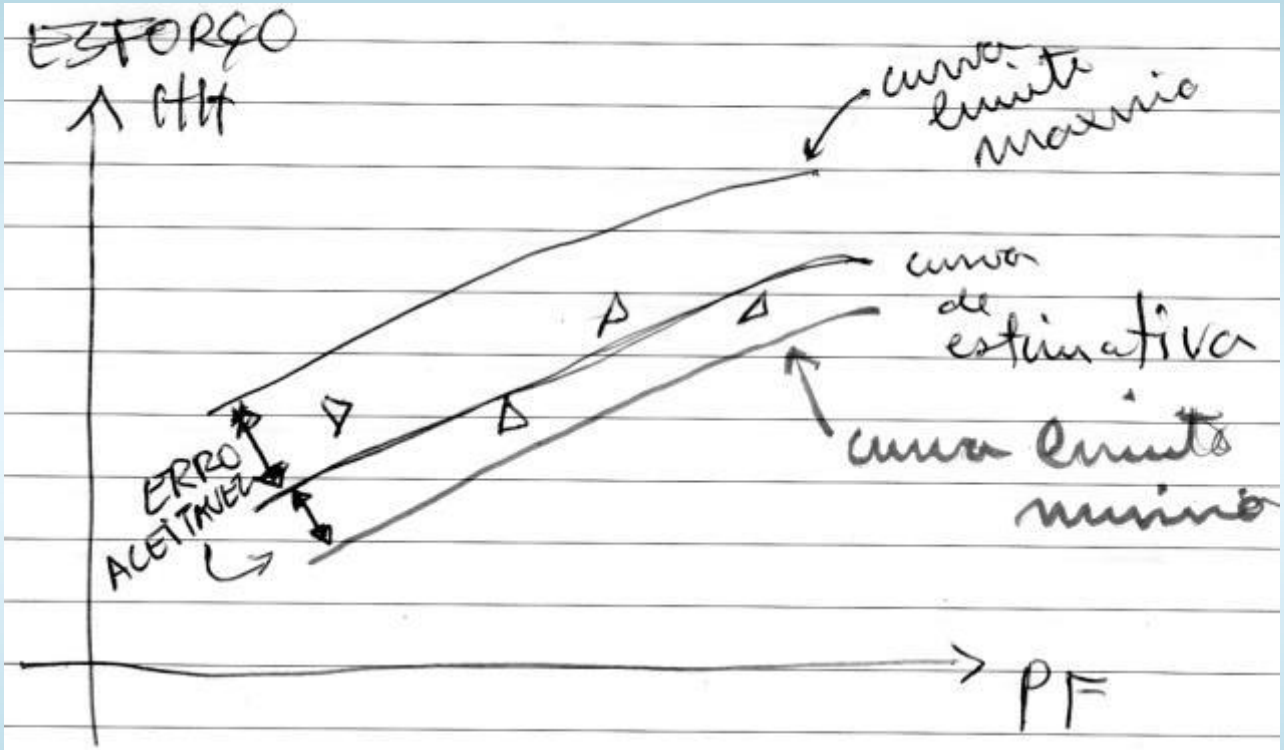
- Assim o erro em um ponto qualquer é o módulo de $f(x) – G(x)$.

Erros das Estimativas Interpoladas

- Existem vários processos matemáticos para tratar a interpolação e o seu erro.

- De uma maneira simplista podemos definir uma curva, no nosso gráfico, que seja limite para considerarmos o erro.

Risco Assumido



- A medida que aumentamos nossa base estatística podemos ir aproximando as duas curvas limites (Mínimo e Máximo) pra perto da estimativa diminuindo a faixa de erro, como mostrado na figura

Risco Assumido

- As interpolações de dados devem ser feitos com muita técnica, ao se estimar uma valor, estamos assumindo um risco. Pode-se aprofundar a teoria matemática. Existe teoria, com métodos bem estabelecidos para se trabalhar com segurança.

- O nível de erro não deve ser desprezado, pois pode levar a altos prejuízos, quando se tem uma base de dados com pouca informação