

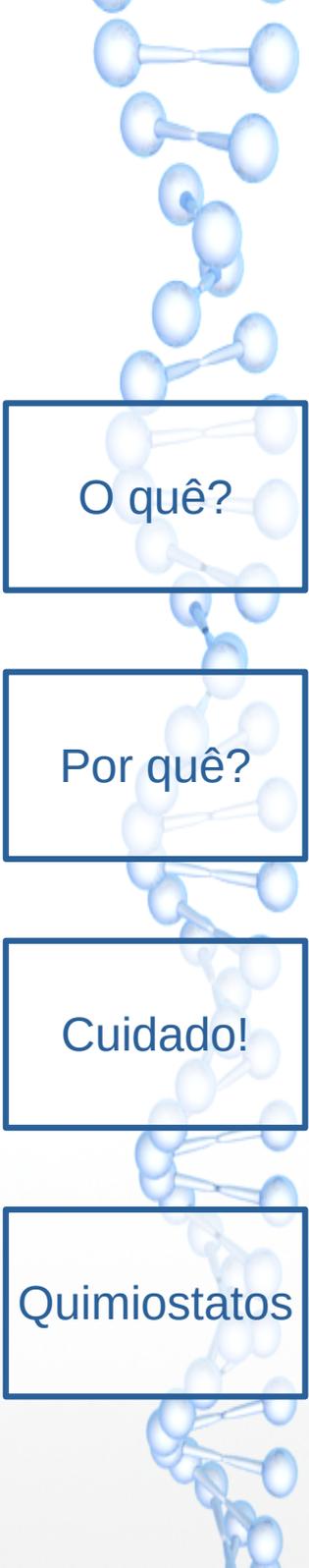
# Bioprocessos microbianos 4

## Reconciliação de dados

Maio/2020



# Sumário



O quê?

O que é reconciliação de dados?

Por quê?

Por que é importante?

Cuidado!

Limiar entre reconciliação e fabricação

Quimiostatos

Exemplo quimiostatos

# O que é reconciliação de dados?

O quê?

A partir de um conjunto de dados experimentais inconsistente com alguma teoria, gerar um conjunto de estimativas para esses mesmos dados que seja consistente com essa mesma teoria evitando enviesar os dados mais ainda

Por quê?

Cuidado!

Quimiostatos

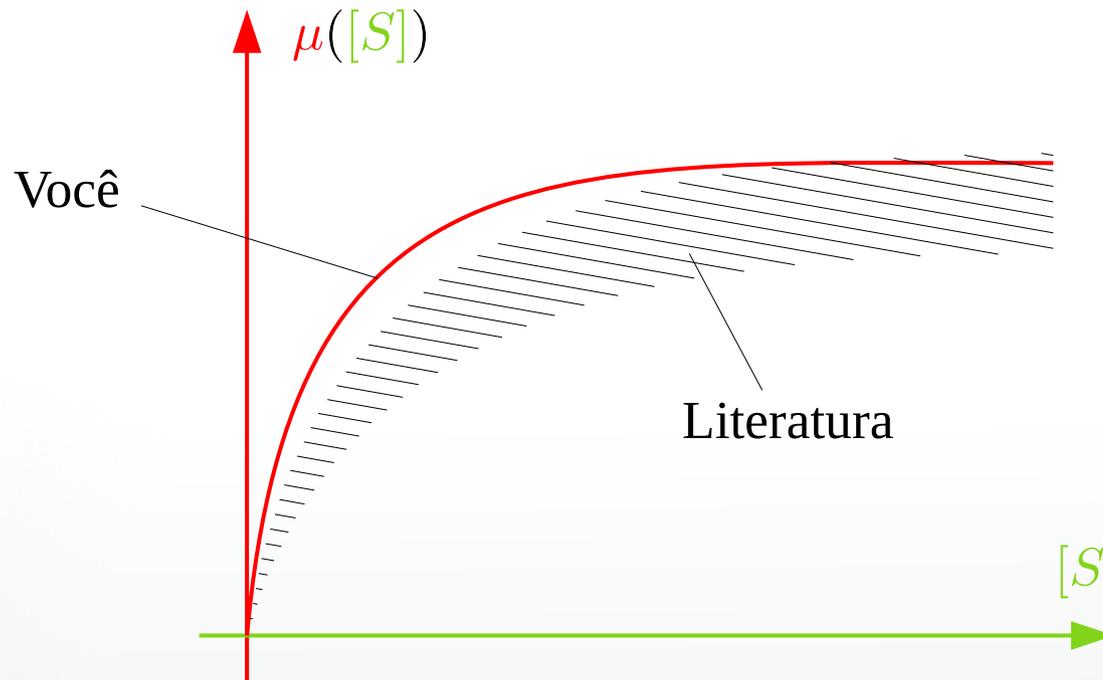
# O que é reconciliação de dados?

## Exemplo motivacional

O quê?

Pela literatura, o seu microorganismo deve seguir o modelo de Monod. Mas quando você fez um ensaio em batelada para verificar os valores dos parâmetros  $\mu_{\max}$  e  $K_S$ , você obteve o seguinte gráfico:

Por quê?



Cuidado!

Quimiostatos

# O que é reconciliação de dados?

## Exemplo motivacional

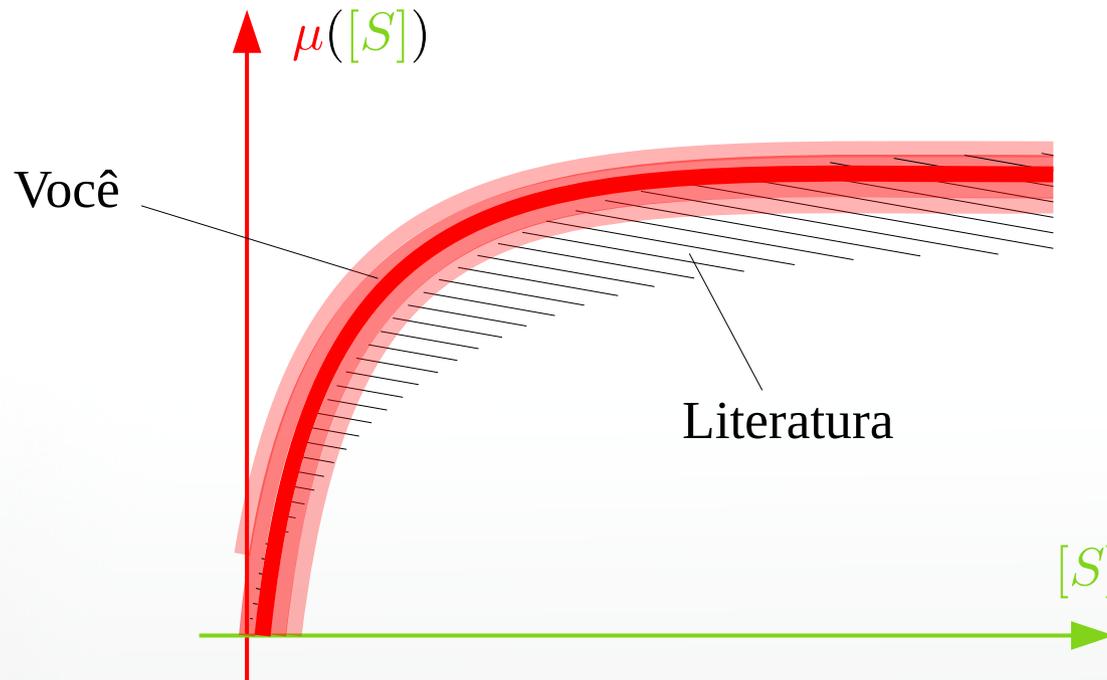
O quê?

Porém, ao fazer uma simulação de monte-carlo variando os valores das suas medidas (concentrações, volumes, tempo, etc) dentro das incertezas delas, você obteve a seguinte área de probabilidades:

Por quê?

Cuidado!

Quimiostatos



# O que é reconciliação de dados?

## Exemplo motivacional

O quê?

Então, você recalculou a curva após selecionar uma estimativa para cada medição de dentro da sua área de possibilidades, obtendo agora uma nova curva que é consistente com aquilo teoricamente esperado:

Por quê?

Você após  
reconciliar  
os dados

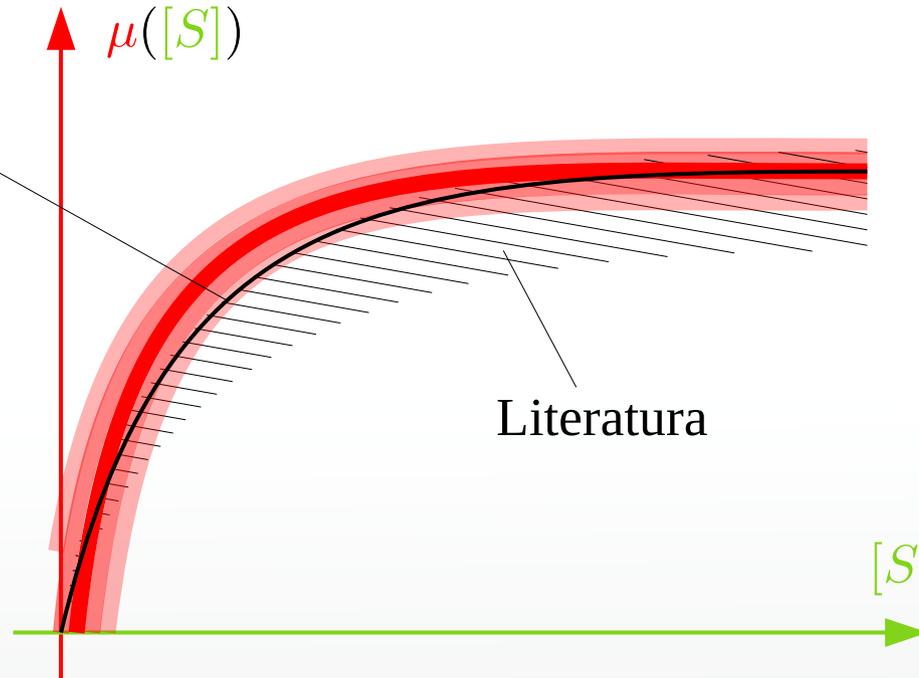
$\mu([S])$

Cuidado!

Literatura

Quimiostatos

$[S]$



# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{array}{rclcl} m_B & = & [B] & \cdot & V \\ 2.5 & = & 0.5 & \cdot & 5 \\ ??? \pm ??? & = & 0.50 \pm 0.05 & \cdot & 5.00 \pm 0.05 \end{array}$$

Cuidado!

Quimiostatos

# O que é reconciliação de dados?

Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{array}{rclcl} m_B & = & [B] & \cdot & V \\ 2.5 & = & 0.5 & \cdot & 5 \\ ??? \pm ??? & = & 0.50 \pm 0.05 & \cdot & 5.00 \pm 0.05 \end{array}$$

Cuidado!

**Método tradicional:** médias e desvios-padrões das variáveis medidas → fórmulas de propagação

Quimiostatos

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{aligned} m_B &= [B] \cdot V \\ 2.5 &= 0.5 \cdot 5 \\ ??? \pm ??? &= 0.50 \pm 0.05 \cdot 5.00 \pm 0.05 \end{aligned}$$

Cuidado!

Quimiostatos

$$\sigma_{m_B} = m_B \cdot \sqrt{\left(\frac{\sigma_{[B]}}{[B]}\right)^2 + \left(\frac{\sigma_V}{V}\right)^2}$$

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{array}{rclcl} m_B & = & [B] & \cdot & V \\ 2.5 & = & 0.5 & \cdot & 5 \\ ??? \pm ??? & = & 0.50 \pm 0.05 & \cdot & 5.00 \pm 0.05 \\ 2.50 \pm 0.25 & = & 0.50 \pm 0.05 & \cdot & 5.00 \pm 0.05 \end{array}$$

Cuidado!

Quimiostatos

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{array}{rcl} m_B & = & [B] \cdot V \\ 2.5 & = & 0.5 \cdot 5 \\ ??? \pm ??? & = & 0.50 \pm 0.05 \cdot 5.00 \pm 0.05 \end{array}$$

Calcular 1000 valores e pegar do primeiro ao terceiro quartis

Gerar 1000 valores aleatórios entre 0.45 e 0.55

Gerar 1000 valores aleatórios entre 4.95 e 5.05

Quimiostatos

# O que é reconciliação de dados?

Incertezas via simulação de Monte-Carlo

O quê?

Método numérico simples para estimar uma distribuição de probabilidades usando um gerador de números aleatórios como o do Excel.

Por quê?

Exemplo:

$$\begin{array}{rcl} m_B & = & [B] \cdot V \\ 2.5 & = & 0.5 \cdot 5 \\ 2.53 \pm 0.13 & = & 0.50 \pm 0.05 \cdot 5.00 \pm 0.05 \end{array}$$

Calcular 1000  
valores e pegar  
do primeiro ao  
terceiro quartis

Gerar 1000  
valores  
aleatórios entre  
0.45 e 0.55

Gerar 1000  
valores  
aleatórios entre  
4.95 e 5.05

Quimiostatos

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Vantagens em relação ao método tradicional de estimativa de incertezas:

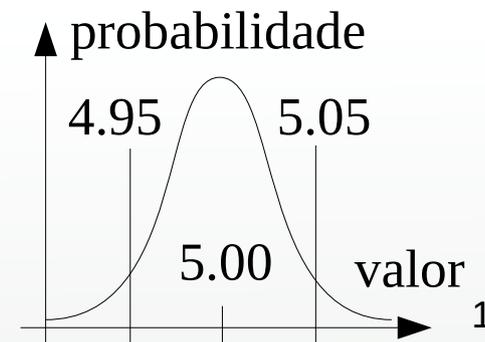
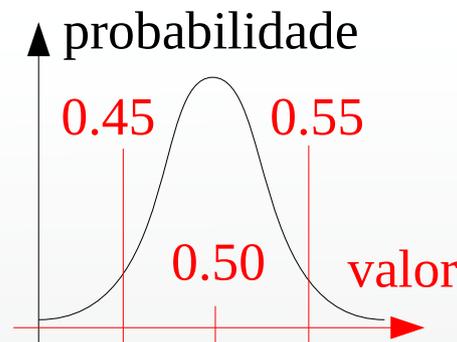
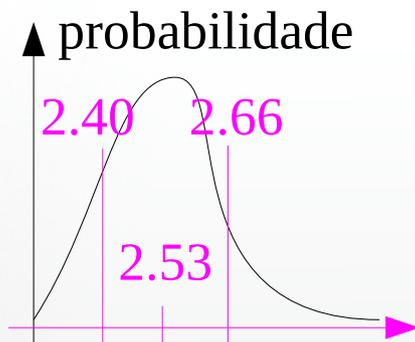
- não precisa saber fórmulas nem cálculo diferencial
- resulta num intervalo de confiança significativo
- funciona com distribuições não-gaussianas

Por quê?

Cuidado!

$$\begin{array}{rcl} m_B & = & [B] \cdot V \\ 2.5 & = & 0.5 \cdot 5 \\ 2.53 \pm 0.13 & = & 0.50 \pm 0.05 \cdot 5.00 \pm 0.05 \end{array}$$

Quimiostatos



# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

Vantagens em relação ao método tradicional de estimativa de incertezas:

- não precisa saber fórmulas nem cálculo diferencial
- resulta num intervalo de confiança significativo
- funciona com distribuições não-gaussianas

Por quê?

$$\begin{aligned} m_B &= [B] \cdot V \\ 2.5 &= 0.5 \cdot 5 \\ 2.53 \pm 0.13 &= 0.50 \pm 0.05 \cdot 5.00 \pm 0.05 \end{aligned}$$

Monte-Carlo

Quimiostatos

$$2.50 \pm 0.25 = 0.50 \pm 0.05 \cdot 5.00 \pm 0.05$$

Método tradicional

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

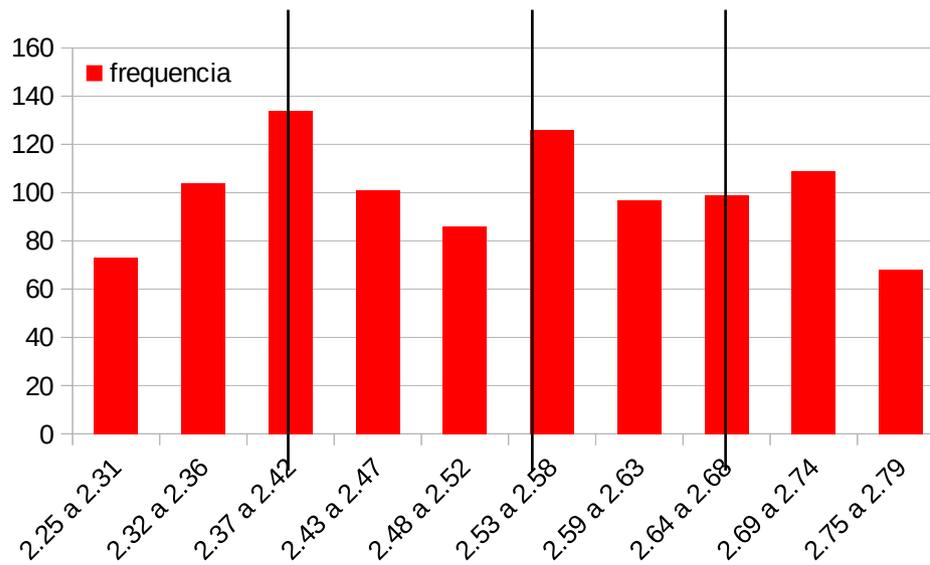
O quê?

Vantagens em relação ao método tradicional de estimativa de incertezas:

- não precisa saber fórmulas nem cálculo diferencial
- resulta num intervalo de confiança significativo
- funciona com distribuições não-gaussianas

Por quê?

Cuidado!



O resultado ( $m_B$ ) não parece ser uma gaussiana.

Mesmo assim, o método continua o mesmo se formos calcular outra variável a partir de  $m_B$  (ex.:  $q_S$ )

Quimiostatos

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

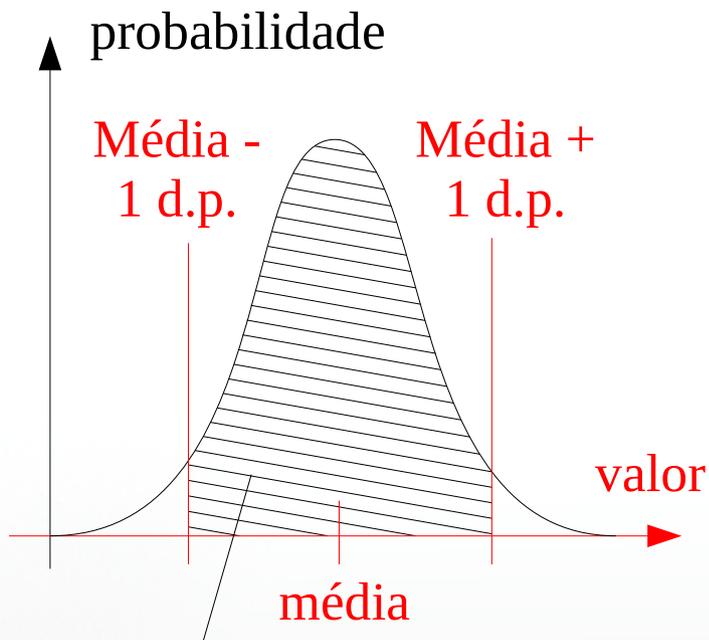
O quê?

Numa distribuição gaussiana, 1 desvio-padrão em torno da média representa uma confiança de ~68%:

Por quê?

Cuidado!

Quimiostatos



# O que é reconciliação de dados?

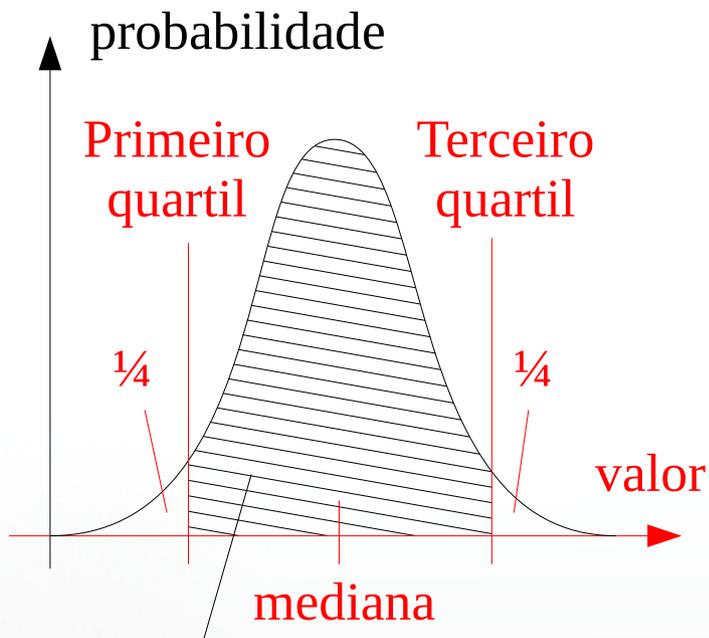
## Incertezas via simulação de Monte-Carlo

O quê?

Primeiro quartil é o valor que divide a amostra em  $\frac{1}{4}$  e  $\frac{3}{4}$ . O segundo quartil é a mediana ( $\frac{1}{2}$  e  $\frac{1}{2}$ ). E o terceiro quartil divide a amostra em  $\frac{3}{4}$  e  $\frac{1}{4}$ :

Por quê?

Cuidado!



Quanto essa área representa da área total da curva?

Quimiostatos

# O que é reconciliação de dados?

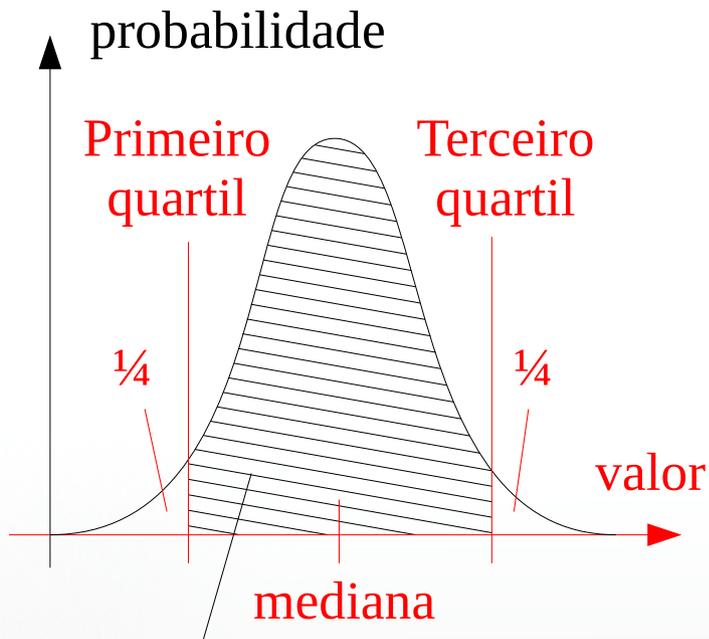
Incertezas via simulação de Monte-Carlo

O quê?

Primeiro quartil é o valor que divide a amostra em  $\frac{1}{4}$  e  $\frac{3}{4}$ . O segundo quartil é a mediana ( $\frac{1}{2}$  e  $\frac{1}{2}$ ). E o terceiro quartil divide a amostra em  $\frac{3}{4}$  e  $\frac{1}{4}$ :

Por quê?

Cuidado!



Quanto essa área representa da área total da curva?

Então, se o intervalo de confiança está entre o primeiro e o terceiro quartis, qual é o valor da confiança?

Quimiostatos

# O que é reconciliação de dados?

Incertezas via simulação de Monte-Carlo

O quê?

- 50% de confiança parece pior do que os 68% de confiança da abordagem tradicional usando média + 1 desvio-padrão

Por quê?

- Mas esse 50% já é da variável final calculada (aqui no exemplo, a massa de biomassa) com todas as incertezas das medições originais todas embutidas!

Cuidado!

Quimiostatos

# O que é reconciliação de dados?

## Incertezas via simulação de Monte-Carlo

O quê?

- Mas esse 50% já é da variável final calculada (aqui no exemplo, a massa de biomassa) com todas as incertezas das medições originais todas embutidas!

Por quê?

- Como o método tradicional de propagação de incertezas **superestima** as incertezas e **assume** uma distribuição para o resultado (ex.: gaussiana, beta, chi-quadrado, t-student, etc), se o resultado não tem essa distribuição, o valor da confiança da variável final pode ser completamente falso!

Cuidado!

Quimiostatos

# O que é reconciliação de dados?

Incertezas via simulação de Monte-Carlo

O quê?

- Como o método tradicional de propagação de incertezas **superestima** as incertezas e **assume** uma distribuição para o resultado (ex.: gaussiana, beta, chi-quadrado, t-student, etc), se o resultado não tem essa distribuição, o valor da confiança da variável final pode ser completamente falso!

Por quê?

Cuidado!

$$2.53 \pm 0.13 = 0.50 \pm 0.05 \cdot 5.00 \pm 0.05$$

Monte-Carlo (50%)

Quimiostatos

$$2.50 \pm 0.25 = 0.50 \pm 0.05 \cdot 5.00 \pm 0.05$$

Tradicional (68% ?)

# Por quê é importante?

O quê?

Geral: ter mais confiança e menos viés nas conclusões que você tira dos resultados experimentais

Por quê?

Especificamente no nosso contexto, o objetivo é aumentar a confiança e reduzir o viés ao:

- validar redes metabólicas
- propor modelos fenomenológicos
- escolher as condições do próximo experimento
- identificar fontes de erro experimental
- diagnosticar problemas em equipamentos
- prever custos de um bioprocessamento numa escala maior

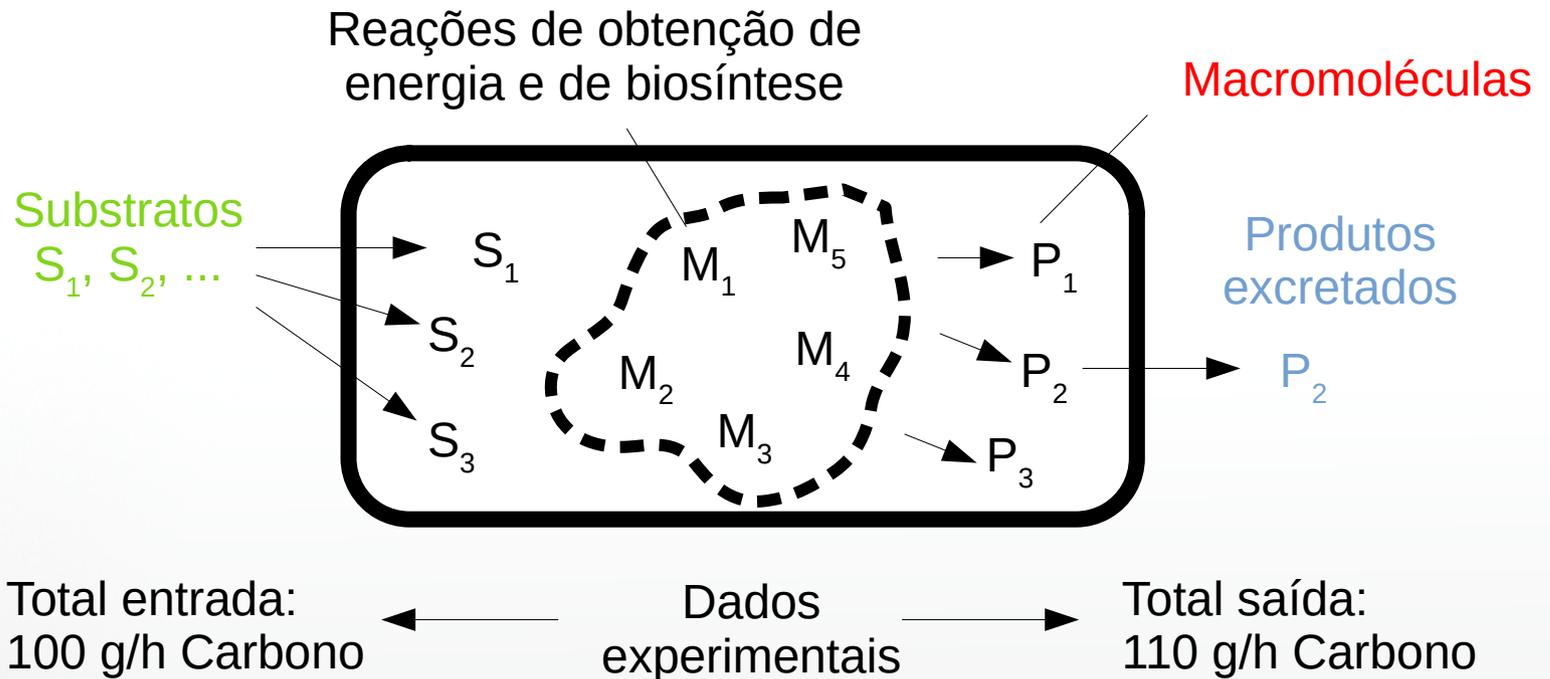
Cuidado!

Quimiostatos

# Por quê é importante?

## Exemplo motivacional

Validação de uma rede metabólica: o nosso modelo das reações internas (aquelas envolvendo os metabólitos  $M_1$ ,  $M_2$ , etc) é no melhor caso tão confiável quanto os dados experimentais usados para validá-lo:



# Reconciliação não é fabricação!

O quê?

Estatística é a arte de torturar os números até eles dizerem o que você quer...

Por quê?

...até a hora em que o biorreator explode, o seu artigo é rejeitado, etc.

Cuidado!

Quimiostatos

# Reconciliação não é fabricação!

Então até onde a estimativa é razoável?

- as restrições teóricas devem ser sólidas e deve haver indícios/evidências fortes de que a teoria se aplica ao seu caso. Ex: balanço de massa

- qualquer estimativa fora da incerteza de uma medição deve ser muito bem justificada. Ex: a coluna do cromatógrafo foi trocada logo depois dessas medições e todas as amostras repetidas após a troca estavam entre 25 e 30% acima dos valores anteriores

O quê?

Por quê?

Cuidado!

Quimiostatos

# Reconciliação não é fabricação!

Então até onde a estimativa é razoável?

- podem haver limites mínimos e máximos teóricos de incerteza, a menos de erros grosseiros (que entram no item anterior). Ex: no Laboratório de Bioprodutos nós temos no mínimo 3% de incerteza nos valores de fatores de conversão calculados a partir de dados experimentais gerados nos nossos biorreatores operando em modo quimiostato

O quê?

Por quê?

Cuidado!

Quimiostatos

# Exercício quimiostatos

## Planilha exercício 6.1

O quê?

a) plotar  $[S]_{\text{out}}(t)$  e  $[B]_{\text{out}}(t)$  para decidir quais pontos considerar em estado estacionário

Por quê?

b) apenas para os pontos em estado estacionário, calcular  $\mu$  e  $q_s$

Cuidado!

Quimiostatos

# Exercício quimiostatos

## Planilha exercício 6.2

O quê?

c) plotar no mesmo gráfico o  $\mu(q_s)$  experimental do item anterior e este  $\mu(q_s)$  teórico:

$$\mu(q_s) = \frac{0.46}{1 + e^{-8 \cdot (q_s - 0.5)}}$$

Cuidado!

d) criar 4 pares de colunas  $M_x$  e  $x_{rec}$ , onde  $x = Fin, FpH, [S]_{out}$  e  $[B]_{out}$ . Nas colunas  $M_x$ , colocar zero. Nas  $x_{rec}$ , colocar:  $x_{rec} = (1 + M_x * s_x)$

Quimiostatos

# Exercício quimiostatos

## Planilha exercício 6.2

O quê?

e) criar mais 2 colunas,  $\mu_{rec}$  e  $q_s_{rec}$ , e calculá-las com as colunas  $x_{rec}$  do item anterior

Por quê?

f) plotar os pontos do item e no mesmo gráfico dos itens b e c. Chutar valores entre -1 e +1 para as colunas  $M_x$  a fim de aproximar os pontos do item e à curva do  $\mu(q_s)$  teórico

Cuidado!

Quimiostatos

# Exercício quimiostatos

## Planilha exercício 6.3

O quê?

Por quê?

Cuidado!

Quimiostatos

g) Cada coluna  $s_x$  é a incerteza da medição  $x$  estimada a partir de especificações dos equipamentos e testes de reprodutibilidade das técnicas analíticas utilizadas. Use os pontos em estado estacionário para determinar se essas estimativas estão subestimadas ou superestimadas para este conjunto de dados experimentais.

h) Repita o item g incluindo os pontos em modo de operação contínuo transiente.

