

Estatística Aplicada a Ciências Ambientais

Incerteza

Daniel Detzel

detzel@ufpr.br



Agenda

Incerteza

- fundamentos

- erros em medições

- medidas de precisão

Quantificação da incerteza

- intervalos de confiança

- amostragem

Hands-on-work: Trabalho

INCERTEZA

fundamentos

As incertezas estão ligadas à **falta de confiança** em um determinado sistema, processo, ou resultado

Podem ser classificadas em:

- epistêmica: causada pela falta de informações

- ontológica: causada pelas variações aleatórias nos sistemas

- ambígua: emerge na busca por consenso em problemas complexos

 - ex.: explicações para um racionamento de água

 - visão 1: sistema de abastecimento subdimensionado

 - visão 2: excesso de consumo

 - (...)

“Qualquer tentativa para lidar com incertezas na gestão de recursos naturais deveria também incluir a **pluralidade de perspectivas** ligadas aos problemas estudados”.
(Coelho, 2019, tradução livre)

Incertezas epistêmicas são minimizadas conhecendo-se mais o sistema mais dados, informações, coletas, etc.

Incertezas ontológicas **não podem ser minimizadas**, apenas mais bem conhecidas

Ex.: lista de incertezas em dados de vazão obtidos via réguas linimétricas

1. Escolha da seção do rio

posição e geometria

2. Medições dos níveis

geralmente a olho nu

3. Método de obtenção da curva-chave (relação cota vs. nível) e intervalo de validade

ajuste numérico e extrapolações

4. Alterações da curva-chave por razões físicas

modificações no leito e/ou interferências na posição das réguas

Ex.: lista de incertezas em dados de qualidade da água

1. Estratégia de monitoramento

- escolha de local
- frequência de medições
- escolha de variáveis

2. Métodos de medição

- técnicas e equipamentos
- precisão das medições
- transporte de amostras

3. Trabalho em laboratório

- armazenamento de amostras
- manejo de amostras
- execução das análises

INCERTEZA

erros em medições

“A próxima medição que você fizer (...) estará corrompida por um erro experimental. Isso é um fato da vida”.

(Berthouex e Brown, 2002)

As medições de um certo valor **não serão idênticas** mesmo que os experimentos sejam os **mais idênticos possíveis**

incerteza ontológica

Erro experimental é o desvio que o valor observado tem em relação ao seu valor **verdadeiro**

Os valores observados x_i irão se diferenciar dos valores verdadeiros χ por um erro ε_i

$$x_i = \chi + \varepsilon_i$$

O erro pode ter componentes sistemáticas τ_i e aleatórias e_i . Portanto:

$$x_i = \chi + (\tau_i + e_i)$$

Incerteza | erros em medições

Erros sistemáticos: causados por uma tendenciosidade consistente em relação ao valor verdadeiro

ex.: falta de calibração de equipamentos, alteração em níveis de referência, etc.

Podem ser identificados por *offsets* (desvios contínuos) nos valores

Erros sistemáticos podem ser removidos, restando apenas erros aleatórios:

$$x_i = \chi + e_i$$

Incerteza | erros em medições

Erros aleatórios: causados pelos processos de obtenção das medições, sejam em campo ou em laboratório

Não podem ser removidos! (apenas minimizados)

Não podem ser previstos com base em erros anteriores

erros aleatórios são avaliados coletivamente usando toda a amostra

A parcela e_i é resultante da soma de diversos processos, cada qual com seu erro aleatório individual

os erros podem ser positivos ou negativos na mesma proporção

Incerteza | erros em medições

Aleatórios

Resultados de
processos
aditivos

Positivos ou
negativos em
igual proporção


$$e_i \sim N(0, s_e)$$

Os erros aleatórios das medições são normalmente distribuídos
lembrando que Gauss deduziu a equação normal justamente para modelar
erros de medições!

Valores típicos para erros de medições (compilados por Coelho, 2019)

Etapas	Variáveis	Estimativas (\pm)
Amostragem em campo	Vazão	6% a 19%
	Constituintes dissolvidos	4% a 47%
	Sólidos suspensos totais, Nitrogênio e Fósforo	4% a 50%
Armazenamento de amostras	Constituintes dissolvidos	2% a 16%
	Nutrientes totais	7% a 9%
Análises laboratoriais	Constituintes dissolvidos	4% a 26%
	Sólidos suspensos totais	1% a 5%
	Nutrientes totais	3% a 32%

INCERTEZA

precisão

Erros aleatórios podem ser quantificados por meio de réplicas dos experimentos

a avaliação é feita considerando a **dispersão** entre as medidas obtidas
grandes dispersões → grandes erros → maior incerteza

A relação dos erros aleatórios com a dispersão das medidas é chamada de **precisão**

resultados precisos possuem menores dispersões e, conseqüentemente, menores erros e menor incerteza

Os erros podem ser mapeados ao analisar as **médias** de diferentes réplicas dos experimentos

Para um só experimento, a precisão pode ser quantificada por meio do **erro padrão da média** (ou simplesmente erro padrão), calculado por:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

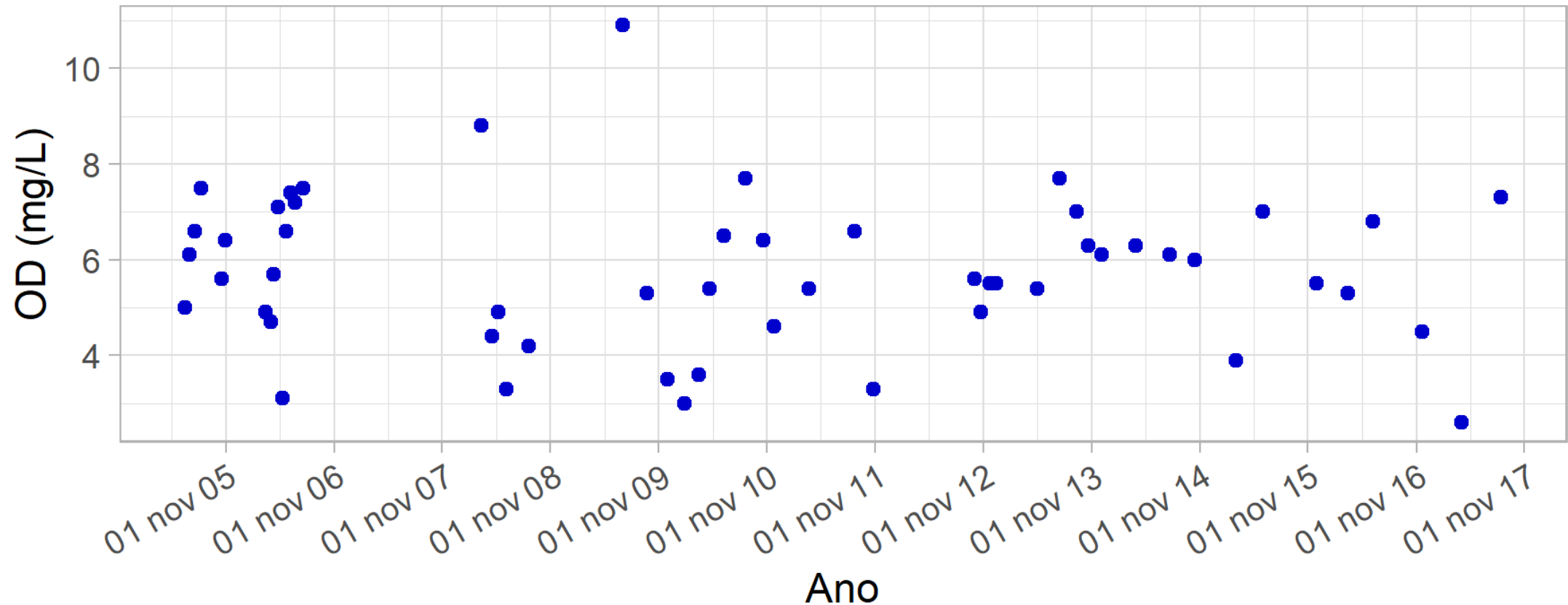
onde

s desvio padrão da amostra

n tamanho da amostra

Incerteza | precisão

Exemplo: Considere a série de Oxigênio Dissolvido medido entre 17 jun 2005 e 14 ago 2017 no posto IG1 (rio Iguaçu)



Ex.: (cont.)

Para esta série, os parâmetros são:

$n = 53$ observações

$\bar{x} = 5,74$ mg/L

$s = 1,57$ mg/L

O erro padrão da média é:

$$s_{\bar{x}} = \frac{1,57}{\sqrt{53}} = 0,21 \text{ mg/L}$$

Ex.: (cont.)

Interpretações sobre os estimadores

$n = 53$ observações | tamanho da amostra utilizada para estimar a verdadeira média do processo OD no posto IG1 do rio Iguaçu

$\bar{x} = 5,74$ mg/L | estimativa da verdadeira média do processo OD (...)

$s = 1,57$ mg/L | estimativa da dispersão das amostras no entorno da média amostral

$s_{\bar{x}} = 0,21$ mg/L | estimativa da dispersão das médias das amostras no entorno da verdadeira média do processo OD. Quantifica a precisão das medições

QUANTIFICAÇÃO DA INCERTEZA

intervalos de confiança

Incerteza | intervalos de confiança

Até então, as diversas estatísticas descritivas trabalhadas se referem a **estimativas pontuais** de parâmetros (a média, a variância, a assimetria, etc.)

Essas estimativas, por si só, não trazem a ideia de variabilidade ou de incerteza dos processos envolvidos

Para isso, empregam-se estimativas intervalares de parâmetros
os intervalos são determinados de modo a fornecer a **probabilidade** de que os verdadeiros parâmetros populacionais estejam neles contidos
por terem esse propósito, são chamados de **intervalos de confiança**

Incerteza | intervalos de confiança

Assim sendo, o **nível de confiança** traduz a probabilidade de que o parâmetro populacional em estudo esteja dentro do intervalo de confiança
o nível de confiança é a probabilidade complementar do nível de significância α

$$\alpha = 1 - \text{nível de confiança}$$

Interpretação: Suponha que diferentes campanhas de medição de uma variável foram feitas e que se deseja saber o valor médio desta variável.

para um nível de confiança de 90% (ou $\alpha = 10\%$), a verdadeira média populacional estará contida dentro do intervalo de confiança em 90% das campanhas

Incerteza | intervalos de confiança

Exemplo: Suponha, **hipoteticamente**, que a média e o desvio padrão **populacionais** da série de OD do rio Iguaçu no posto IG1 sejam conhecidas e valham 5,74 mg/L e 1,57 mg/L, respectivamente

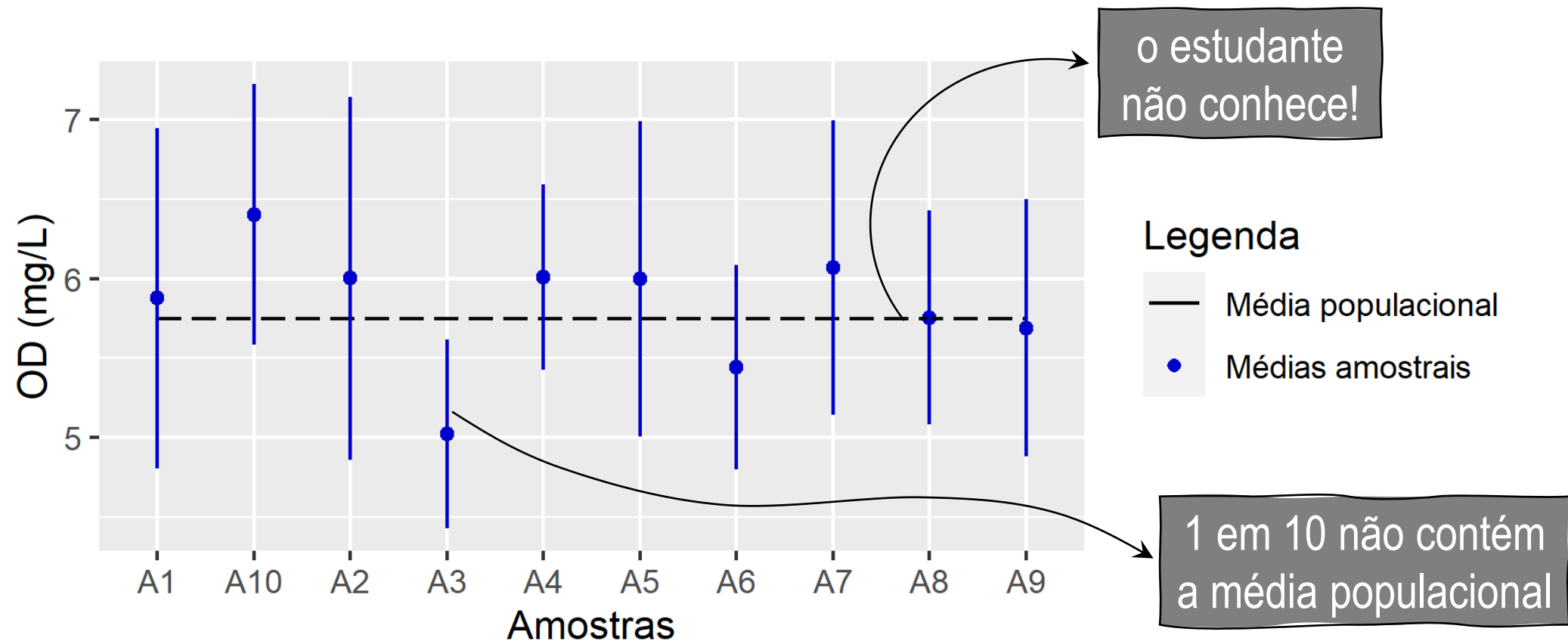
Um estudante dispõe de recursos para 10 campanhas de medição durante o seu curso e que 12 réplicas são realizadas em cada uma dessas campanhas

Ao final das campanhas, o estudante calcula os intervalos de confiança para um nível de 10% (o estudante não conhece os valores populacionais)

Incerteza | intervalos de confiança

Exemplo: (cont.)

Assumindo que a série tenha distribuição normal, os resultados seriam:



Incerteza | intervalos de confiança

O intervalo de confiança para a média populacional μ é dado por:

$$\bar{x} + t_{(\alpha/2, n-1)} \sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{x} + t_{(1-\alpha/2, n-1)} \sqrt{\frac{s^2}{n}}$$

onde

\bar{x}

média amostral

s^2

variância amostral

$t_{(\alpha/2, n-1)}, t_{(1-\alpha/2, n-1)}$

valores retirados da distribuição t-Student com significância α e $n - 1$ graus de liberdade

Incerteza | intervalos de confiança

Exemplo: Estimar o intervalo de confiança para série de OD no posto IG1 do rio Iguaçu e $\alpha = 5\%$

Dados calculados anteriormente:

$n = 53$ observações

$\bar{x} = 5,74$ mg/L

$s = 1,57$ mg/L

Valores de t :

$$t_{(\alpha/2, n-1)} = t_{(0,05/2, 53-1)} = -2,001$$

$$t_{(1-\alpha/2, n-1)} = t_{(1-0,05/2, 53-1)} = 2,001$$

No R:

```
qt(0.025, 52)
```

```
[1] -2.006647
```

```
qt(0.975, 52)
```

```
[1] 2.006647
```

Incerteza | intervalos de confiança

Assim:

$$5,74 + (-2,001) \sqrt{\frac{1,57^2}{53}} \leq \mu \leq 5,74 + 2,001 \sqrt{\frac{1,57^2}{53}}$$

$$5,31 \leq \mu \leq 6,18$$

No R:

```
t.test(amostra, conf.level = 0.95)$"conf.int"  
[1] 5.311128 6.179438
```

Incerteza | intervalos de confiança

Os intervalos de confiança são tipicamente simétricos no entorno do parâmetro de interesse

há possibilidade de determinar intervalos de confiança assimétricos

Existem equações diferentes para intervalos de confiança de outros parâmetros (variâncias, percentis, etc.)

contudo, o procedimento é similar ao mostrado

Para casos assim, recomenda-se consulta a Helsel et al. (2020, cap. 3)

QUANTIFICAÇÃO DA INCERTEZA

amostragem

Os métodos mostrados possuem um desdobramento interessante e altamente aplicável, ligado à [amostragem](#)

A equação do intervalo de confiança pode ser reescrita na forma:

$$\bar{x} \pm t_{(\alpha/2, n-1)} \sqrt{\frac{s^2}{n}} \leq \mu \rightarrow t_{(\alpha/2, n-1)} \sqrt{\frac{s^2}{n}} \leq \mu \pm \bar{x}$$

A porção $\mu \pm \bar{x}$ representa a (im)precisão da estimativa amostral no entorno do valor populacional

Assumindo que isso seja o erro cometido pelo analista ao estimar a média, pode-se fazer:

$$e = \mu \pm \bar{x}$$

Assim:

$$t_{(\alpha/2, n-1)} \sqrt{\frac{s^2}{n}} \leq e$$

Pode-se, ainda, assumir uma segunda premissa:

$$t_{(\alpha/2, n-1)} \rightarrow z_{\alpha/2}$$

onde $z_{\alpha/2}$ é retirado da distribuição normal padrão para significância α .
Portanto:

$$z_{\alpha/2} \sqrt{\frac{s^2}{n}} \leq e$$

Isolando-se n , chega-se a:

$$n = \left(\frac{Z_{\alpha/2} \cdot s}{e} \right)^2$$

Com isso, tem-se uma maneira de dimensionar a amostra necessária para estimar a média populacional do parâmetro de interesse

importante: é uma estimativa **aproximada**, dadas as premissas assumidas

Exemplo: qual é o tamanho da amostra necessária para estimar a média de OD no posto IG1 com um erro de $\pm 0,20$ mg/L e $\alpha = 5\%$?

Para $\alpha = 5\%$, $z_{\alpha/2} = 1,96 [q_{\text{norm}}(0.025)]$

$$\therefore n = \left(\frac{1,96 \cdot 1,57}{0,20} \right)^2$$

$$n \cong 237$$

Incerteza | amostragem

É uma quantidade grande de medições! Pode-se aumentar α para 10%:

Para $\alpha = 10\%$, $z_{\alpha/2} = 1,64 [\text{qnorm}(0.05)]$

$$\therefore n = \left(\frac{1,64 \cdot 1,57}{0,20} \right)^2$$

$$n \cong 165$$

Ainda assim é alto. Então, talvez seja preciso ser mais modesto no erro que se deseja cometer... (para $e = 0,50$ mg/L e $\alpha = 10\%$, $n = 27$)

Revisão

Dados de ciências ambientais contém erros, o que agrega certo grau de incerteza para as medições

A precisão de um estimador pode ser mensurada pelo erro padrão da média

quanto menor o seu valor, menor a incerteza

A incerteza pode ser quantificada por meio de intervalos de confiança
mensuram a probabilidade de o verdadeiro parâmetro populacional estar
contido no intervalo calculado

Os intervalos de confiança fornecem meios para se dimensionar o tamanho da amostra necessária para estimar um parâmetro de interesse



Estatística Aplicada a Ciências Ambientais

Daniel Detzel
detzel@ufpr.br