

Universidade Federal Fluminense



## Gerência de Redes e Engenharia de Tráfego

Professor Cledson de Sousa

Versão: 0.1.111– 13 de maio/2024 – Até MPLS.



## Conteúdo


<b>1 O Conceito de Camadas . . . . .</b>	<b>3</b>
<b>2 Variação e Incertezas . . . . .</b>	<b>4</b>
<b>Bibliografia . . . . .</b>	<b>6</b>

## Prefácio

Caros futuros engenheiros,

Bem-vindos ao curso de Gerência de Redes e Engenharia de Tráfego, uma disciplina criada para introduzir os conceitos fundamentais e práticas avançadas da controle de tráfego. administração e gerência de modernas infraestruturas de rede. Este curso, estruturado com uma mistura equilibrada de aulas em classe e extra-classe, cobre uma carga horária total de 60 horas, das quais 40 horas são dedicadas a conceitos teóricos e 20 horas a aplicações fora de sala. Então o objetivo é proporcionar aos alunos uma compreensão abrangente dos modelos de gerência de redes, monitoramento, auditoria e a engenharia de tráfego necessária para lidar com o encaminhamento do tráfego.

Esta apostila, embora tenha um caráter acadêmico, não possui a pretensão de ser extremamente rigorosa, especialmente na forma. O autor empenhou-se em dar o devido crédito a todas as fontes utilizadas; no entanto, por vezes, estende-se o texto sem as devidas citações específicas dos autores originais. Então peço que aceitem as escusas do autor, pois muitas das fontes são os textos originais presentes nos locais onde as figuras foram extraídas. Além disso, diversas explicações são fruto do esforço do autor em dissecar os diagramas, que frequentemente simplificam o verdadeiro encaminhamento dos pacotes.

Ao longo do texto há diversos sinais de parada , em notas de margem, gráficos, figuras e outros sinais. Estes sinais lá estão para o leitor desopilar, se você chegou em um desses sinais parabéns! É por que você já leu o suficiente, nesse momento é bom parar, descontraí, visitar os sites e sair um pouco do texto, naturalmente deve-se retornar ao texto tão logo possível. Essa estratégia funciona comigo! Um curso leve, produtivo, enriquecedor e sem reprovações. É o que desejo!

### **Sobre o autor:**

O autor recebeu seu título de Doutor em Computação pela Universidade Federal Fluminense (UFF) em 2019, e seus títulos de graduação e mestrado em Engenharia de Telecomunicações pela mesma universidade, em 1997 e 2013. Com mais de 30 anos de experiência na indústria de telecomunicações, desde 2021 é Professor Adjunto do Departamento de Engenharia de Telecomunicações da Universidade Federal Fluminense. Seus interesses de pesquisa atuais incluem redes de sensores sem fio, SDN, rádios cognitivos e CSI.

### **Plataformas:**

- LinkedIn: <https://www.linkedin.com/in/cledsonsousa>
- Página pessoal: <https://cledsonsousa.github.io>
- Currículo Lattes: <http://lattes.cnpq.br/7195080748145566>

## 1 O Conceito de Camadas

O conceito de camadas (*layering*) é fundamental para a visualização de dados. Vamos explorar como esse conceito facilita essas representações:

### Separação de Elementos Visuais:

Camadas permitem separar diferentes elementos visuais em uma visualização de dados. Por exemplo, ao representar um conjunto de dados, uma camada pode ser usada para mostrar os valores médios, enquanto outra camada pode ser usada para mostrar a variação desses valores (como barras de erro ou intervalos de confiança). Isso ajuda a distinguir claramente entre os dados principais e as informações auxiliares que podem por exemplo representar incerteza.

### Flexibilidade na Visualização:

O uso de camadas oferece flexibilidade na criação de gráficos complexos. Diferentes tipos de representações visuais, como pontos de dados, linhas de tendência, e intervalos de confiança, podem ser adicionados ou removidos conforme necessário. Isso permite que os visualizadores ajustem a complexidade da visualização com base em suas necessidades específicas.

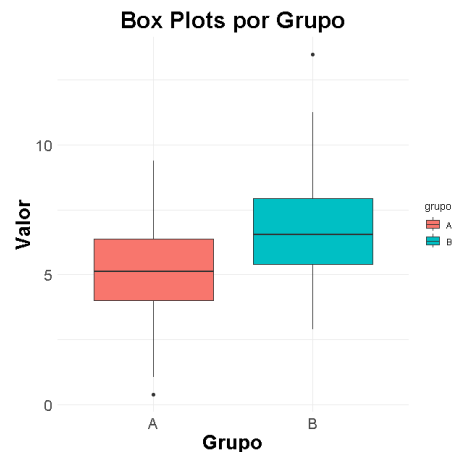


Figura 1

### Exemplo Prático:

*Box plots* como o apresentados na Figura 1, por exemplo, utilizam camadas. A camada do *box plot* mostra a mediana e os quartis, enquanto camadas adicionais indicam *outliers*.

## 2 Variação e Incertezas

O conceito de camadas também facilita a representação da variação estatística e da incerteza. É importante se lembrar que dados experimentais tipicamente apresentam alguma incerteza e a visualização e representação destes dados deve "caracterizar a magnitude dessa incerteza em relação aos dados reais [1].

Os autores infelizmente mostram resultados que em muitos trabalhos as figuras publicadas não atendem a esse padrão, especialmente à medida que a dimensionalidade dos dados aumenta [2]. Quando o objetivo de uma visualização é comparar uma quantidade medida ou derivada entre categorias ou condições, deve-se incluir um elemento (são as *geom* do *ggplot*) gráfico retratando a quantidade e um segundo elemento retratando a incerteza da quantidade. Note que, sem a representação da incerteza, uma comparação visual precisa não é possível, os leitores podem tirar conclusões incorretas ou mal informadas.

A variação e a incerteza podem ser retratadas com uma variedade de *geoms*, mas são mais comumente exibidas com barras de erro. Infelizmente, não há um padrão único para o que a barra de erro deve representar já que há uma pluralidade esmagadora de significados possíveis, como desvio padrão (DP) da amostra, erro padrão da média (EPM), Intervalo de Confiança (IC) paramétrico de  $100(1-\alpha)\%$ , intervalo de probabilidade bayesiano, um intervalo de previsão, etc. Cada quantidade tem sua própria interpretação estatística

Portanto, ao usar barras de erro, certifique-se de que (1) a quantidade codificada pela barra é consistente com o objetivo da visualização e (2) a quantidade é definida de forma inequívoca. Em relação ao primeiro ponto, oferecemos as seguintes diretrizes ao usar barras de erro para retratar a variação de uma estimativa de parâmetro ou a variação dos dados.

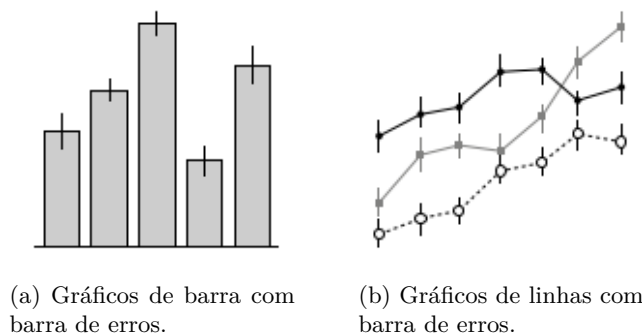


Figura 2: as barras normalmente para comparar categorias e as linhas para comparar grupos e tendências.

Se o interesse for estimar um parâmetro populacional, como a média ou a variância, então a variação da estimativa (isto é, a distribuição amostral da estatística) é desejada. Exemplos de barras de erro adequadas incluem o EPM ou um IC paramétrico ou *bootstrap* de 95%, como visto em visualizações que

ênfatisam comparações (Figura 2). ICs paramétricos devem ser usados apenas se os dados atenderem às suposições do modelo subjacente, caso contrário, um bootstrap (ou outra estratégia para aproximar a distribuição amostral) deve ser usado<sup>1</sup>.

#### IC um conceito muitas vezes mal entendido [3] e [4]

Um intervalo de confiança de  $100(1-\alpha)\%$  para um parâmetro populacional é um intervalo calculado a partir dos dados amostrais que, em  $100(1-\alpha)\%$  das amostras possíveis, conteriam o verdadeiro valor do parâmetro.

**Isto é:** Suponha que você calcule a média de alturas de uma amostra de 100 pessoas e obtenha uma média de 170 cm com um desvio padrão de 10 cm. Um intervalo de confiança de 95% pode ser calculado, e você pode obter algo como (168 cm, 172 cm). Isso significa que, se você repetisse esse experimento várias vezes, 95% dos intervalos de confiança calculados conteriam a verdadeira média populacional (nesse caso 170 cm).

<sup>1</sup>E o que é  $\alpha$ ? o nível de significância:  $\alpha$  é a probabilidade de cometer um erro tipo I, que ocorre quando rejeitamos a hipótese nula ( $H_0$ ) quando ela é verdadeira. **Em outras palavras, é a taxa de falso positivo permitida.**

Um outro exemplo de interpretação geométrica dos intervalos de confiança são a sua sobreposição ou não entre duas amostras.

#### A sobreposição dos ICs [5] e [6]

Um intervalo de confiança de  $100(1-\alpha)\%$  para um parâmetro populacional é um intervalo calculado a partir dos dados amostrais que, em  $100(1-\alpha)\%$  das amostras possíveis, conteriam o verdadeiro valor do parâmetro.

**Isto é:** Suponha que você calcule a média de alturas de uma amostra de 100 pessoas e obtenha uma média de 170 cm com um desvio padrão de 10 cm. Um intervalo de confiança de 95% pode ser calculado, e você pode obter algo como (168 cm, 172 cm). Isso significa que, se você repetisse esse experimento várias vezes, 95% dos intervalos de confiança calculados conteriam a verdadeira média populacional (nesse caso 170 cm).

Enquanto ICs de 95% não sobrepostos indicam uma diferença significativa (sob um modelo de probabilidade normal), o inverso não é verdadeiro — dependendo do tamanho da amostra, as barras de IC podem se sobrepor em até 50% e ainda atender aos critérios de significância.

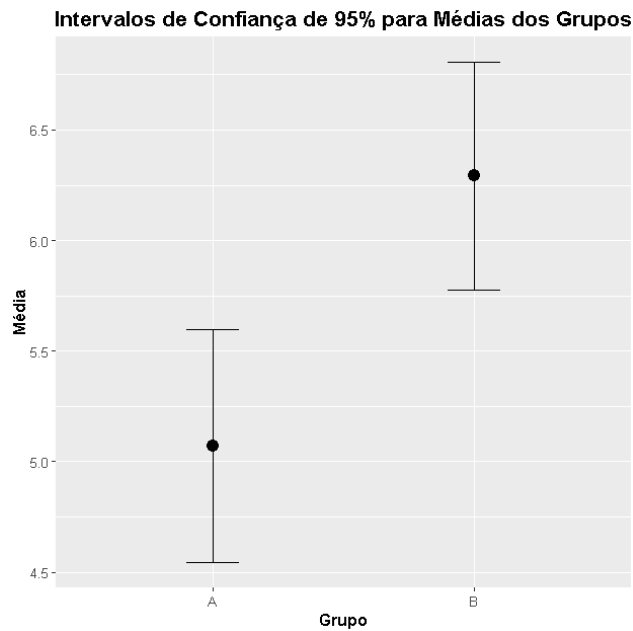


Figura 3: ICs de 95% não sobrepostos indicam uma diferença significativa (sob um modelo de probabilidade normal).

## Bibliografia

- [1] Howard Wainer. Depicting error. *The American Statistician*, 50(2):101–111, 1996.
- [2] Elena A Allen, Erik B Erhardt, and Vince D Calhoun. Data visualization in the neurosciences: overcoming the curse of dimensionality. *Neuron*, 74(4):603–608, 2012.
- [3] S. Belia, F. Fidler, J. Williams, and G. Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10(4):389–396, 2005.
- [4] R. Hoekstra, R. D. Morey, J. N. Rouder, and E. J. Wagenmakers. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5):1157–1164, 2014.
- [5] G. Cumming, F. Fidler, and D. L. Vaux. Error bars in experimental biology. *The Journal of Cell Biology*, 177(1):7–11, 2007.
- [6] M. Krzywinski. Points of view: Elements of visual style. *Nature Methods*, 10(5):371–371, 2013.