

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Apresenta  o da disciplina

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br

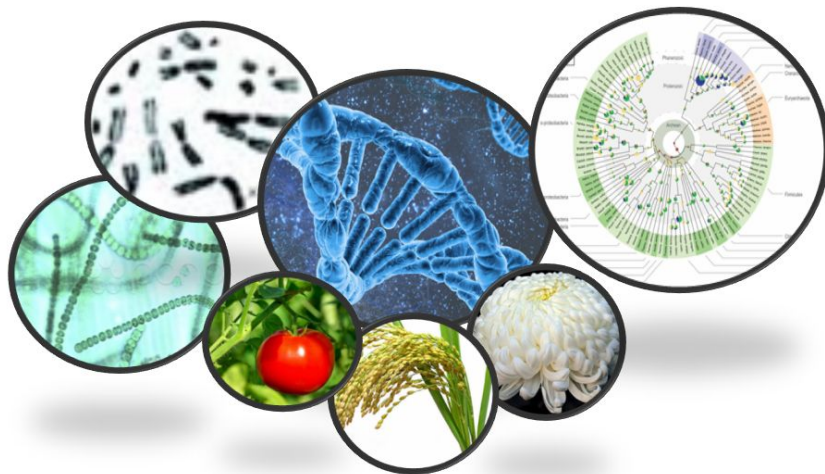


Sobre mim



Prof. Tetsu Sakamoto

- Biólogo/Bioinformata
- tetsu@imd.ufrn.br
- Sala A224
- Horários de atendimento: 24T34



Bioestatística

IMD0601

O que é e por
quê?

Bioestatística

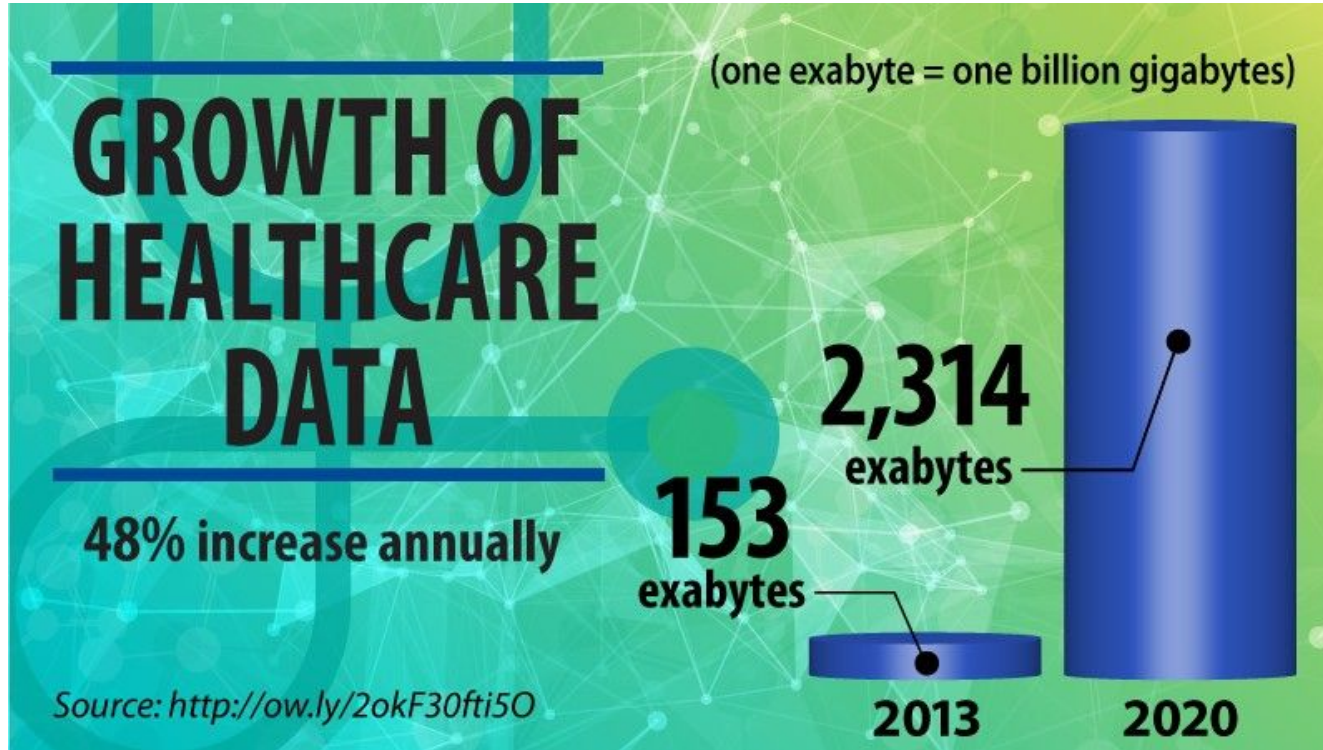
IMD0601

O que é e por quê?

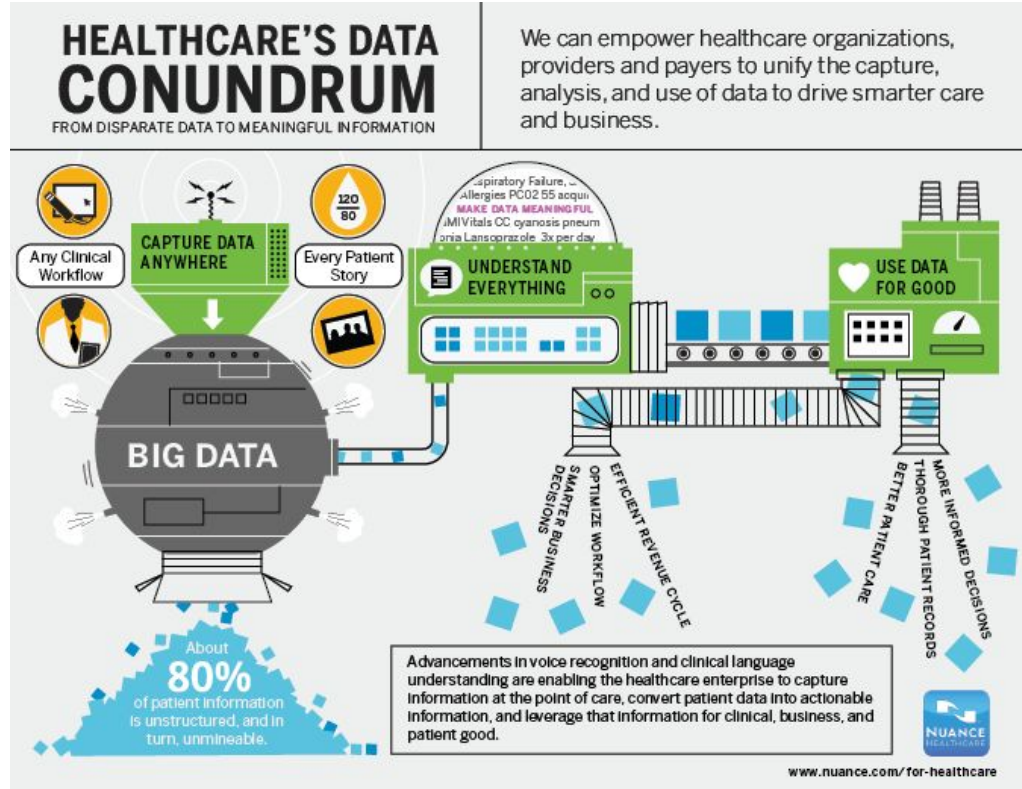
Estatística - uma área da ciência que procura responder questões baseando-se em dados.

- Desenhar o experimento;
 - Coletar dados de forma apropriada;
 - Analisar os dados e checar as hipóteses;
 - Extrair conclusões confiáveis;
-

Dados biológicos



Dados biológicos



Passo a passo de um experimento biológico

1. Especifique a pergunta biológica;
2. Transforme a pergunta na forma de hipóteses biológicas nula e alternativa;
3. Transforme a pergunta na forma de hipóteses estatísticas nula e alternativa;
4. Determine quais variáveis são importantes para a pergunta;

Passo a passo de um experimento biológico

5. Determine o tipo de cada variável;
6. Desenhe o experimento que controle ou randomize as variáveis de confusão;
7. Baseado no número e tipo de variáveis, no ajuste a uma distribuição de probabilidade e na hipótese a ser testada, escolha o melhor teste estatístico a ser utilizado;

Passo a passo de um experimento biológico

8. Se possível, realize a potência estatística para verificar um bom tamanho amostral para o experimento;
9. Realize o experimento;

Passo a passo de um experimento biológico

10. Examine os dados e verifique se eles cumprem com os requisitos do teste estatístico escolhido (teste de normalidade e de homocedasticidade). Se não, escolha outro teste;
11. Aplique o teste estatístico escolhido e interprete os dados;
12. Apresente os seus resultados efetivamente na forma de gráficos e/ou tabelas;

Exemplo

Varrelli e Eanes (2001)

- Mediu conteúdo de glicogênio em indivíduos de *Drosophila melanogaster*;
- As drosófilas são polimórficas no locus que codifica a enzima PGM (fosfoglucomutase) nos sítios 52 (V ou A) e 484 (V ou L);
- Todas as quatro combinações possíveis estavam presentes na população (V-V, V-L, A-V, A-L);

Exemplo

1. Especifique a pergunta biológica;

O polimorfismo no locus PGM influencia no conteúdo de glicogênio nas drosófilas?

2. Transforme a pergunta na forma de hipóteses biológicas nula e alternativa;

Hipótese nula: polimorfismo em PGM não afeta o conteúdo do glicogênio;

Hipótese alternativa: polimorfismo em PGM afeta o conteúdo do glicogênio;

3. Transforme a pergunta na forma de hipóteses estatísticas nula e alternativa;

Hipótese nula: drosófilas com diferentes sequências de PGM possui a mesma média do conteúdo de glicogênio;

Hipótese alternativa: drosófilas com diferentes sequências de PGM não possui a mesma média do conteúdo de glicogênio;

Exemplo

4. Determine quais variáveis são importantes para a pergunta;

O conteúdo de glicogênio e a sequência do PGM.

5. Determine o tipo de cada variável;

Conteúdo de glicogênio - variável quantitativa;

Sequência do PGM - variável qualitativa.

6. Desenhe o experimento que controle ou randomize as variáveis de confusão;

Outras variáveis que podem influenciar no experimento, como a idade ou o frasco de origem, foram controladas (foram utilizados drosófilas da mesma idade) ou randomizadas (drosófilas de diferentes frascos foram amostradas de forma aleatória).

Exemplo

7. Baseado no número e tipo de variáveis, no ajuste a uma distribuição de probabilidade e na hipótese a ser testada, escolha o melhor teste estatístico a ser utilizado;

Como o objetivo é comparar as médias de uma das variáveis entre grupos classificados por uma variável qualitativa, e como existem mais de duas categorias, um teste apropriado seria o ANOVA one-way

8. Se possível, realize a potência estatística para verificar um bom tamanho amostral para o experimento;

Para realizar esse tipo de análise, é requerido previamente uma estimativa do desvio padrão do conteúdo de glicogênio e do tamanho efetivo. Neste experimento, os pesquisadores tentaram amostrar o máximo de drosófilas possível em um tempo hábil;

Exemplo

9. Realize o experimento;

O conteúdo de glicogênio foi medido em drosófilas de diferentes genótipos.

10. Examine os dados e verifique se eles cumprem com os requisitos do teste estatístico escolhido (teste de normalidade e de homocedasticidade). Se não, escolha outro teste;

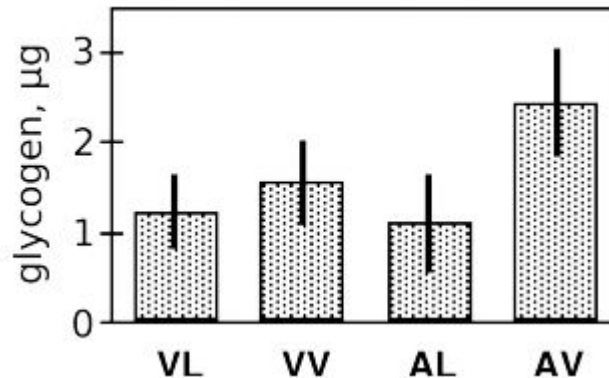
O teste ANOVA assume que os dados sigam uma distribuição normal e que seja homocedástico. Se ele não segue um desses requisitos, testes não paramétricos são mais apropriados.

Exemplo

11. Aplique o teste estatístico escolhido e interprete os dados;

O teste ANOVA foi realizado e resultado do p-valor foi menor que 0,05. A interpretação é que drosófilas com diferentes sequências de PGM possuem diferentes médias de conteúdo glicogênico.

12. Apresente os seus resultados efetivamente na forma de gráficos e/ou tabelas;



Objetivos

IMD0601 - Bioestatística

Abordar e aplicar as diferentes temáticas da estatística em um contexto biológico.

Estrutura da disciplina

IMD0601 - Bioestatística

Carga horária: 60 horas

Data: 11/02/2019 a 26/06/2019

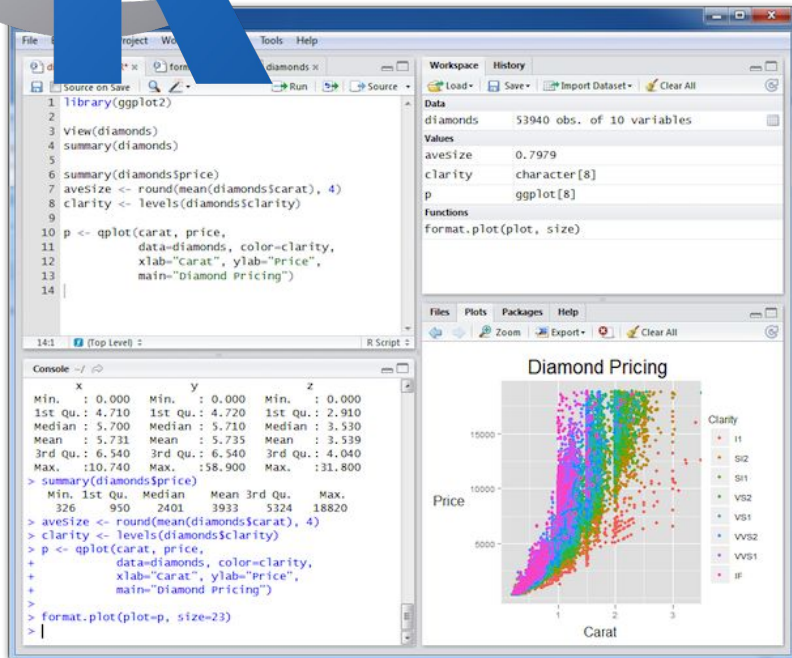
Horário: 24T12

Local: A304

Avaliações: 3 trabalhos

Aula de reposição

Cronograma e temas



Introdução ao R

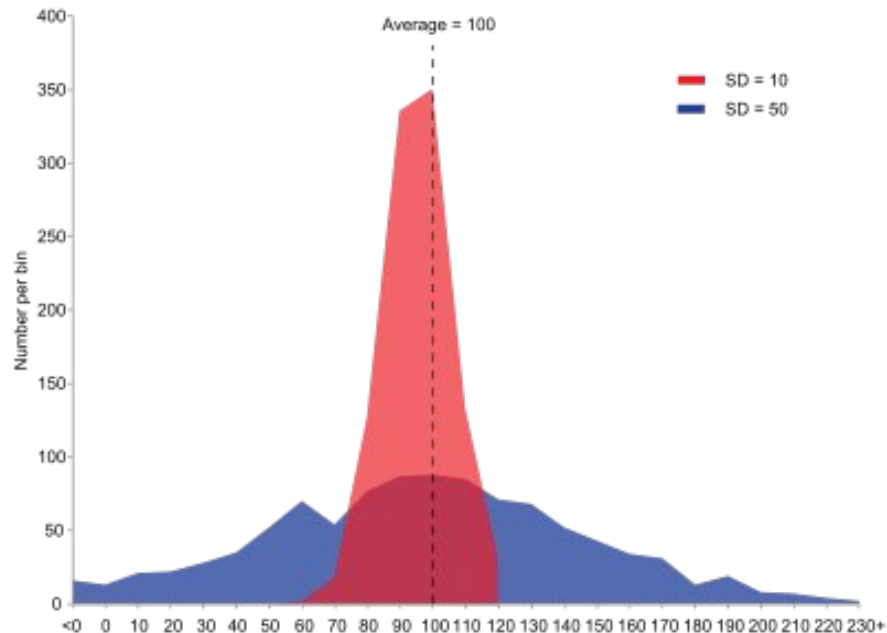
Organização dos dados

Visualização dos dados

$$\text{Mean}(\text{population}) = \mu = \frac{\sum_{i=1}^k f_i x_i}{n}$$

$$\text{StandardDeviation}(\text{population}) = \sigma = \sqrt{\frac{\sum_{i=1}^k f_i (x_i - \mu)^2}{n}}$$

$$\text{Variance}(\text{population}) = \sigma^2 = \frac{\sum_{i=1}^k f_i (x_i - \mu)^2}{n}$$



Estatística descritiva

Medidas de Tendência Central

Medidas de Dispersão

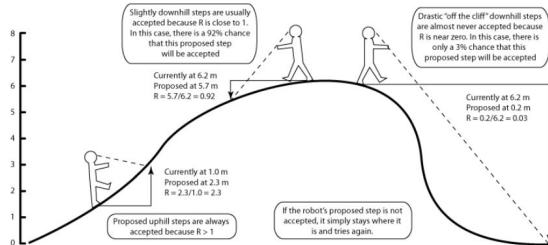
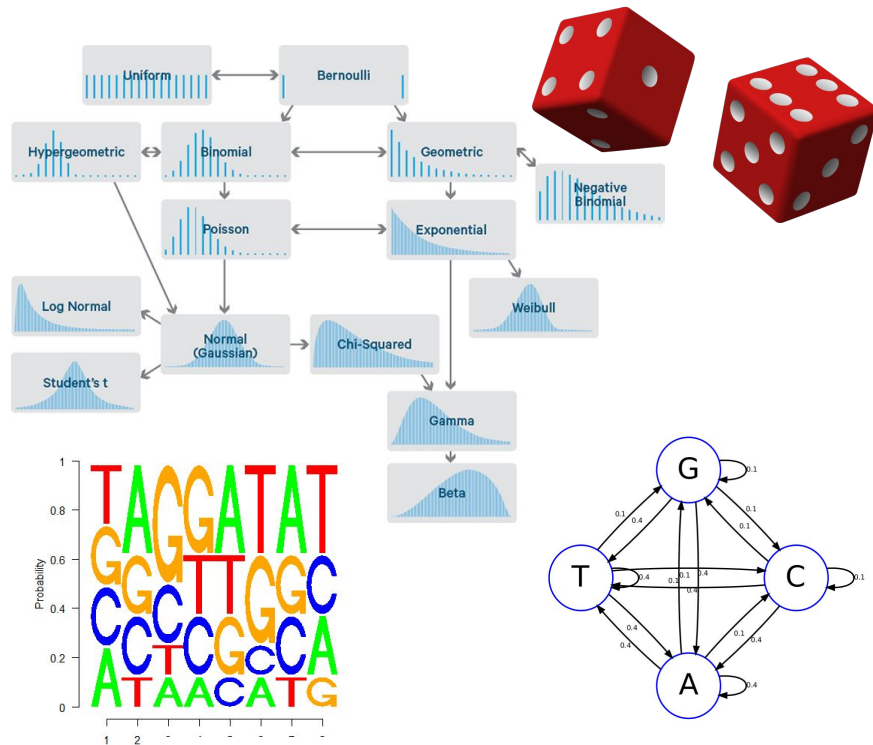


Illustration of MCMC method process (Lewis, 2011)

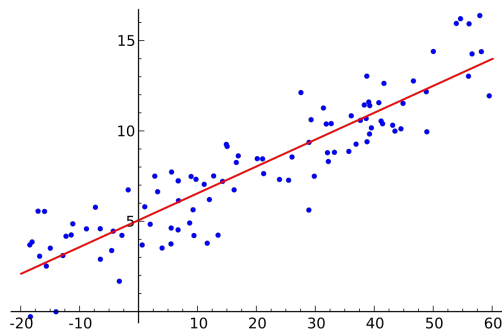
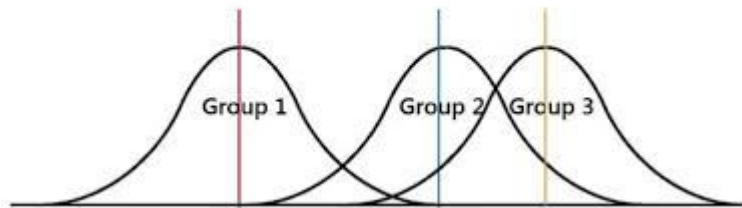
Probabilidade

Distribuições discretas e contínuas

Inferência Bayesiana

Cadeias de Markov

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct



Estatística inferencial

Teste de hipóteses

Testes não paramétricos

ANOVA

Correlação e regressão linear

Referências bibliográficas

- Seefeld, Kim e Linder, Ernst, **Statistics Using R with Biological Examples**, 2007;
- Shahbaba, Babak, **Biostatistics with R**, 2012;
- McDonald, Jonh H., **Handbook of Biological Statistics**, 2014;

Perguntas?

Tem git instalados em suas máquinas?

Verificando se git está instalado

git (<https://git-scm.com/>)

Abra um terminal e dê o seguinte comando:

```
> git help
```

Os arquivos e os slides da aula estarão em **github.com/tetsufmbio/IMD0601**. Para clonar o repositório no seu computador, dê o seguinte comando:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

Para atualizar o git, dê o seguinte comando:

```
> git pull
```



git



GitHub

Tem R e Rstudio instalados em suas máquinas?

Verificando se R e Rstudio estão instalados

R (<https://www.r-project.org/>)

Abra um terminal e dê o seguinte comando:

```
> R
```

Rstudio (<https://www.rstudio.com/products/RStudio/>)

Abra um terminal e dê o seguinte comando:

```
> rstudio
```

Não tem git, R ou Rstudio?

Baixe o instalador do **Anaconda**
(www.anaconda.com/distribution/);

Dê a permissão de execução
(`chmod +x Anaconda*`)

Execute o instalador;



Não tem git, R ou Rstudio?

Instalando git via Anaconda

```
conda install -c anaconda git
```

Instalando R e Rstudio via Anaconda

```
conda install -c anaconda r rstudio
```

