



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METRÓPOLE DIGITAL

IMD0033 – PROBABILIDADE

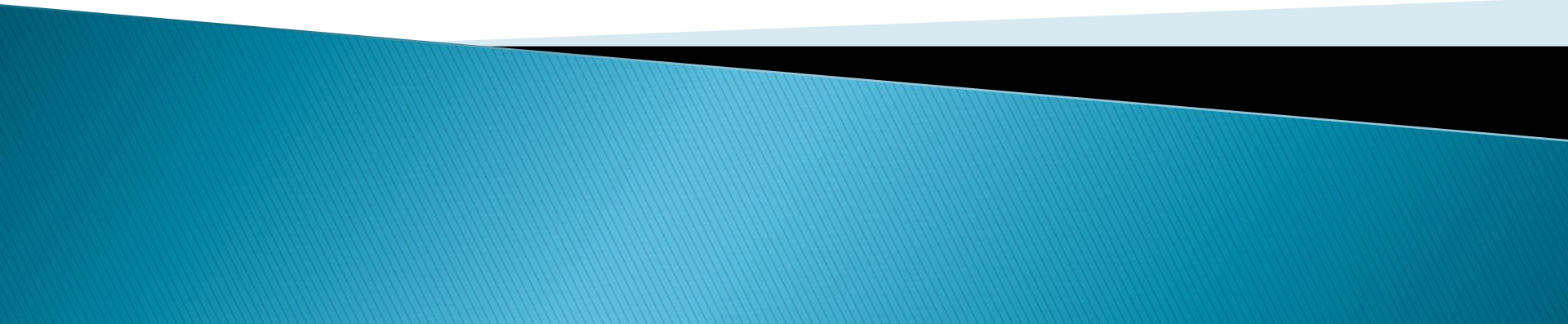
Apresentação da disciplina

Sobre mim



- ▶ Tetsu Sakamoto
 - tetsu@imd.ufrn.br
 - Sala A224

**Por que Probabilidade (e
estatística)?**

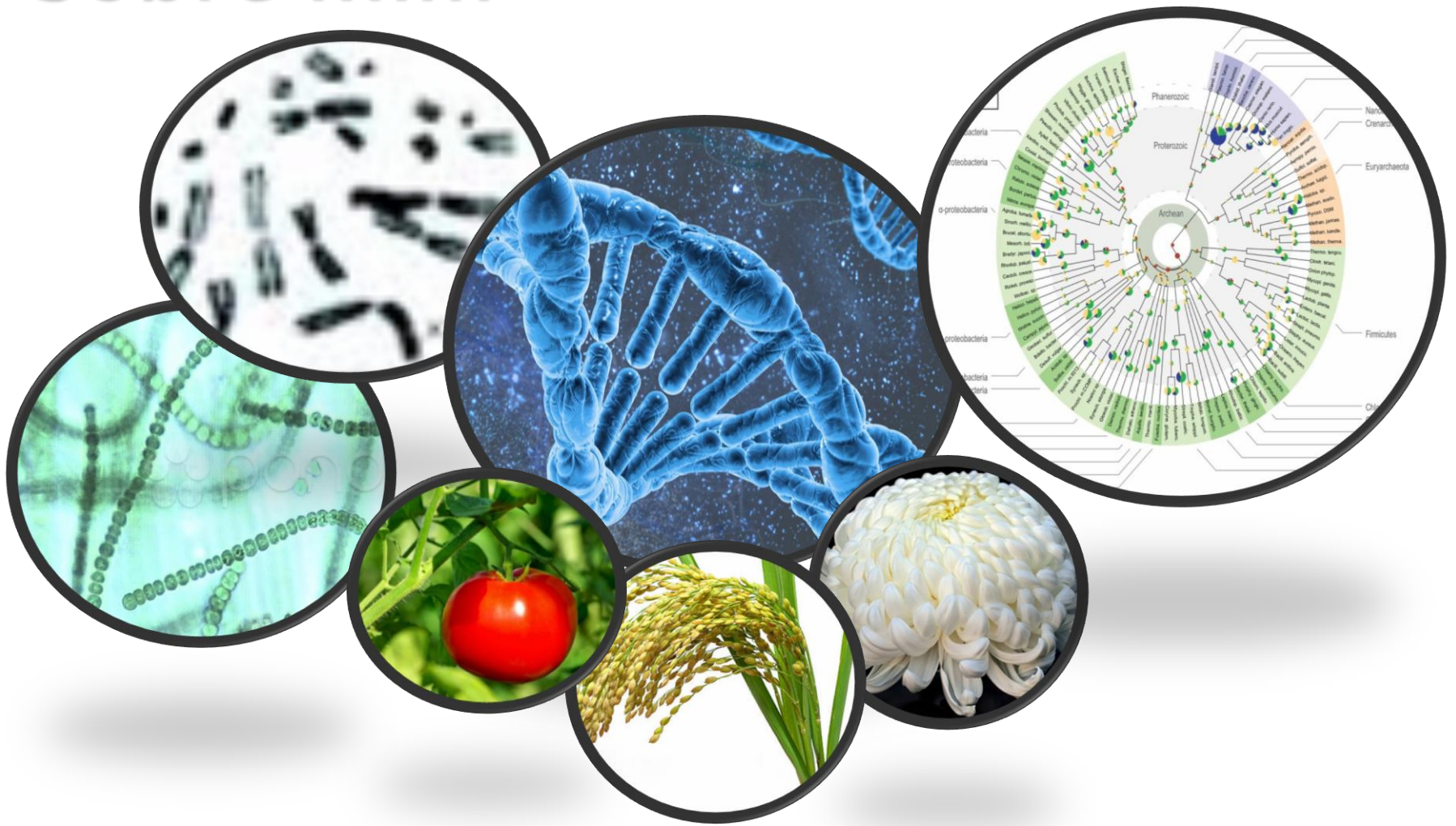


Por que Probabilidade (e Estatística)?

- ▶ Cada um vive em uma ilha de conhecimento envolto por um mar de incertezas.



Sobre mim



Por que Probabilidade (e Estatística)?

- ▶ Uma forma racional (matemático) de lidar com as incertezas;
- ▶ Cada vez mais precisos e acurados...



BIG DATA

Por que Probabilidade (e Estatística)?

- ▶ Ajuda a montar modelos que realizam:
 - Predições;
 - Tomadas de decisões;



MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

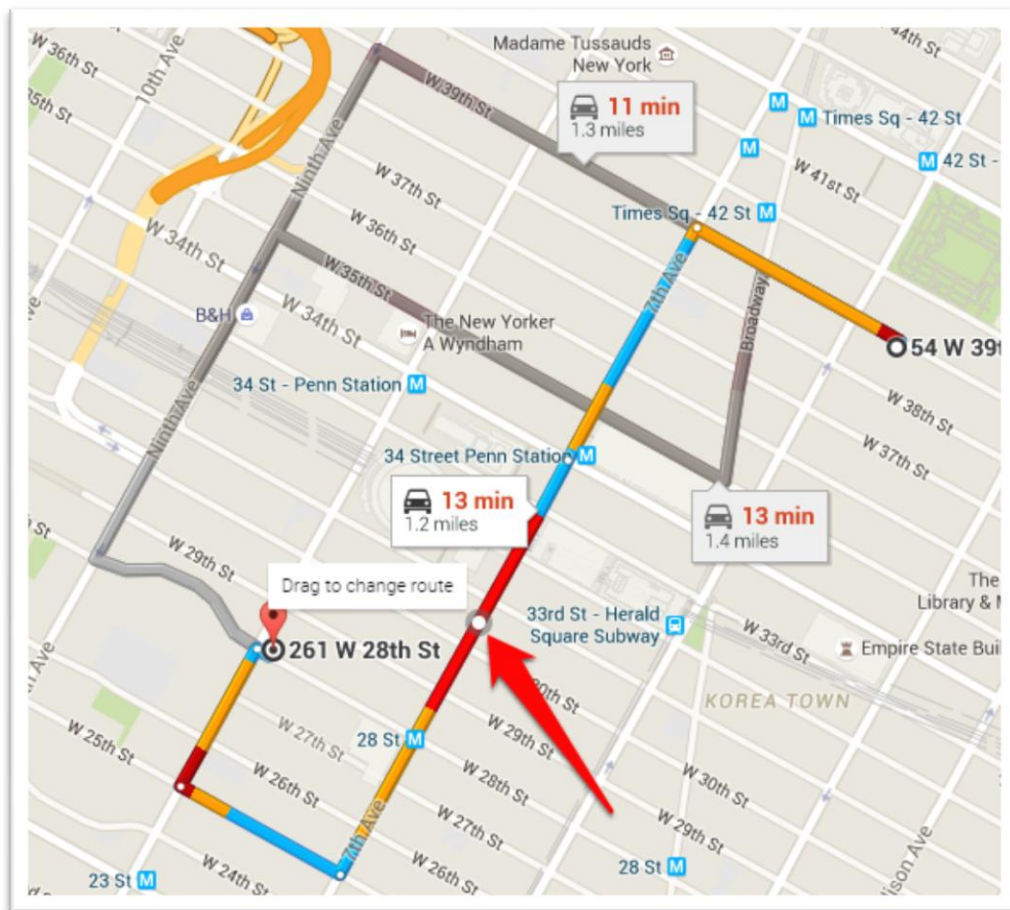
- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



Exemplos de aplicação



- ▶ **Certeza:**
 - Determinar o menor caminho;
- ▶ **Incerteza:**
 - Determinar o caminho mais rápido;

Exemplos de aplicação



- ▶ **Certeza:**
 - Após o falecimento, a família recebe um seguro;
- ▶ **Incerteza:**
 - Quando ocorrerá o falecimento?

Objetivo da disciplina

- ▶ Abordar noções básicas de probabilidade (e estatística).

Estrutura da disciplina

- ▶ Data:
 - 31 de julho a 4 de dezembro;
 - Terça e quinta (20:35h ~ 22:15h)
- ▶ Local:
- ▶ 3 avaliações;
 - Na aula seguinte a avaliação → discussão.
- ▶ Linguagem de programação: Python.

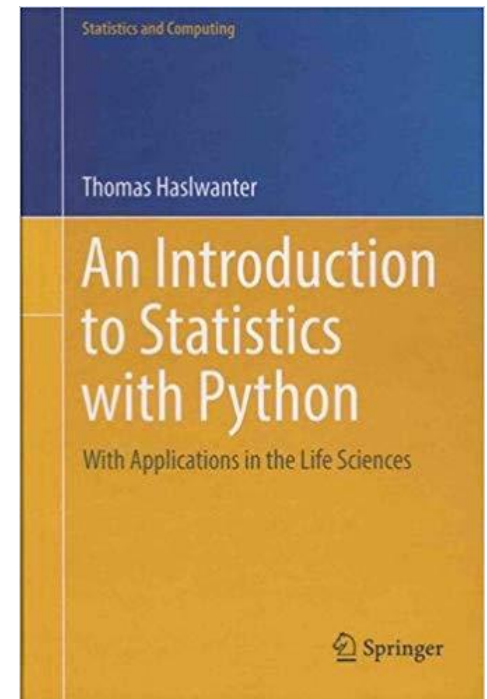
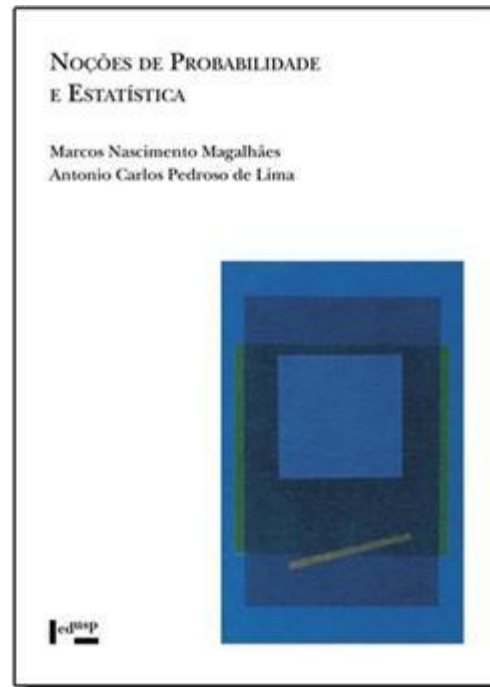
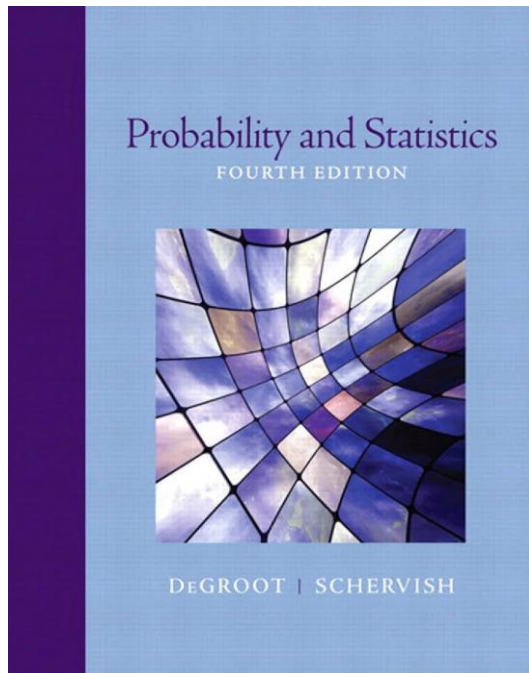
Cronograma de aula

<i>dia</i>	<i>mês</i>	<i>programação</i>
31	julho	Apresentação da disciplina
2	agosto	0. Introdução à Python
7	agosto	
9	agosto	1. Estatística Descritiva. Definição; Variáveis qualitativas e quantitativas; Medidas de tendência central, de dispersão, de achatamento e separatrizes; Representação e organização de dados.
14	agosto	
16	agosto	
21	agosto	
23	agosto	2. Probabilidade. Conceitos fundamentais; Definição de Probabilidade; Regras e teoremas básicos da probabilidade; Probabilidade do evento complementar; Teorema da soma; Probabilidade condicional; Teorema do produto; Independência probabilística.
28	agosto	
30	agosto	
4	setembro	
6	setembro	
11	setembro	
13	setembro	
18	setembro	
20	setembro	Prova 1

Cronograma de aula

<i>dia</i>	<i>mês</i>	<i>programação</i>
25	setembro	3. Variáveis aleatórias. Conceitos e definições; Variável aleatória discreta; Variável aleatória contínua
27	setembro	
2	outubro	
4	outubro	
9	outubro	
11	outubro	
16	outubro	Prova 2
18	outubro	4. Distribuições de probabilidades discretas. Bernoulli; Binomial; Geométrica; Poisson; Hipergeométrico.
23	outubro	
25	outubro	
30	outubro	
1	novembro	5. Distribuições de probabilidades contínuas. Normal; Exponencial; Qui-quadrado; t de Student.
6	novembro	
8	novembro	
13	novembro	
15	novembro	Feriado
20	novembro	Prova 3
22	novembro	discussão da prova
27	novembro	prova de reposição
29	novembro	discussão da prova de reposição
4	dezembro	consolidação

Referências



Dúvidas?

