Metodologia de Pesquisa Jurimétrica

O que estamos fazendo aqui?

Neste semestre, tentaremos responder à pergunta:

Como desenvolver um projeto de jurimetria?

Nossa jornada começa com as definições básicas de jurimetria e, através de diversos exemplos, discutiremos todos os aspectos metodológicos envolvidos.

No final, você será capaz de

- pensar um problema cientificamente,
- construir boas perguntas de pesquisa,
- planejar um estudo jurimétrico e
- desenhar o passo-a-passo para realizá-lo na prática.

Se quiser, você poderá até realizar algumas análises e programas!

VAMOS APRENDER IA??!?!!11

- Sim, mas provavelmente não da forma que você espera 😉
- Tem muito *hype* em torno da Inteligência Artificial.

- Vamos discutir o tema a partir de aplicações.

As aulas

- As aulas acontecerão às sextas, das 10h15 às 11h55.

Uso do tempo de aula (aproximado)

Horario	Atividade
10h15 - 10h30	Dúvidas e correção do quiz da aula anterior
10h30 - 11h30	Aula expositiva
11h30 - 11h55	Dúvidas e quiz

Dinâmica das aulas

- Pode me interromper para dúvidas e comentários. O plano de aula é apenas um guia, mas a aula vai para onde a gente quiser.
- Algumas dúvidas serão respondidas na hora. Outras serão respondidas mais tarde na própria aula ou em aulas futuras.

Materiais

- Os slides, quizzes etc serão disponibilizados na página do curso.
- Outros materiais poderão ser disponibilizados no **Teams**.
- Perguntas podem ser enviadas pelo Teams ou no e-mail jaztrecenti@pucsp.br

Por onde vamos?

Jurimetria não é software, é processo. Vamos aprender como isso funciona e o que devemos levar em conta.

Plano de aulas

1. Aspectos gerais da pesquisa jurimétrica

Data: 04/03/2022

Tipo: aula

Bimestre: 1

Atividades:

- Metodologia geral de pesquisa científica
- Formulação de hipóteses
- Fontes de dados
- Estudo prospectivo e retrospectivo

Obs: aula 1, não tem quiz

2. Estatística básica

Data: 11/03/2022

Tipo: aula

Bimestre: 1

Atividades:

- Base de dados arrumada
- Tipos de variáveis
- Estatísticas de posição
- Estatísticas de variabilidade
- Quiz

3. Introdução à programação em R

Data: 18/03/2022

Tipo: aula

Bimestre: 1

Atividades:

- Vantagens de usar programação
- R como calculadora
- Tipos de objetos
- Estatísticas de posição e variabilidade no R
- Quiz

4. Introdução à programação em R

Data: 25/03/2022

Tipo: aula

Bimestre: 1

Atividades:

- Vantagens de usar programação
- R como calculadora
- Tipos de objetos
- Estatísticas de posição e variabilidade no R
- Quiz

5. Pesquisa jurimétrica utilizando dados da jurisprudência

Data: 01/04/2022

Tipo: aula

Bimestre: 1

Atividades:

- Consulta de jurisprudência do TJSP
- Análise descritiva
- Exemplo: boletim
- Quiz

6. Pesquisa jurimétrica utilizando classificação manual

Data: 08/04/2022

Tipo: aula

Bimestre: 1

Atividades:

- Formulário de classificação
- Controle de inconsistências
- Exemplo: observatório das falências
- Quiz

7. Prova 1

Data: 29/04/2022

Tipo: prova

Bimestre: 1

Atividades:

-

8. Probabilidade e inferência

Data: 06/05/2022

Tipo: aula

Bimestre: 1

Atividades:

- O que é probabilidade
- O que é inferência estatística
- Viés e variância
- Amostragem
- Quiz

9. Regressão linear

Data: 13/05/2022

Tipo: aula

Bimestre: 2

Atividades:

- Modelos de regressão
- Variável resposta e explicativa
- Intercepto e coeficiente angular
- Qualidade do ajuste
- Quiz

10. Pesquisa jurimétrica utilizando dados agregados do judiciário

Data: 20/05/2022

Tipo: aula

Bimestre: 2

Atividades:

- Justiça em números
- Módulo de produtividade mensal
- Datajud
- Exemplo: litigiosidade e desenvolvimento
- Quiz

11. Modelagem preditiva

Data: 27/05/2022

Tipo: aula

Bimestre: 2

Atividades:

- Duas culturas
- IA no direito
- Sobreajuste
- Treino/teste
- Quiz

12. Tópicos de pesquisa jurimétrica

Data: 03/06/2022

Tipo: aula

Bimestre: 2

Atividades:

- Pesquisa jurimétrica utilizando diários de justiça
- Pesquisa jurimétrica usando sistemas administrativos e dados do executivo
- Pesquisa jurimétrica usando dados de leis
- Pesquisa jurimétrica utilizando inteligência artificial
- Quiz

13. Prova 2

Data: 10/06/2022

Tipo: prova

Bimestre: 2

Atividades:

-

14. Sub

Data: 17/06/2022

Tipo: prova

Bimestre: 2

Atividades:

-

Obs: emenda de feriado

15. (extra) exemplo de projeto no R

Data: 24/06/2022

Tipo: aula

Bimestre: 2

Atividades:

Modelagem preditiva e inteligência artificial

16. Exame

Data: 01/07/2022

Tipo: prova

Bimestre: 2

Atividades:

-

Notas

Para cada bimestre:

- **Quiz do final da aula**: 50%
    - Cada quiz vale 1,0 na nota do bimestre.
    - Cada quiz terá ~5 questões de múltipla escolha.
    - Poderemos ter questões extras, valendo **pontos extras**.
    - Os quizes podem ser feitos individualmente ou em grupo, mas a entrega é individual.
    - O prazo de cada quiz é de 1 semana: 23h59 da quinta-feira anterior à aula.

- **Prova**: 50%
    - Cada prova vale 5,0 na nota do bimestre
    - A prova poderá ter questões extras, valendo pontos extras.
    - Dependendo do andar das aulas, poderemos substituir alguma prova por um trabalho em grupo.

Guia para os slides

Slides sobre cases: **marca azul**

Slides sobre conceitos: **marca rosa**

Slides sobre metodologia: **marca verde**

Aspectos gerais da pesquisa jurimétrica

O que é jurimetria?

- Jurimetria é a **estatística** aplicada ao **direito**.
- Fazer jurimetria é pensar o direito de forma **concreta**, colocando nosso objeto de estudo no **tempo** e no **espaço**.

Estudo clássico

Princípios jurídicos e pensamento de especialistas sobre o valor de indenização por dano moral a ser concedido no contexto de direito do consumidor.

Estudo jurimétrico

Valores típicos de dano moral em sentenças de primeira instância de processos envolvendo direito do consumidor no Tribunal de Justiça de São Paulo em 2021.

A jurimetria **complementa** o estudo clássico do direito, considerando leis, **princípios** jurídicos e pensamentos abstratos como **hipóteses** que se manifestam ou não na **realidade** através das **decisões** judiciais.

Definição de hipóteses

Definição de hipóteses

Exemplo (em sala)

Ciclo da ciência de dados

Sabemos onde estão os dados. O que fazer?

Nosso dia-a-dia

Download

Arrumação

Visualização / Modelagem

Resumo

Conceitos

- Pensar jurimetria é pensar de forma **concreta**
- A jurimetria é complementar ao estudo clássico do direito, e tem limites
- Para fazer jurimetria, precisamos de dados!

Metodologia

- Antes de realizar a coleta, precisamos definir nossas perguntas de pesquisa. Entender a teoria é muito importante!
- Os parâmetros da coleta precisam ser bem especificados.
- Algumas perguntas podem surgir durante a análise dos dados.

Estatística básica

O que é estatística?

- Conjunto de ferramentas e técnicas para tratar dados.
- Fazemos estatísticas para extrair informações sumarizadas de dados.
- Também pode ser pensada como a *metaciência*, ou seja, ela organiza a forma que a ciência funciona.

Para que serve estatística?

- Usamos estatística para fazer afirmações sobre o desconhecido.
- Detectamos *tendências* para fazer previsões.
- Organizamos processos complexos de decisão.
- Aprendemos a controlar a incerteza que existe no mundo.

Perguntas respondidas com estatística

- **Advogando**: quanto cobrar por esse processo?
- **No tribunal**: o estoque de processos está caindo?
- **Na minha empresa**: vale a pena fazer acordo?
- **No legislativo**: essa lei aumentou a litigiosidade?

Perguntas não respondidas com estatística*

- **Advogando**: qual tese aplicar nesse caso?
- **No tribunal**: quais temas do direito devem ser especializados?
- **Na minha empresa**: como prevenir novos conflitos?
- **No legislativo**: essa lei é justa?

O que é um modelo estatístico?

- É uma redução da realidade a partir de uma fórmula matemática.
- Em outras palavras, é a formalização de uma teoria.
- A fórmula é uma representação do *mecanismo de geração dos dados*.

Exemplos:

- Leis da física (gravidade, por exemplo)
- Distribuição das alturas da turma

O que é um modelo estatístico?

- Fórmulas matemáticas sempre apresentam erros
- A estatística assume que esses erros existem e colocam isso nas fórmulas, usando *distribuições de probabilidades*
- Por isso, aprender estatística envolve também aprender sobre teoria de probabilidades

Medidas de posição

Média

- É a soma dos valores dividida pela quantidade de valores
- Está presente em tudo: desde as notas da faculdade até as redes neurais

Mediana

- São úteis quando temos dados assimétricos
- Exemplo da Oprah

Limitações

- Conjuntos de dados muito diferentes levam a estatísticas de posição iguais.
- Exemplo das notas

Medidas de variabilidade

Variância e desvio padrão

- **Variância**: É a soma dos valores em relação à média (ao quadrado), dividida pela quantidade de observações

- **Desvio padrão**: É a raiz quadrada da variância. A vantagem do desvio padrão é que está na mesma escala dos dados. Por exemplo, se temos uma variável em reais, o desvio padrão está em reais.

Programação

Conteúdo

Introdução ao R
R
Manipulação de dados
tidyr
dplyr
ggplot2
Visualização de dados
haven
readr
readxl
Importação de dados
knitr
rmarkdown
Comunicação

Introdução ao R

- R e RStudio
- Trabalhando com projetos
- Objetos e funções
- Pacotes

Importação

- Arquivos de texto
- Arquivos do excel
- Arquivos de outros softwares (SAS, SPSS)
- Conexão com banco de dados

Ilustração por Allison Horst
clique aqui para mais informações

Transformação

- Selecionar colunas
- Filtrar linhas
- Criar ou modificar colunas
- Agrupar e sumarizar
- Juntar tabelas

Ilustração por Allison Horst
clique aqui para mais informações

Visualização

Comunicação

- Relatórios automatizados
- Dashboards estáticos

Referência

As vantagens de analisar dados usando linguagens de programação

As principais linguagens para ciência de dados são

Código Aberto

O que significa ser código aberto?

1. Acesso gratuito.
2. Estudantes podem usar as mesmas ferramentas que os profissionais.
3. Possibilita a existência de uma comunidade ativa.
4. Você pode corrigir problemas e aprimorar a linguagem.
5. Você pode desenvovler suas próprias ferramentas.

O código é texto

Podemos copiar e colar

Reprodutibilidade

O código é uma linguagem

Podemos documentar a nossa análise

```
image_read("https://jeroen.github.io/images/frink.png") |>
  image_rotate(270) |>
  image_background("blue", flatten = TRUE) |>
  image_border("red", "10x10") |>
  image_annotate("Linguagens de programação ainda\nsão linguagens!",
                 color = "white", size = 30)
```

Entrada

Saída

A relevância do R

O ciclo da ciência de dados

O ciclo da ciência de dados com o R

Sugestões

Pratique

O objetivo de hoje é ajudar a quebrar inércia da etapa inicial de aprendizado. Assim como um idioma, o único jeito de aprender uma linguagem de programação é praticando.

Pedindo ajuda

- Documentação do R
- Discourse Curso-r
- Google
- Stack Overflow

Regras, boas práticas e estilo

Na grande maioria das vezes, você poderá resolver uma tarefa de diversas maneiras. Isso porque, apesar de existirem várias **regras**, linguagens de programação também possuem **boas práticas** e **estilos**.

- **Regras**: devem ser seguidas para o código funcionar (sintaxe, vocabulário)
- **Boas práticas**: recomenda-se seguir para a criação de códigos legíveis (espaçamento, nomes, organização)
- **Estilo**: cada um pode escolher aquele com o qual se sentir mais à vontade (tipos de indentação, formatação)

Pesquisa jurimétrica usando dados de jurisprudência

(estudos retrospectivos)

Objetivos de aprendizagem

- Entender sobre estudos prospectivos e retrospectivos
- Entender quando realizar uma análise de jurisprudência
- Fazer uma análise de jurisprudência live coding

Fontes de dados

- **Diários de Justiça**: PDFs gigantescos contendo toda a movimentação judiciária. Não é fácil lista processos a partir desses arquivos. É útil para estudos prospectivos.
- **Consulta de jurisprudência**: Ferramentas de pesquisa dos tribunais. É mais fácil de consultar, mas ficamos reféns da qualidade da ferramenta de pesquisa.
- **DataJud**: No futuro, teremos uma base de dados acessível e completa disponibilizada pelo CNJ (Resolução 331/2020).

Estudos prospectivos e retrospectivos

- Estudos **prospectivos** acompanham processos desde seu **nascimento**, identificado pela data de distribuição, até sua morte, identificada pela sentença, trânsito em julgado, arquivamento, etc.
    - Importante em análise de tempos.
    - Análise de sobrevivência.
- Já estudos **retrospectivos** são aqueles que indexam os processos por alguma data de **desfecho**, como uma sentença, um acórdão ou a data de arquivamento.
    - Importante para análise de decisões.
    - Modelos preditivos.

Estudo prospectivo e retrospectivo

Estrutura de dados

- **Capa**: contém classe, assunto, vara, comarca, status, indicador de processo digital, local físico, entre outras.
- **Partes**: contém nomes dos autores e dos réus, advogado e tipo de participação.
- **Movimentações**: contém datas, títulos e conteúdo de todas as movimentações públicas do processo. São movimentações desde despachos simples, remessas e conclusos até ata de audiências e sentenças completas.
- **Pasta digital**: contém todos os autos do processo, incluindo petição inicial, contestação e outros documentos não publicáveis. Os documentos só estão disponíveis em processos digitais, com acesso somente a advogados e pesquisadores credenciados.

Tempos em estudos retrospectivos

O que observamos O que é de verdade

Análise de tempos

- Quando o estudo é prospectivo, utilizamos uma técnica chamada **análise de sobrevivência**.
- A técnica é capaz de lidar corretamente com casos que ainda estão ativos no momento da análise. Denominamos esses casos como **censura**.
- A censura é capaz de lidar com informações incompletas. No nosso caso, sabemos que um processo ativo nascido em janeiro de 2021 durou **pelo menos 7 meses** até agora.

Exemplo!

Apoio

Núcleo de Direito, Tecnologia e Jurimetria / PUC-SP

Associação Brasileira de Jurimetria