

INTRODUÇÃO À JURIMETRIA

The background is a light gray with several abstract elements. On the left, there are wavy, horizontal lines. In the center, there is a faint outline of a human brain. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). On the far right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board.

METODOLOGIA: TÉCNICAS DE APLICAÇÃO

Considerações iniciais



- A metodologia que vai ser apresentada é uma metodologia natural, no sentido de que deve aparecer naturalmente no dia-a-dia de uma **análise externa**.
- A separação das etapas da metodologia é artificial, pois na vida real essas etapas possuem intersecções entre si.
- Ainda sim, faz sentido apresentar as etapas da metodologia de forma separada, pois analiticamente isso nos ajuda a compreender os problemas metodológicos que vamos enfrentar

Análises Externas vs Análises Internas

Análise Interna

- **O que é:** Análise realizada com base em um banco de dados interno.
- **Vantagens:** Informações estruturadas, presença informações sobre resultados e estratégias
- **Desvantagens:** Viés, migrações de sistema, estratégia de armazenamento de dados

Análise Externa

- **O que é:** Análises realizadas com bases de dados públicas, como as dos tribunais.
- **Vantagens:** Informações confiáveis, possibilidade de analisar grandes volumes de informações
- **Desvantagens:** Diferentes sistemas, ausência de informações relevantes, acesso às informações etc

O CAMINHO DE UMA PESQUISA

The background is a light gray with several abstract elements. On the left, there are wavy, horizontal lines. In the center, there is a faint outline of a human brain. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). On the far right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board.

O caminho de uma pesquisa



1. Definição dos objetivos

- 1.1. Análise Confirmatória vs Análise Exploratória
- 1.2. Tradução: Linguagem jurídica -> linguagem analítica

2. Listagem dos processos

- 2.1. Definição do escopo temporal, temático e regional
 - 2.1.1. TPUs e cifra oculta

3. Extração e estruturação dos dados

- 3.1. Extração automática
- 3.2. Extração manual

4. Tratamento dos dados

- 4.1. Controle de inconsistências

O caminho de uma pesquisa



5. Análise Descritiva

5.1. Visualização analítica (gráficos e tabelas)

5.2. Criação de hipóteses

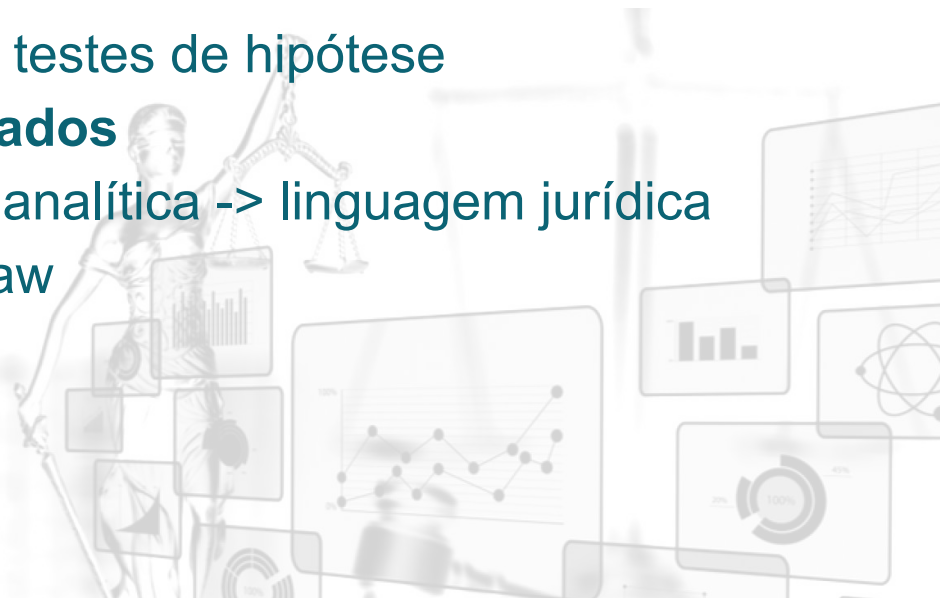
6. Análise Inferencial

6.1. Modelos estatísticos e testes de hipótese

7. Comunicação dos resultados

7.1. Tradução: Linguagem analítica -> linguagem jurídica

7.2. Legal Design/Visual Law



O ciclo da ciência de dados



Importar



Arrumar

(Armazenar os dados
consistentemente)



Transformar

(Criar novas variáveis e
agregações)

Visualizar

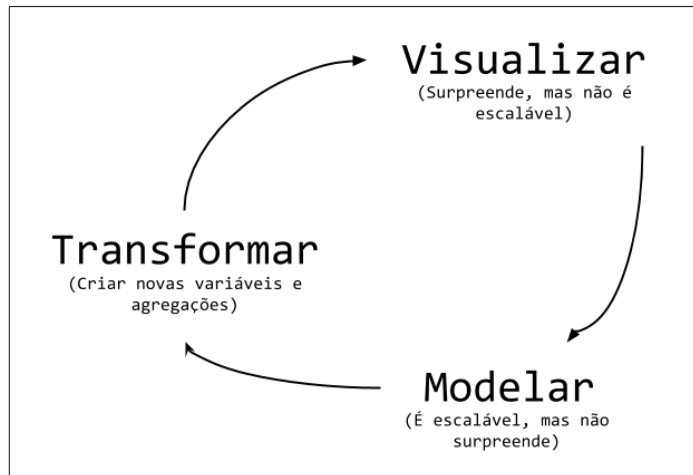
(Surpreende, mas não é
escalável)

Modelar

(É escalável, mas não
surpreende)

Comunicar

Automatizar



Definição dos objetivos



- **Input:** Hipótese, ideia, impressão, interesse...
- **Output:** Perguntas e questões-chave
- **Objetivo:** Planejamento objetivo da análise
- **Métodos e técnicas:** Experiência/subjetividade da parte interessada no estudo
- **Problema:** Análise exploratória

Listagem de processos



- **Input:** Escopo temporal, regional e temático
- **Output:** Lista com os identificadores dos elementos de estudo (contratos, processos etc)
- **Objetivo:** Delimitar todos os elementos que estarão dentro do escopo da pesquisa
- **Métodos e técnicas:** Web Scrapping, APIs, Informações internas...
- **Problema:** Viés, cifra oculta, acesso à informação de interesse, restrições de sistemas etc

Extração de dados



- **Input:** Uma lista de objetos (processos, contratos etc)
- **Output:** Base de dados bruta
- **Objetivo:** Coletar os dados que serão relevantes para a análise
- **Métodos e técnicas:** Web scrapping, classificação manual, regex, NLP...
- **Problema:** Sites não raspáveis, reCAPTCHA, segredo de justiça, criatividade humana

Tratamento dos dados



- **Input:** Uma base de dados bruta
- **Output:** Base de dados tratada
- **Objetivo:** Analisar a qualidade dos dados e eventuais limitações técnicas da análise.
- **Métodos e técnicas:** Algoritmos de detecção de erros lógicos, de padronização de categorias...
- **Problema:** Criatividade humana, erros não inconsistentes

Análise Descritiva



- **Input:** Uma base de dados tratada
- **Output:** Resumo analítico da base de dados
- **Objetivo:** Estudar a distribuição dos dados, avaliar a existência de padrões e tendências, comunicação...
- **Métodos e técnicas:** Criação de gráficos, tabelas, resumos analíticos
- **Problema:** DCN (Detecção de Coisa Nenhuma) - não existir relação entre as informações

Análise Inferencial



- **Input:** Uma base de dados tratada + Conjunto de hipóteses e/ou previsões de interesse
- **Output:** Hipóteses testadas (análise de significância), modelos estatísticos...
- **Objetivo:** Demonstrar a significância estatística de uma hipótese, desenvolver modelos preditivos etc
- **Métodos e técnicas:** Machine Learning, Deep Learning, IA
- **Problema:** Não existir relação entre as informações/modelos de baixa performance

Comunicação dos resultados



- **Input:** Relatório/Resumo da análise, com metodologia, resultados e conclusões do estudo
- **Output:** Material de divulgação e comunicação dos resultados para o público alvo
- **Objetivo:** Apresentar os resultados da análise em uma linguagem amigável, focada no usuário final
- **Métodos e técnicas:** Legal Design, Visual Law...
- **Problema:** Documentação da análise/planejamento

Pausa



Antes de prosseguirmos com as definições de cada etapa de uma análise, vamos responder à seguinte pergunta:



COMO O DIREITO AJUDA A ESTATÍSTICA?

Preliminares: Regras básicas



Armazenamento das informações para fins analíticos:

- Cada **linha** é uma observação única (**unidade amostral**)
 - Uma observação é única SE a unidade amostral dela não se repetir em mais nenhuma linha da base de dados.
- Cada **coluna** é uma informação/característica sobre essa observação (**unidade observacional**)

Preliminares: Caso prático



- Olhe este processo na consulta de processos do esaj do TJSP: 1054460-52.2016.8.26.0100
- Vamos desenhar o formato disso juntos, seguindo as regras básicas

e-SAJ Consulta de Processos do 1ºGrau				
←				
1054460-52.2016.8.26.0100	Extinto			
Classe Recuperação Judicial	Assunto Recuperação judicial e Falência	Foro Foro Central Cível	Vara 2ª Vara de Falências e...	Juiz PAULO FURTADO DE OLIVEIRA FILHO
Distribuição 31/05/2016 às 11:37 - Direcionada	Controle 2016/000380	Área Cível	Valor da ação R\$ 100.000,00	Outros assuntos Empresas

Resultado esperado



	id_processo	status	classe	assunto
1	10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência
2	10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência
3	00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência
4	10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência
5	10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência
6	00237449820128260100	Extinto	Recuperação Judicial	Administração judicial
7	00187684820128260100	Extinto	Recuperação Judicial	Administração judicial
8	10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência
9	00323077120128260071	Suspenso	Recuperação Judicial	Empresas
10	00010360720138260363	NA	Recuperação Judicial	Empresas

Calma, as coisas podem complicar...

e-SAJ | Consulta de Processos do 1º Grau



1054460-52.2016.8.26.0100 **Extinto**

Classe

Recuperação Judicial

Assunto

Recuperação judicial e Falência

Foro

Foro Central Cível

Vara

2ª Vara de
Falências e...

Juiz

PAULO FURTADO DE OLIVEIRA
FILHO

Distribuição

31/05/2016 às 11:37 -
Direcionada

Controle

2016/000380

Área

Cível

Valor da ação

R\$ 100.000,00

Outros assuntos

Empresas

PARTES DO PROCESSO

Reqte

Supermercado Novo Capão Ltda

Advogado: João Apolinario da Silva Filho

Calma, as coisas podem complicar...



id_processo	nome	parte	papel
10544605220168260100	Rachelle Bustamante	Reqte	Reqte
10544605220168260100	Zuraara el-Bilal	Reqte	Advogado
10046286820168260482	Armando Ponce-Campa	Reqte	Reqte
10046286820168260482	Samantha Hale	Reqte	Advogado
10046286820168260482	Dalton Durst	AdmTerc	AdmTerc
10046286820168260482	Raeven West	AdmTerc	Advogado
10046286820168260482	Desiree Jefferson	Interesdo	Interesdo
10046286820168260482	Michael Hogan	Interesdo	Advogado
10046286820168260482	Dylan Long	Perito	Perito
10046286820168260482	Nicole Varnet	Advogada	Advogada
00037451620128260568	Chris Zhong	Reqte	Reqte
00037451620128260568	Jacob Weiss	Reqte	Advogada

Nova variável: Número de partes por processo



id_processo	nome	parte	papel
10544605220168260100	Rachelle Bustamante	Reqte	Reqte
10544605220168260100	Zuraara el-Bilal	Reqte	Advogado
10046286820168260482	Armando Ponce-Campa	Reqte	Reqte
10046286820168260482	Samantha Hale	Reqte	Advogado
10046286820168260482	Dalton Durst	AdmTerc	AdmTerc
10046286820168260482	Raeven West	AdmTerc	Advogado
10046286820168260482	Desiree Jefferson	Interesdo	Interesdo
10046286820168260482	Michael Hogan	Interesdo	Advogado
10046286820168260482	Dylan Long	Perito	Perito
10046286820168260482	Nicole Varnet	Advogada	Advogada
00037451620128260568	Chris Zhong	Reqte	Reqte
00037451620128260568	Jacob Weiss	Reqte	Advogada

n_partes = 2

n_partes = 8

n_partes = 2

Nova variável: Número de partes por processo



	id_processo	status	classe	assunto
1	10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência
2	10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência
3	00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência
4	10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência
5	10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência
6	00237449820128260100	Extinto	Recuperação Judicial	Administração judicial
7	00187684820128260100	Extinto	Recuperação Judicial	Administração judicial
8	10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência
9	00323077120128260071	Suspenso	Recuperação Judicial	Empresas
10	00010360720138260363	NA	Recuperação Judicial	Empresas

Nova variável: Número de partes por processo



id_processo	status	classe	assunto	n_partes
00010360720138260363	NA	Recuperação Judicial	Empresas	8
00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência	10
00187684820128260100	Extinto	Recuperação Judicial	Administração judicial	7
00237449820128260100	Extinto	Recuperação Judicial	Administração judicial	3
00323077120128260071	Suspenso	Recuperação Judicial	Empresas	9
10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência	2
10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência	4
10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência	8
10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência	8
10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência	2



OK, MAS COMO O DIREITO AJUDA A ESTATÍSTICA?

Entendendo os dados de movimentações

e-SAJ | Consulta de Processos do 1º Grau



1054460-52.2016.8.26.0100

Extinto

Classe

Recuperação Judicial

Assunto

Recuperação judicial e Falência

Foro

Foro Central Cível

Vara

2ª Vara de
Falências e...

Juiz

PAULO FURTADO DE OLIVEIRA
FILHO

MOVIMENTAÇÕES

Data

Movimento

06/09/2016

Arquivado Definitivamente

06/09/2016

Trânsito em Julgado às partes - com Baixa

06/09/2016



Certidão de Trânsito em Julgado com Baixa Expedida
Certidão - Trânsito em Julgado com Baixa - Processo Digital

05/09/2016

Sentença Registrada

11/08/2016

Certidão de Publicação Expedida

Relação :0240/2016 Data da Disponibilização: 11/08/2016 Data da Publicação: 12/08/2016 Número do Diário: 2177 Página: 770/786

Entendendo os dados de movimentações

id_processo	id_mov	data	movimento	descricao	tem_anexo
10544605220168260100	1	2016-05-31	Decisão	Vistos. 1 Indefiro a concessão do benefício de justiça gratuit...	TRUE
10544605220168260100	2	2016-05-31	Conclusos para Decisão	NA	FALSE
10544605220168260100	3	2016-05-31	Distribuído por Direcionamento (movimentação exclusiva d...	Há suspeita de repetição da ação. Confronte os dados do pr...	FALSE
10544605220168260100	4	2016-06-01	Remetido ao DJE	Relação: 0168/2016 Teor do ato: Vistos. 1 Indefiro a concess...	FALSE
10544605220168260100	5	2016-06-02	Certidão de Publicação Expedida	Relação :0168/2016 Data da Disponibilização: 02/06/2016 D...	FALSE
10544605220168260100	6	2016-06-28	Pedido de Extinção Juntada	Nº Protocolo: WJMJ.16.40565843-0 Tipo da Petição: Pedido ...	FALSE
10544605220168260100	7	2016-08-01	Conclusos para Sentença	NA	FALSE
10544605220168260100	8	2016-08-05	Extinto o Processo sem Resolução do Mérito por Desistência	Vistos. Homologo, para que produza seus efeitos jurídicos e...	TRUE
10544605220168260100	9	2016-08-10	Remetido ao DJE	Relação: 0240/2016 Teor do ato: Vistos. Homologo, para qu...	FALSE
10544605220168260100	10	2016-08-11	Certidão de Publicação Expedida	Relação :0240/2016 Data da Disponibilização: 11/08/2016 D...	FALSE
10544605220168260100	11	2016-09-05	Sentença Registrada	NA	FALSE
10544605220168260100	12	2016-09-06	Arquivado Definitivamente	NA	FALSE
10544605220168260100	13	2016-09-06	Trânsito em Julgado às partes - com Baixa	NA	FALSE
10544605220168260100	14	2016-09-06	Certidão de Trânsito em Julgado com Baixa Expedida	Certidão - Trânsito em Julgado com Baixa - Processo Digital	TRUE

Um estatístico quer extrair dessas informações de movimentações do processo 105446 uma única coluna chamada “data_sentenca”. Ajude-o

Entendendo os dados de movimentações

id_processo	id_mov	data	movimento	descricao	tem_anexo
10544605220168260100	1	2016-05-31	Decisão	Vistos. 1 Indefiro a concessão do benefício de justiça gratuit...	TRUE
10544605220168260100	2	2016-05-31	Conclusos para Decisão	NA	FALSE
10544605220168260100	3	2016-05-31	Distribuído por Direcionamento (movimentação exclusiva d...	Há suspeita de repetição da ação. Confronte os dados do pr...	FALSE
10544605220168260100	4	2016-06-01	Remetido ao DJE	Relação: 0168/2016 Teor do ato: Vistos. 1 Indefiro a concess...	FALSE
10544605220168260100	5	2016-06-02	Certidão de Publicação Expedida	Relação :0168/2016 Data da Disponibilização: 02/06/2016 D...	FALSE
10544605220168260100	6	2016-06-28	Pedido de Extinção Juntada	Nº Protocolo: WJMJ.16.40565843-0 Tipo da Petição: Pedido ...	FALSE
10544605220168260100	7	2016-08-01	Conclusos para Sentença	NA	FALSE
10544605220168260100	8	2016-08-05	Extinto o Processo sem Resolução do Mérito por Desistência	Vistos. Homologo, para que produza seus efeitos jurídicos e...	TRUE
10544605220168260100	9	2016-08-10	Remetido ao DJE	Relação: 0240/2016 Teor do ato: Vistos. Homologo, para qu...	FALSE
10544605220168260100	10	2016-08-11	Certidão de Publicação Expedida	Relação :0240/2016 Data da Disponibilização: 11/08/2016 D...	FALSE
10544605220168260100	11	2016-09-05	Sentença Registrada	NA	FALSE
10544605220168260100	12	2016-09-06	Arquivado Definitivamente	NA	FALSE
10544605220168260100	13	2016-09-06	Trânsito em Julgado às partes - com Baixa	NA	FALSE
10544605220168260100	14	2016-09-06	Certidão de Trânsito em Julgado com Baixa Expedida	Certidão - Trânsito em Julgado com Baixa - Processo Digital	TRUE

O problema aqui é que um estatístico não sabe a diferença entre uma sentença e uma decisão interlocutória!

Entendendo os dados de movimentações



Movimento
15/03/2018 19:47:27 - Concedida a Medida Liminar
24/10/2017 13:21:57 - Conclusos para decisão
23/10/2017 08:16:50 - Juntada de Petição
12/09/2017 12:17:28 - Processo Suspenso/Sobrestado
12/09/2017 00:04:40 - Juntada de Certidão de decurso de prazo
01/09/2017 08:28:30 - Juntada de Certidão de Intimação
01/09/2017 08:27:50 - Juntada de Certidão de Intimação
31/08/2017 09:41:17 - Juntada de Certidão de Intimação
29/08/2017 17:18:40 - Expedição de expediente
29/08/2017 16:30:55 - Processo Suspenso ou sobrestado por recurso especial repetitivo
27/06/2017 12:41:01 - Conclusos para julgamento
10/06/2017 00:00:21 - Juntada de Certidão de Intimação
02/06/2017 08:43:26 - Juntada de Certidão de Intimação
02/06/2017 08:02:28 - Juntada de Certidão de Intimação
30/05/2017 17:54:01 - Não Concedida a Medida Liminar
29/05/2017 11:52:55 - Conclusos para decisão
11/05/2017 16:58:30 - Juntada de Petição
07/05/2017 00:01:04 - Juntada de Certidão de Intimação
03/05/2017 16:54:36 - Juntada de Contestação
03/05/2017 16:39:07 - Juntada de Certidão de Intimação
03/05/2017 09:03:15 - Juntada de Certidão de Intimação
28/04/2017 08:06:54 - Juntada de Certidão de Intimação
28/04/2017 08:06:53 - Juntada de Certidão de Intimação
27/04/2017 14:37:49 - Expedição de expediente
27/04/2017 14:36:55 - Expedição de expediente
27/04/2017 12:34:26 - Proferido despacho de mero expediente
24/04/2017 16:37:44 - Conclusos para despacho
24/04/2017 16:37:44 - Distribuído por Sorteio para 6ª VARA FEDERAL - Substituto

De onde extrair os dados?



- João, um Estatístico, precisa analisar os processos de judicialização da saúde relacionados ao fornecimento de medicamentos que não possuem registro na Anvisa, dentro do Distrito Federal
- Para tanto, ele busca no Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) por todos os processos de primeiro grau com o Assunto “Fornecimento de Medicamentos”.

João tomou uma decisão acertada?

De onde extrair os dados?



Medicamento sem registro na Anvisa – remessa dos autos à Justiça Federal

“2. Nos termos do tema fixado em sede de repercussão geral (Embargos de Declaração no RE 855178 - Tema 793), no âmbito do Supremo Tribunal Federal, é necessária a formação de litisconsórcio passivo necessário, a demonstrar a necessidade de proposição da ação perante a União, **atraindo a competência da Justiça Federal, nos casos de demanda de fornecimento de medicamentos que não possuem registro na Anvisa.**

3. Diante da necessidade de integrar a lide a União Federal, esta Justiça do Distrito Federal não tem mais competência para prosseguir no julgamento, e os autos deverão ser remetidos à Justiça Federal de 1ª Instância, porquanto firmada a necessidade de formação de litisconsórcio com polo passivo necessário em situações como a dos autos, em que o medicamento requerido não está padronizado.”

Acórdão 1339944, 07028254920208070018, Relator: MARIO-ZAM BELMIRO, Oitava Turma Cível, data de julgamento: 19/5/2021, publicado no PJe: 25/5/2021.

Como a Estatística ajuda o Direito? (Aula passada)



- Através de análises descritivas
- Através de medidas-resumo
- Através das análises de impactos que ajudam na reformulação de leis
- Através das análises internas que embasam decisões estratégicas, como priorização de acordos, provisionamento, mudança de políticas internas etc

Como o Direito ajuda a Estatística?



- A análise de dados não é um processo “reto”, nem objetivo. Embora exista um caminho a se seguir, ela pressupõe a tomada constante de decisões metodológicas.
- Quando esses dados se tratam de dados jurídicos, essas decisões precisam ser embasadas pelo direito.
- Assim, o Direito ajuda a Estatística, principalmente, por meio de conhecimentos de direito processual



PAUSA PARA PERGUNTAS

The background is a light gray with several abstract elements. On the left, there are wavy, horizontal lines. In the center, there is a faint outline of a human brain. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). On the far right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board (PCB) layout.

1. DEFINIÇÃO DOS OBJETIVOS DO ESTUDO

Análise Confirmatória VS Análise Exploratória



- **Análise Confirmatória:** Abordagem estatística utilizada para testar hipóteses ou teorias pré-definidas em um conjunto de dados. Utiliza técnicas como testes de hipóteses, análise de regressão e ANOVA, sendo crucial para a validação de modelos ou teorias.
- **Análise Exploratória:** Utilizada quando a compreensão das variáveis ou relações em um conjunto de dados é incerta. Seu objetivo é descobrir padrões, tendências, relações e características nos dados.

Importante: A definição dos objetivos da pesquisa só faz sentido ocorrer **ANTES** da listagem dos processos **QUANDO** a análise tiver caráter **CONFIRMATÓRIO**

Tradução: Juridiquês para Estatistiquês



- Quando nos propomos a investigar a realidade de forma quantitativa, partimos de alguma problemática inicial, de uma pergunta de pesquisa.
- As explicações para a nossa pergunta envolvem, sempre, conceitos
- Os conceitos são elaborações teóricas e, portanto, não são materiais, ou seja, eles não existem no mundo real, eles não podem ser tocados, sentidos.
- Assim, se queremos acessar os conceitos por meio de uma investigação empírica, precisamos traduzir o conceito em algo real, palpável.
- Disso, surge o problema que vamos investigar:

*Como passar das **definições teóricas** às **definições operacionais** com a segurança de que as variáveis escolhidas medem o conceito que se quer traduzir operacionalmente?*

Identificando a operacionalização



- Em toda operacionalização, eu preciso ter clareza de duas coisas: (1) o que é o conceito e (2) o que é a variável:
 - O conceito é sempre fruto de elaboração teórica.
 - A variável é sempre uma medida.
 - Exemplos de palavras que são usualmente encontradas para se referir às variáveis: média, proporção, razão, taxa, percentual, valor, contagem, frequência...

Identificando a operacionalização



A operacionalização

- A operacionalização diz respeito à adequação entre a variável e o conceito. É apenas com a clareza do que é o conceito e o que é a variável que conseguimos construir a operacionalização.
- Se olharmos apenas para o conceito, não estaremos olhando para a operacionalização, mas estaremos apenas em uma discussão teórica.
- Se olharmos apenas para a variável, tampouco estaremos olhando para a operacionalização, mas estaremos olhando para os dados apenas. E os dados sem teoria carecem de significação.

Como as variáveis se relacionam com os conceitos?



- A relação que se estabelece entre as variáveis e os conceitos que elas representam é uma relação de “implicação”. Dizemos que as variáveis são implicações observáveis dos conceitos.
- Dizemos que é uma implicação porque, caso o conceito exista ou se verifique na realidade, então algo que decorre dele vai se manifestar no mundo empírico. Essa manifestação do conceito no mundo empírico é material e, por isso, ela é observável.
- Em métodos empíricos-quantitativos, nós precisamos de coisas que sejam observáveis para medir. Por isso é que nós medimos as definições operacionais dos conceitos, e não os conceitos em si. E com essas medições das variáveis (e não dos conceitos em si) é que queremos concluir algo sobre os conceitos que nos interessam.

Exemplo 1: Extinção dos dinossauros

*A extinção dos dinossauros pode ser estudada cientificamente: hipóteses alternativas podem ser desenvolvidas e testadas com respeito às suas implicações observáveis. Uma hipótese sobre a extinção dos dinossauros, desenvolvida por Luis Alvarez e seus colaboradores em Berkeley no fim dos anos 1970s (W. Alvarez and Asaro, 1990), **propõe que houve uma colisão cósmica**: um meteorito se chocou com a Terra voando a, aproximadamente, 72.000 km/h, criando uma explosão maior do que uma guerra nuclear mundial. Se essa hipótese estiver correta, então seria uma implicação observável que o iridium (um elemento muito comum em meteoritos, mas muito raro na Terra) poderia ser encontrado em alguma camada particular da crosta terrestre que corresponda aos sedimentos de 65 milhões de anos atrás; de fato, a descoberta de iridium nas camadas previstas tem sido usada como evidência para a teoria de Alvarez (...). Ainda existem outras implicações observáveis. Por exemplo, deveria ser possível também encontrar uma cratera de meteorito em algum lugar da Terra (e muitas pessoas já encontraram).*

Exemplo 2: Vulnerabilidade juvenil



Conceito de 2000, criado pela Fábrica de Cultura da Secretaria de Estado da Cultura, no município de São Paulo. A vulnerabilidade juvenil é um conceito medido a partir das seguintes variáveis:

- Taxa anual de crescimento populacional entre 1991 e 2000;
- Percentual de jovens, de 15 a 19 Anos, no total da população dos distritos;
- Taxa de mortalidade por homicídio da população masculina de 15 a 19 anos;
- Percentual de mães adolescentes, de 14 a 17 anos;
- Valor do rendimento nominal médio mensal, das pessoas com rendimento, responsáveis pelos domicílios particulares permanentes;
- Percentual de jovens de 15 a 17 anos que não frequentam a escola.

Fonte: Márcia Lima. A operacionalização de conceitos - parte 1. Slides para a aula Métodos e Técnicas de Pesquisa I – 2015

Exemplo 3: Segurança jurídica



Podemos pensar também um exemplo hipotético de como a gente poderia operar um conceito muito importante no Direito: A “segurança jurídica”. Abaixo, tem algumas propostas de “implicações observáveis” da segurança jurídica

- Taxa de litigiosidade
- Tempo de duração dos contratos entre particulares e Administração Pública

Limitações das conclusões



- Sabendo qual é a nossa operacionalização – ou seja, tendo clareza de qual é o nosso conceito, de qual é a nossa variável e de como a variável e o conceito se relacionam – podemos pensar nas suas implicações para as conclusões do estudo.
- Nenhuma operacionalização é perfeita/ideal, mas também não é porque a operacionalização não é ideal que devemos descartá-la por completo.
- A régua para julgar uma operacionalização é a confiança subjetiva que temos nas conclusões.

Exemplo: Litigante habitual (Galanter, 1941)

- **Pergunta de interesse:** Litigantes Habituais estão em vantagem nos litígios judiciais?
- **Conceito:** Vantagem em litígio pelos LHs.
 - O Litigante Habitual (LH) é um ator, no Judiciário, que está envolvido em muitos litígios de natureza similar. Ele se contrapõe ao Participante Eventual (PE), que aparece apenas de vez em quando em conflitos judiciais.
- **Variável:** Taxa de vitória dos LHs.
 - O que vamos considerar como um LH? Abaixo, seguem duas sugestões de mensuração desse conceito.

Operacionalização:

- Sugestão 1 (restritiva): Toda parte que aparecer mais de 1.000 vezes no Judiciário é um LH. Então, iremos medir a taxa de vitória das partes que aparecem mais de 1.000 vezes no Judiciário contra partes que aparecem menos do que mil vezes.
- Sugestão 2 (ampla): Toda pessoa jurídica de direito privado é um LH. Então iremos medir a taxa de vitória das PJs, considerando apenas litígios de PJ x PF.

Exemplo: Litigante habitual (Galanter, 1941)

- O estudo apresenta duas possíveis conclusões: (1) as vantagens dos LHs levam a mais decisões favoráveis a eles; (2) as vantagens dos LHs não resultam em mais decisões favoráveis a eles.
- A sugestão de operacionalização restritiva afeta apenas a conclusão do tipo (2), que se refere à igualdade de vantagens entre LH e PE. Se a pesquisa constatar que não há diferença significativa entre LH e PE, isso pode ocorrer por duas razões: (i) realmente não há diferença significativa entre eles ou (ii) a suposta igualdade de vantagens é resultado da inclusão de diversos LHs no grupo de PEs, tornando os grupos comparados semelhantes, já que o grupo de LHs contém apenas LHs, enquanto o grupo de PEs contém PEs e LHs.
- Portanto, a validade da conclusão (2) pode ser comprometida. No entanto, a conclusão oposta, a conclusão (1), não é afetada. Assim, mesmo com uma operacionalização restritiva, se o estudo identificar uma diferença significativa a favor de PE ou LH, essa conclusão será válida.

Reportando incertezas

- Sempre haverá incerteza no processo de operacionalização. A incerteza é do quanto que a variável escolhida está, de fato, representando conceito de interesse. Isso acontece porque sequer é possível de encontrar a melhor operacionalização de um conceito, ou seja, nunca haverá uma correspondência exata entre o conceito e a variável que está sendo medida para representar o conceito.
- A melhor prática então não é fazer uma busca incessante pela melhor operacionalização possível, mas é reportar e explicitar com clareza qual foi a operacionalização utilizada e quais as incertezas neste processo.

Um pesquisador que falhe em encarar o problema da incerteza diretamente está ou dizendo que ele/ela sabe de tudo perfeitamente ou de que ele/ela não tem a menor ideia do quão certos ou incertos são os seus resultados.

Esforço de objetivação

- Outra forma de encarar a incerteza da operacionalização é entender que essa incerteza é um processo subjetivo, pois quanto maior for a incerteza sobre os dados, mais a interpretação deles irá depender da pessoa que está realizando a pesquisa.
- Entender a operacionalização como um processo subjetivo nos leva a ter de pensar na seguinte pergunta: como mitigar a subjetividade na ciência, que se pretende ser objetiva?

Para tanto, devemos também explicitar, não somente as incertezas sobre operacionalização realizada, ou as limitações acerca das conclusões, mas também os nossos pressupostos teóricos pessoais.

Fonte: PIRES, Álvaro P. Sobre algumas questões epistemológicas de uma metodologia geral para as ciências sociais.
In: POUPART, Jean (Org.). A pesquisa qualitativa: Enfoques epistemológicos e metodológicos. Petrópolis: Vozes. 2010

Conclusões



- Em pesquisas, é importante saber transformar os conceitos em variáveis, isto é, em unidades de medidas observáveis, como frequência/quantidade, taxa/proporção, média, mediana etc
- É importante ter conhecimento sobre as incertezas geradas em um processo de operacionalização. O estudo deve sempre documentar as suposições, imposições ou qualquer outra decisão acerca dessa “tradução”.
 - Exemplo: Testar a hipótese de que determinada política pública trouxe economia para o Estado em processos de determinado tema.
 - **Conceito/hipótese:** Houve economia.
 - **Variável:** Valor médio (comparar antes vs depois)
 - **Incerteza:** Variabilidade em torno da média (isso também deve ser estudado)

Diferenças para análises internas

- Nenhuma. A tradução e definição dos objetivos depende de um trabalho conjunto entre o Estatístico e a equipe jurídica por trás do estudo.

2. LISTAGEM DE PROCESSOS: DEFININDO O ESCOPO TEMPORAL

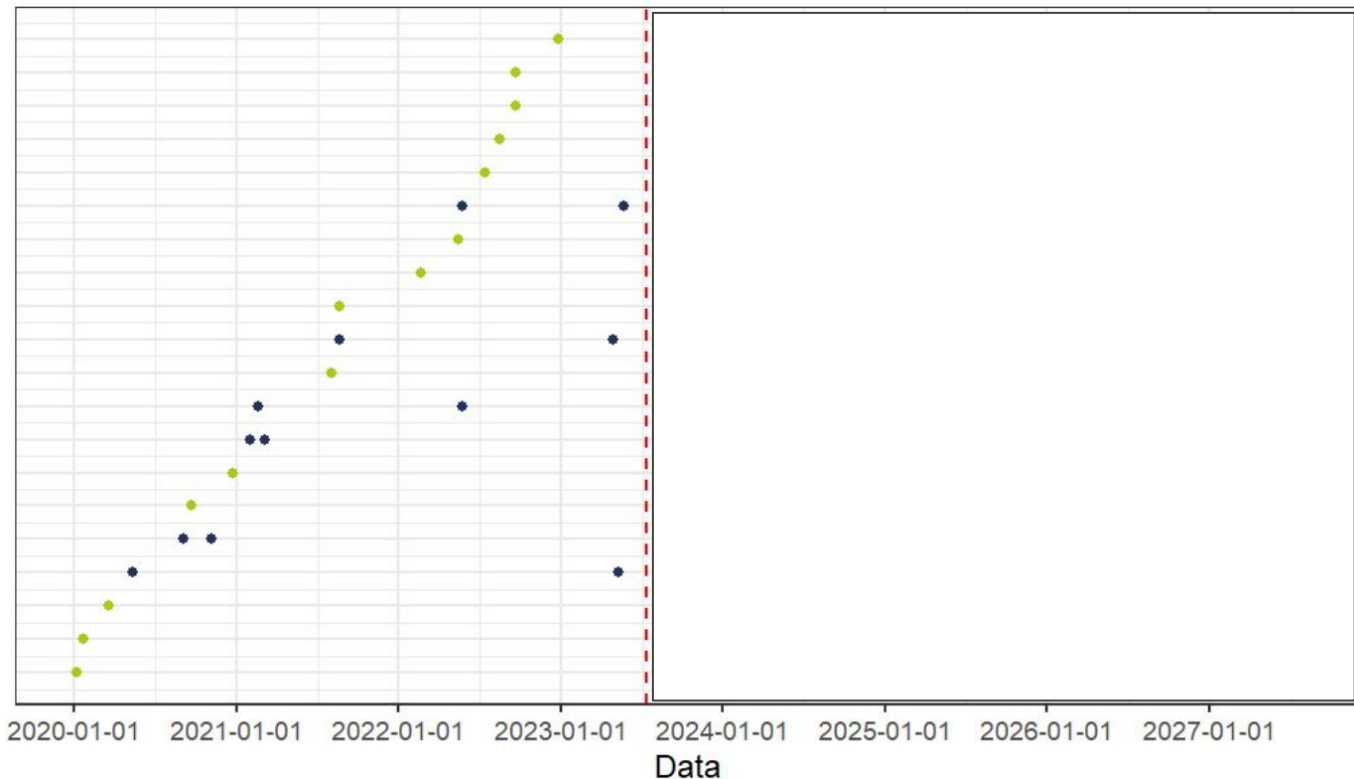
The background is a light gray with a complex, abstract design. On the left, there are wavy, organic lines. In the center, a stylized brain is depicted. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). Further to the right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board (PCB) layout.

Análise de tempos



Vamos começar com uma pesquisa fictícia que quer estudar o perfil de julgamento de processos distribuídos entre janeiro de 2020 e julho de 2023.

Análise de tempos



Tempo médio dos processos, considerando apenas os processos encerrados (azuis): 439 dias.

Tempo médio de todos os processos: 1.102 dias

● Iniciou e encerrou no período

● Iniciou no período, mas não se encerrou

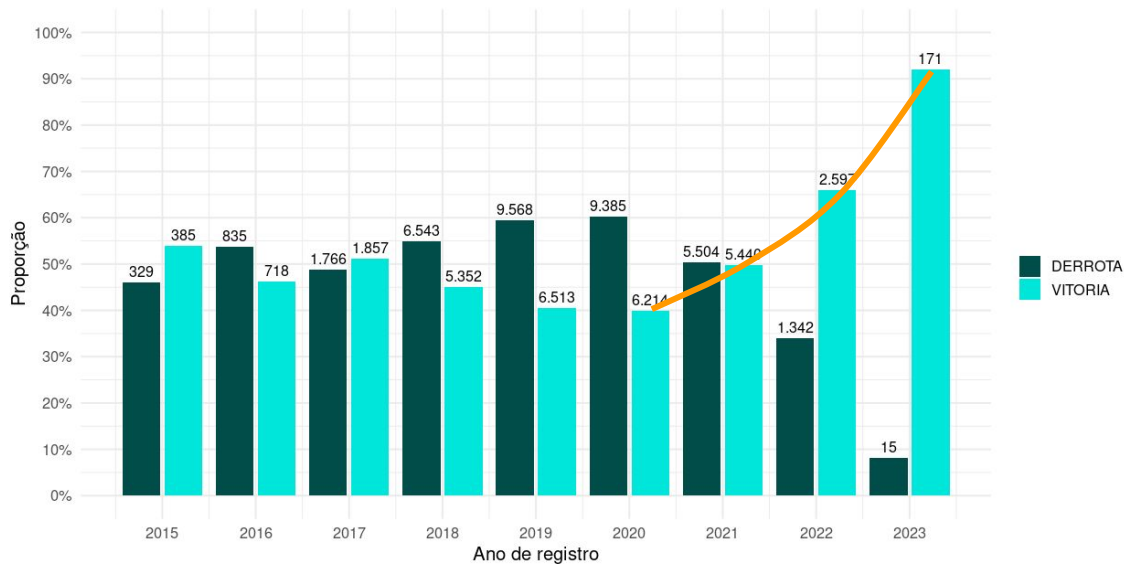
Análise de tempos



Há dois tipos de escopo em uma pesquisa:

- **Pesquisa retrospectiva:** São as pesquisas cujo escopo temporal considera apenas casos finalizados/julgados até determinada data. Isso é feito buscando os processos pela decisão/sentença/acórdão
- **Pesquisa prospectiva:** São as pesquisas cujo escopo temporal considera todos os casos iniciados a partir de uma data, independente de seu fim. Isso é feito buscando os processos de acordo com a sua data de distribuição

Ainda assim temos que tomar cuidado...



- **Objetivo:** Desenvolver um modelo estatístico para estimar o risco de perda de um processo.
- **Tipo de Estudo:** Retrospectivo

O gráfico acima mostra a evolução da taxa de derrota/vitória em processos cíveis **encerrados por data de registro** da ação no sistema interno da companhia.

Análise de tempos



Conclusão parcial: Temos um **problema sério**: a gente SEMPRE vai enviar as análises de tempo sob algumas situações:

1. Pesquisas retrospectivas, com limitação de data de distribuição
2. Pesquisas prospectivas, em que exista pelo menos um processo ainda não se encerrou.

A solução para isso é uma metodologia estatística chamada **Análise de Sobrevivência**.

A análise de sobrevivência considera o tempo decorrido até o momento do estudo.

- Para os eventos que já ocorreram, tem-se o tempo total entre o início e fim.
- Para os eventos que ainda não ocorreram, tem-se a informação de que, até a data do estudo, aquele evento **ainda** não ocorreu.

Análise de tempos: Curiosidade



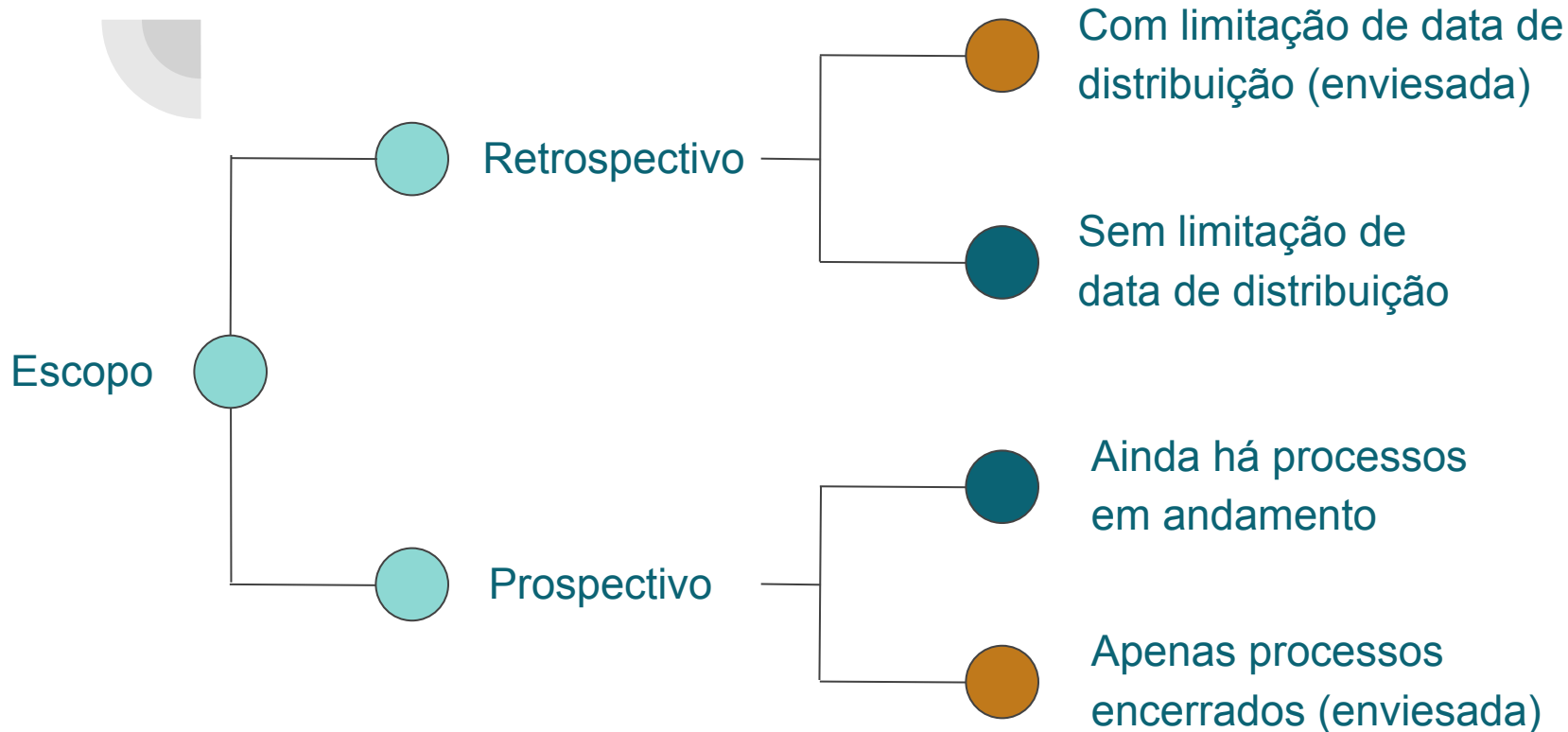
Análise de sobrevivência:

- Surgiu na medicina, em estudos epidemiológicos
- A pergunta dos médicos era: se eu fizer um tratamento, a chance de sobrevivência do paciente aumenta? Quanto?
- O problema era que, se o tratamento funcionasse, o paciente poderia viver ainda por muito tempo além do período disponível para o estudo e, portanto, o médico não conseguia medir o tempo de sobrevida do paciente

Informações estatísticas:

- “Não-morte” do paciente = censura
- Data do fim do estudo = data da censura

Conclusões



P.S. Não tem problema enviesar a análise, desde que seja o objetivo.

The background is a light gray with a complex, abstract design. On the left, there are wavy, organic lines. In the center, a stylized brain is depicted with a spiral of binary code (0s and 1s) winding around it. To the right, there are geometric, circuit-like lines with small circles at the ends, resembling a printed circuit board. The overall aesthetic is technological and intellectual.

2. LISTAGEM DE PROCESSOS: DEFININDO O ESCOPO TEMÁTICO

TPUs



- **O que são:** São as Tabelas Processuais Unificadas (TPUs).
- **Objetivo:** As Tabelas Processuais Unificadas (TPUs) buscam sistematizar uniformizar a taxonomia e terminologia de “classes, assuntos, movimentações e documentos processuais aplicáveis a todos os órgãos do Poder Judiciário”.
- **Construção:** Desde a Resolução nº 46 de 18 de dezembro de 12 de 2007.
- **Gestão:** Um Comitê do CNJ, chamado Comitê Gestor da Numeração Única e das Tabelas Processuais Unificadas
- **Relevância para a Jurimetria:** No mundo ideal, a escolha do **escopo temático** poderia ser **facilitada** por um conjunto de assuntos, classes e movimentações específicas. Exemplo: Direito do Consumidor, Aviação, Telecom etc...

➔ CONSULTA PÚBLICA DE ASSUNTOS

Pesquisar: ☒ Assunto ☐ Glossário

nome ou código

Pesquisar

Versão 05/07/2023

- + 12775 DIREITO À EDUCAÇÃO
- + 9985 DIREITO ADMINISTRATIVO E OUTRAS MATÉRIAS DE DIREITO PÚBLICO
- + 10110 DIREITO AMBIENTAL
- + 12734 DIREITO ASSISTENCIAL
- + 899 DIREITO CIVIL
- + 9633 DIREITO DA CRIANÇA E DO ADOLESCENTE
- + 12480 DIREITO DA SAÚDE
- + 1156 DIREITO DO CONSUMIDOR
- + 864 DIREITO DO TRABALHO
- + 11428 DIREITO ELEITORAL
- + 10739 DIREITO ELEITORAL E PROCESSO ELEITORAL DO STF
- + 6191 DIREITO INTERNACIONAL

➔ DETALHAMENTO

➔ TEMPORALIDADE

Justiça Estadual

- ☐ 1º Grau ☐ 2º Grau ☐ Juizado Especial ☐ Turmas Recursais
- ☐ Juizado Especial da Fazenda Pública
- ☐ Turma Estadual de Uniformização

Competência Militar

- ☐ 1º Grau ☐ 2º Grau

Justiça Federal

- ☐ 1º Grau ☐ 2º Grau ☐ Juizado Especial ☐ Turmas Recursais
- ☐ Turma regional de unifor. ☐ Turma nacional de unifor. ☐ CJF

Funcionamento



As TPUs se referem a 4 tabelas:

- Classe
- Assunto
- Movimentações
- Documentos Processuais

Classes

Movimentos

Assuntos

Documentos
Processuais

Funcionamento



Cada tabela é composta pela junção de 26 tabelas menores (exceto a tabela de documentos processuais)

Data Versão	Justiça	Tabela (Classes)
05/07/2023	Justiça do Trabalho	Trabalho 1º Grau ▼ Gerar Excel
05/07/2023	Justiça Federal	Federal 1º Grau ▼ Gerar Excel
05/07/2023	Justiça Estadual	Estadual 1º Grau ▼ Gerar Excel
05/07/2023	Justiça Militar	Militar da União 1º ▼ Gerar Excel
05/07/2023	Justiça Eleitoral	Eleitoral Zonas Eleitorais ▼ Gerar Excel
05/07/2023	Superiores	STF ▼ Gerar Excel

Funcionamento

Justiça do trabalho:

1. Trabalho 1º grau
2. Trabalho 2º grau
3. Trabalho TST

Justiça Federal:

4. Federal 1º grau
5. Federal 2º grau
6. Federal Juizados Especiais
7. Federal Turmas Recursais
8. Federal Turma Nacional de Uniformização
9. Federal Turma Regional de Uniformização

Justiça Estadual

10. Estadual 1º grau
11. Estadual 2º grau
12. Estadual Juizados Especiais
13. Estadual Turmas Recursais de Uniformização
14. Juizados Especiais da Fazenda Pública
15. Turma Estadual de Uniformização

Justiça Militar:

16. Militar da União 1º
17. Militar da União STM
18. Militar Estadual 1º
19. Militar Estadual TJM

Justiça Eleitoral:

20. Eleitoral Zonas Eleitorais
21. Eleitoral TRE
22. Eleitoral TSE

Superiores:

23. STF
24. STJ
25. CJF
26. CNJ

Funcionamento



- Cada tabela é composta pela junção de 26 tabelas menores (exceto a tabela de documentos processuais)
- A junção das 26 tabelas menores, tirando as duplicatas, equivale a 1 TPU

Versão 05/07/2023

+ -	385	EXECUÇÃO PENAL E DE MEDIDAS ALTERNATIVAS
+ -	1198	PROCEDIMENTOS ADMINISTRATIVOS
+ -	547	PROCEDIMENTOS DE INFÂNCIA E JUVENTUDE
+ -	11099	PROCEDIMENTOS PRÉ-PROCESSUAIS DE RESOLUÇÃO CONSENSUAL
+ -	2	PROCESSO CÍVEL E DO TRABALHO
+ -	268	PROCESSO CRIMINAL
+ -	11427	PROCESSO ELEITORAL
+ -	11028	PROCESSO MILITAR
+ -	5	SUPERIOR TRIBUNAL DE JUSTIÇA
+ -	1310	SUPREMO TRIBUNAL FEDERAL


Funcionamento





- Cada tabela é composta pela junção de 26 tabelas menores (exceto a tabela de documentos processuais)
- A junção das 26 tabelas menores, tirando as duplicatas, equivale a 1 TPU
- 1 TPU é uma Tabela, seja de Classe, Assunto, Movimentos ou Documentos, que possui 6 níveis hierárquicos
 - O nível 1 é sempre o nível mais genérico
 - O nível mais específico pode variar
- Cada classificação possui um código próprio, tendo ela nível mais específico ou não


 385 EXECUÇÃO PENAL E DE MEDIDAS ALTERNATIVAS


 386 Execução da Pena

 12775 DIREITO À EDUCAÇÃO

 12795 Acesso

 12805 Acesso sem Conclusão do Ensino Médio

 14177 Ciência Sem Fronteiras

 12808 Cobrança de Taxa de Matrícula

Funcionamento



- Cada tabela é composta pela junção de 26 tabelas menores (exceto a tabela de documentos processuais)
- A junção das 26 tabelas menores, tirando as duplicatas, equivale a 1 TPU
- 1 TPU é uma Tabela, seja de Classe, Assunto, Movimentos ou Documentos, que possui 6 níveis hierárquicos
 - O nível 1 é sempre o nível mais genérico
 - O nível mais específico pode variar
- Cada classificação possui um código próprio, tendo ela nível mais específico ou não
- Quem atribui uma Classe ou Assunto a um processo é o próprio operador do Direito que provoca o Judiciário
 - Quem registra as Movimentações e Documentos é um serventuário da Justiça

Funcionamento



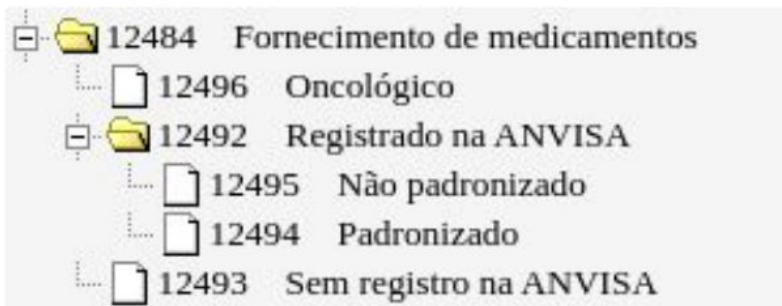
Momento Marketing: Para mais informações, acompanhe o blog da ABJ

Problemas na listagem dos processos

1. Os códigos das TPUs mudam.

Atualmente: 12484 - Fornecimento de medicamentos

- Código genérico



Antigamente: 11884 - Fornecimento de medicamentos

- Código específico



Mudança: Isso mudou na versão do dia 04/07/2019, em decorrência da pesquisa realizada pelo Insper: [Judicialização da Saúde no Brasil: Perfil das Demandas, Causas e Propostas de Solução](#)

Implicação: Uma pesquisa sobre Judicialização da Saúde com escopo temporal de 2015-2021 que usar apenas o código 12484 irá excluir os processos anteriores à 2019.

Problemas na listagem dos processos



2. Cifra oculta

A cifra oculta é a quantidade não observada de determinado dado. Esse problema surge, quando falamos das Classes e Assuntos processuais, por causa da forma como esses dados são gerados.

- O advogado pode parar a classificação em qualquer nível da taxonomia, e não necessariamente no nível mais específico
- O processo pode não se enquadrar em uma categoria de nível mais específico
- Pode haver divergência de interpretação da classificação entre advogados
- Um mesmo tipo de processo pode se enquadrar em mais de uma categoria de Classe ou Assunto

Problemas na listagem dos processos

2. Cifra oculta

Tudo isso gera problema de interpretação nos assuntos genéricos

Implicação: Não podemos assegurar a listagem completa de todos os processos, impossibilitando a realização de pesquisas populacionais. Sendo assim, nos resta conduzir pesquisas amostrais, independentemente do tamanho da amostra.

Sempre teremos que fazer inferências sobre a população.

P.S. Outra opção é buscar por uma classe/assunto genérica e criar classificadores automáticos com base em modelos de Machine/Deep Learning e/ou utilização de ferramentas NLP.

O problema aqui se torna lidar com uma grande quantidade de informações e, na maioria dos casos, irrelevantes para a pesquisa.

Problemas na listagem dos processos



3. Sistemas de pesquisa processual

Por não existir (ainda) um sistema unificado de pesquisa de processos, cada sistema possui informações e padronizações específicas, bem como limitações.

- Limitações no volume, qualidade e disponibilidade das informações podem impactar uma listagem automática de processos, sendo necessário partir para uma análise manual.


Exemplo 1: O mesmo tribunal pode ter mais do que 1 sistema de consulta processual, cada qual com suas especificações.

Exemplo 2: No Esaj não é possível buscar processos por data de distribuição.

Problemas na listagem dos processos

3. Sistemas de pesquisa processual

Exemplo 3: O Tribunal de Justiça do Paraná possui reCAPTCHA.

 **TJPR** | cível

Consulta Pública

Ao eleger uma forma de pesquisa, obrigatório preencher todos os campos marcados com * .

* Número Processo: Formato: **NNNNNNDDAA8160000**

Acesse também a relação de **Processos Prioritários** (de acordo com o novo CPC)

OU

* Número Antigo:

* Ano:

* Vara:

Competência:

Nome:

Distribuição:

OU

* OAB do Advogado:


Número

Partes

**Classe Processual
(Assunto Principal)**

Vara

Nenhum registro encontrado.

 protegido por reCAPTCHA

Conclusões



- Não se esqueça dos códigos históricos quando os códigos das TPUs forem a base para a listagem dos processos.
- Em pesquisas, discuta metodologicamente a cifra oculta e como ela, potencialmente, pode estar enviesando seus dados.
- Lembre-se de determinar muito bem seu escopo regional e sistemas de busca, pois dependendo dos objetivos da pesquisa, o acesso à informação pode ser complicado.

Diferenças para análises internas

- Em uma análise interna, o interessado normalmente já possui um sistema de gerenciamento processual, onde armazena as informações necessárias para o estudo. A listagem dos processos é muito mais fluida.

3. EXTRAÇÃO E ESTRUTURAÇÃO DOS DADOS

The background is a light gray with several abstract elements. On the left, there are wavy, horizontal lines. In the center, there is a faint outline of a human brain. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). On the far right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board (PCB) layout.

Tipos de extração



Quando o estudo chega nesta etapa, já foram definidos i) o escopo temporal, ii) o escopo regional e iii) o escopo temático da sua pesquisa e os processos de interesse já foram listados.

Exemplo: Analisar o perfil de julgamento de todos os julgados do primeiro grau do TJSP que falam sobre Recuperação Judicial e falência.

Escopo Temporal: Desde sempre

Escopo Regional: **TJSP** (Consulta de Julgados do Primeiro Grau - CJPG)

Escopo Temático: Recuperação Judicial e falência

Tipos de extração



O objetivo é transformar as informações de todos esses processos em uma **base de dados bruta**.

Para isso, a extração pode ser feita de duas formas:

1. **Automática:** Através de técnicas de raspagem das informações e/ou utilização de ferramentas de Processamento de Linguagem Natural (NLP)
2. **Manual:** Por meio de pesquisadores qualificados.

Tipos de extração

Extração automática:

- Requer conhecimento de computação
- É possível quando:
 - As informações estiverem disponíveis na capa do processo
 - As informações estiverem disponíveis nas movimentações
 - As informações estiverem disponíveis no texto das sentenças e decisões
- Mas isso ainda não é garantia de que será de fato possível

Extração manual:

- Não requer conhecimento de computação
- É possível sempre, mas é sempre limitado pelo tempo e pela escala
- É recomendável quando:
 - As informações estiverem disponíveis nos documentos das parte
 - As informações, por mais que estejam disponíveis nas sentenças e decisões, são muito difíceis de extrair

Voltando ao exemplo de início



Input:

e-SAJ | Consulta de Processos do 1º Grau



1054460-52.2016.8.26.0100

Extinto

Classe

Recuperação Judicial

Assunto

Recuperação judicial e Falência

Foro

Foro Central Cível

Vara

2ª Vara de
Falências e...

Juiz

PAULO FURTADO DE OLIVEIRA
FILHO

Distribuição

31/05/2016 às 11:37 -
Direcionada

Controle

2016/000380

Área

Cível

Valor da ação

R\$ 100.000,00

Outros assuntos

Empresas

Voltando ao exemplo de início



Output:

	id_processo	status	classe	assunto
1	10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência
2	10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência
3	00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência
4	10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência
5	10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência
6	00237449820128260100	Extinto	Recuperação Judicial	Administração judicial
7	00187684820128260100	Extinto	Recuperação Judicial	Administração judicial
8	10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência
9	00323077120128260071	Suspenso	Recuperação Judicial	Empresas
10	00010360720138260363	NA	Recuperação Judicial	Empresas

Extração automática

```
function(url) {  
  
  # Lendo o HTML do site  
  webpage <- read_html(url)  
  
  # Extraíndo os nomes dos livros  
  book_names <- webpage %>%  
    html_nodes(".product_pod h3 a") %>%  
    html_attr("title")  
  
  # Extraíndo os preços dos livros  
  book_prices <- webpage %>%  
    html_nodes(".product_price .price_color") %>%  
    html_text()  
  
  # Extraíndo a classificação (rating) dos livros  
  book_ratings <- webpage %>%  
    html_nodes(".star-rating") %>%  
    html_attr("class") %>%  
    gsub("star-rating ", "", .)  
  
  # Organizando os dados extraídos em um dataframe  
  df_books <- data.frame(  
    Name = book_names,  
    Price = book_prices,  
    Rating = book_ratings,  
    stringsAsFactors = FALSE  
  )  
}
```

Extração manual


Seção 2 de 7

Informações processuais

Informações sobre o processo

Data de distribuição *

Data de distribuição do processo. Se houver mais de uma, utilizar a mais antiga, desde que seja uma data válida

Mês, dia, ano 

Classe *


Classe processual seguindo a Res. 46 do CNJ







Texto de resposta curta

Assunto *

Assunto processual seguindo a Res. 46 do CNJ

Texto de resposta curta



Conclusões

- A partir de uma listagem de processos, realizar a extração das informações disponíveis.
- É possível realizar a extração automática?
 - Sim
 - Desenvolver/Utilizar o robô de extração
 - Estruturação das informações
 - Não
 - Amostragem aleatória da lista de processos
 - Construção de um formulário de classificação
 - Treinamento dos pesquisadores
 - Análise/Extração manual
 - Estruturação das informações

Diferenças para análises internas

- Os sistemas de gerenciamento geralmente mantêm informações processuais de uma organização ou departamento jurídico em uma base de dados estruturada. Isso permite a automatização na geração de relatórios e na utilização de APIs para acesso a essas informações de forma automatizada e organizada.

4. TRATAMENTO DOS DADOS

The background is a light gray with a complex, abstract design. On the left, there are wavy, organic lines. In the center, a stylized brain is depicted. To the right of the brain, there is a circular arrangement of binary code (0s and 1s). Further to the right, there are circuit-like lines with small circles at the ends, resembling a printed circuit board (PCB) layout.

Análise de qualidade



A curadoria dos dados é o processo de identificar, mapear e corrigir eventuais inconsistências que podem aparecer nos dados jurídicos internos e que podem trazer danos graves ao escritório, departamento jurídico ou entidade pública, além de prejudicar as análises jurimétricas e a implementação de ferramentas analíticas que utilizam esses dados como fonte de informação.

Fonte: DALEFFI, B; TRECENTI, J; LENTE, C. Curadoria de dados Jurídicos: A parte essencial da Jurimetria que quase ninguém menciona. Terranova Consultoria. 2021.

Análise de qualidade

- **O que é:** Analisar a integridade das informações.
- **Objetivos:** Mapear inconsistências e criar um plano de ação.
- **Metodologias:**
 - Comparações lógicas:
 - Data de distribuição > Data da sentença
 - Status “Extinto” com última movimentação “Aguardando Julgamento”
 - Lacunas:
 - Parte contrária ausente
 - Assunto ausente
- **Lembrete:** Garbage in, Garbage out

Análise de qualidade

O que fazer?

	id_processo	status	classe	assunto
1	10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência
2	10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência
3	00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência
4	10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência
5	10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência
6	00237449820128260100	Extinto	Recuperação Judicial	Administração judicial
7	00187684820128260100	Extinto	Recuperação Judicial	Administração judicial
8	10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência
9	00323077120128260071	Suspenso	Recuperação Judicial	Empresas
10	00010360720138260368	NA	Recuperação Judicial	Empresas

Análise de qualidade

E agora?

id_processo	classe	assunto	valor_causa
1	Recuperação Judicial	Recuperação Judicial e falência	NA
2	Recuperação Judicial	Recuperação Judicial e falência	457893.0
3	Recuperação Judicial	Recuperação Judicial e falência	681214.6
4	Recuperação Judicial	Recuperação Judicial e falência	496250.9
5	Recuperação Judicial	Recuperação Judicial e falência	NA
6	Recuperação Judicial	Recuperação Judicial e falência	962960.6
7	Recuperação Judicial	Recuperação Judicial e falência	687897.5
8	Recuperação Judicial	Recuperação Judicial e falência	NA
9	Recuperação Judicial	Recuperação Judicial e falência	670830.8
10	Recuperação Judicial	Recuperação Judicial e falência	456038.5

Análise de qualidade

E agora?

id_processo	id_mov	data	movimento	descricao	tem_anexo
10544605220168260100	1	2016-05-31	Decisão	Vistos. 1 Indefiro a concessão do benefício de justiça gratuit...	TRUE
10544605220168260100	2	2016-05-31	Conclusos para Decisão	NA	FALSE
10544605220168260100	3	2016-05-31	Distribuído por Direcionamento (movimentação exclusiva d...	Há suspeita de repetição da ação. Confronte os dados do pr...	FALSE
10544605220168260100	4	2016-06-01	Remetido ao DJE	Relação: 0168/2016 Teor do ato: Vistos. 1 Indefiro a concess...	FALSE
10544605220168260100	5	2016-06-01	NA	Relação :0168/2016 Data da Disponibilização: 02/06/2016 D...	FALSE
10544605220168260100	6	2016-06-28	Pedido de Extinção Juntada	Nº Protocolo: WJMJ.16.40565843-0 Tipo da Petição: Pedido ...	FALSE
10544605220168260100	7	2016-08-01	Conclusos para Sentença	NA	FALSE
10544605220168260100	8	2016-08-05	Extinto o Processo sem Resolução do Mérito por Desistência	Vistos. Homologo, para que produza seus efeitos jurídicos e...	TRUE
10544605220168260100	9	2016-08-10	Remetido ao DJE	Relação: 0240/2016 Teor do ato: Vistos. Homologo, para qu...	FALSE
10544605220168260100	10	2016-08-11	Certidão de Publicação Expedida	Relação :0240/2016 Data da Disponibilização: 11/08/2016 D...	FALSE
10544605220168260100	11	2016-09-05	Sentença Registrada	NA	FALSE
10544605220168260100	12	2016-09-06	Arquivado Definitivamente	NA	FALSE
10544605220168260100	13	2016-09-06	NA	NA	FALSE
10544605220168260100	14	2016-09-06	Certidão de Trânsito em Julgado com Baixa Expedida	Certidão - Trânsito em Julgado com Baixa - Processo Digital	TRUE

Conceito de inconsistências em dados



- **Definição:** Inconsistências em dados são quaisquer disparidades, contradições ou erros em um conjunto de dados.
- **Impacto:** Inconsistências podem distorcer os resultados, levando a interpretações equivocadas e, conseqüentemente, a decisões erradas.

Origens das inconsistências



- **Erros na entrada de dados:** causados por falhas humanas ou técnicas no momento de inserção.
- **Falhas na extração:** erros que surgem durante a extração automática ou manual de informações de fontes diversas.
- **Inconsistências temporais:** decorrentes de mudanças ao longo do tempo, mas não atualizadas no banco de dados.
- **Contradições entre fontes:** diferenças entre informações de diferentes fontes.

Identificação das inconsistências



- **Análise exploratória:** Uso de ferramentas e técnicas estatísticas para identificar possíveis outliers ou valores que não fazem sentido no contexto.
 - Exemplo: aula passada
- **Ferramentas de qualidade de dados:** softwares específicos que ajudam a detectar, corrigir e prevenir erros em conjuntos de dados.
 - Exemplo: detectar datas erradas, valores vazios etc
- **Cruzamento de dados:** Comparar informações de diferentes fontes para encontrar disparidades.
 - Exemplo: Cruzar a base de dados interna com uma base de dados pública.

Resolução das inconsistências



- **Limpeza de dados:** Processo de identificar e corrigir (ou remover) erros e inconsistências em dados para melhorar sua qualidade.
- **Verificação manual:** Em alguns casos, especialmente quando os dados são críticos, pode ser necessário revisar manualmente.
- **Padronização:** Estabelecer regras e padrões para entrada de dados pode ajudar a reduzir inconsistências futuras.

Voltando aos exemplos

	id_processo	status	classe	assunto
1	10544605220168260100	Extinto	Recuperação Judicial	Recuperação judicial e Falência
2	10046286820168260482	NA	Recuperação Judicial	Recuperação judicial e Falência
3	00037451620128260568	Extinto	Recuperação Judicial	Recuperação judicial e Falência
4	10234819020158260602	Tramitação prioritária	Falência de Empresários, Sociedades Empresariais, Microem...	Recuperação judicial e Falência
5	10004264020168260032	Extinto	Recuperação Judicial	Recuperação judicial e Falência
6	00237449820128260100	Extinto	Recuperação Judicial	Administração judicial
7	00187684820128260100	Extinto	Recuperação Judicial	Administração judicial
8	10023675520148260562	Extinto	Recuperação Judicial	Recuperação judicial e Falência
9	00323077120128260071	Suspenso	Recuperação Judicial	Empresas
10	00010360720138260368	NA	Recuperação Judicial	Empresas

- Por construção, o esaj não faz a classificação de status de processos ativos.
- Portanto, basta criar uma rotina automática que classifica os status vazios como “Ativo”

Voltando aos exemplos

id_processo	classe	assunto	valor_causa
1	Recuperação Judicial	Recuperação Judicial e falência	NA
2	Recuperação Judicial	Recuperação Judicial e falência	457893.0
3	Recuperação Judicial	Recuperação Judicial e falência	681214.6
4	Recuperação Judicial	Recuperação Judicial e falência	496250.9
5	Recuperação Judicial	Recuperação Judicial e falência	NA
6	Recuperação Judicial	Recuperação Judicial e falência	962960.6
7	Recuperação Judicial	Recuperação Judicial e falência	687897.5
8	Recuperação Judicial	Recuperação Judicial e falência	NA
9	Recuperação Judicial	Recuperação Judicial e falência	670830.8
10	Recuperação Judicial	Recuperação Judicial e falência	456038.5

- [Fácil] Substituir os vazios pelo valor médio ou mediano
- [Difícil] Consultar na petição inicial o valor da causa
- [Difícil] Análise manual

Voltando aos exemplos

id_processo	id_mov	data	movimento	descricao	tem_anexo
10544605220168260100	1	2016-05-31	Decisão	Vistos. 1 Indefiro a concessão do benefício de justiça gratuit...	TRUE
10544605220168260100	2	2016-05-31	Conclusos para Decisão	NA	FALSE
10544605220168260100	3	2016-05-31	Distribuído por Direcionamento (movimentação exclusiva d...	Há suspeita de repetição da ação. Confronte os dados do pr...	FALSE
10544605220168260100	4	2016-06-01	Remetido ao DJE	Relação: 0168/2016 Teor do ato: Vistos. 1 Indefiro a concess...	FALSE
10544605220168260100	5	2016-06-01	NA	Relação :0168/2016 Data da Disponibilização: 02/06/2016 D...	FALSE
10544605220168260100	6	2016-06-28	Pedido de Extinção Juntada	Nº Protocolo: WJMJ.16.40565843-0 Tipo da Petição: Pedido ...	FALSE
10544605220168260100	7	2016-08-01	Conclusos para Sentença	NA	FALSE
10544605220168260100	8	2016-08-05	Extinto o Processo sem Resolução do Mérito por Desistência	Vistos. Homologo, para que produza seus efeitos jurídicos e...	TRUE
10544605220168260100	9	2016-08-10	Remetido ao DJE	Relação: 0240/2016 Teor do ato: Vistos. Homologo, para qu...	FALSE
10544605220168260100	10	2016-08-11	Certidão de Publicação Expedida	Relação :0240/2016 Data da Disponibilização: 11/08/2016 D...	FALSE
10544605220168260100	11	2016-09-05	Sentença Registrada	NA	FALSE
10544605220168260100	12	2016-09-06	Arquivado Definitivamente	NA	FALSE
10544605220168260100	13	2016-09-05	NA	NA	FALSE
10544605220168260100	14	2016-09-06	Certidão de Trânsito em Julgado com Baixa Expedida	Certidão - Trânsito em Julgado com Baixa - Processo Digital	TRUE

- Algumas inconsistências são irreversíveis. Nestes casos, precisamos documentar as inconsistências e assumir limitações nas análises.
- Em muitos casos, as inconsistências irreparáveis se manifestam de maneira aleatória. Dependendo do volume de dados, podemos simplesmente desconsiderar a observação inconsistente da análise.

Prevenção das inconsistências



- **Treinamento:** Garantir que as equipes envolvidas saibam a importância da precisão e como inserir dados corretamente.
- **Validações:** Implementar validações de entrada para assegurar que os dados inseridos estão dentro de um formato ou padrão esperado.
- **Revisão periódica:** Estabelecer revisões regulares dos bancos de dados para identificar e corrigir inconsistências.

É importante manter um registro das alterações feitas durante o controle de inconsistências, para que os pesquisadores possam referenciar e justificar decisões tomadas durante a análise.

Conclusão

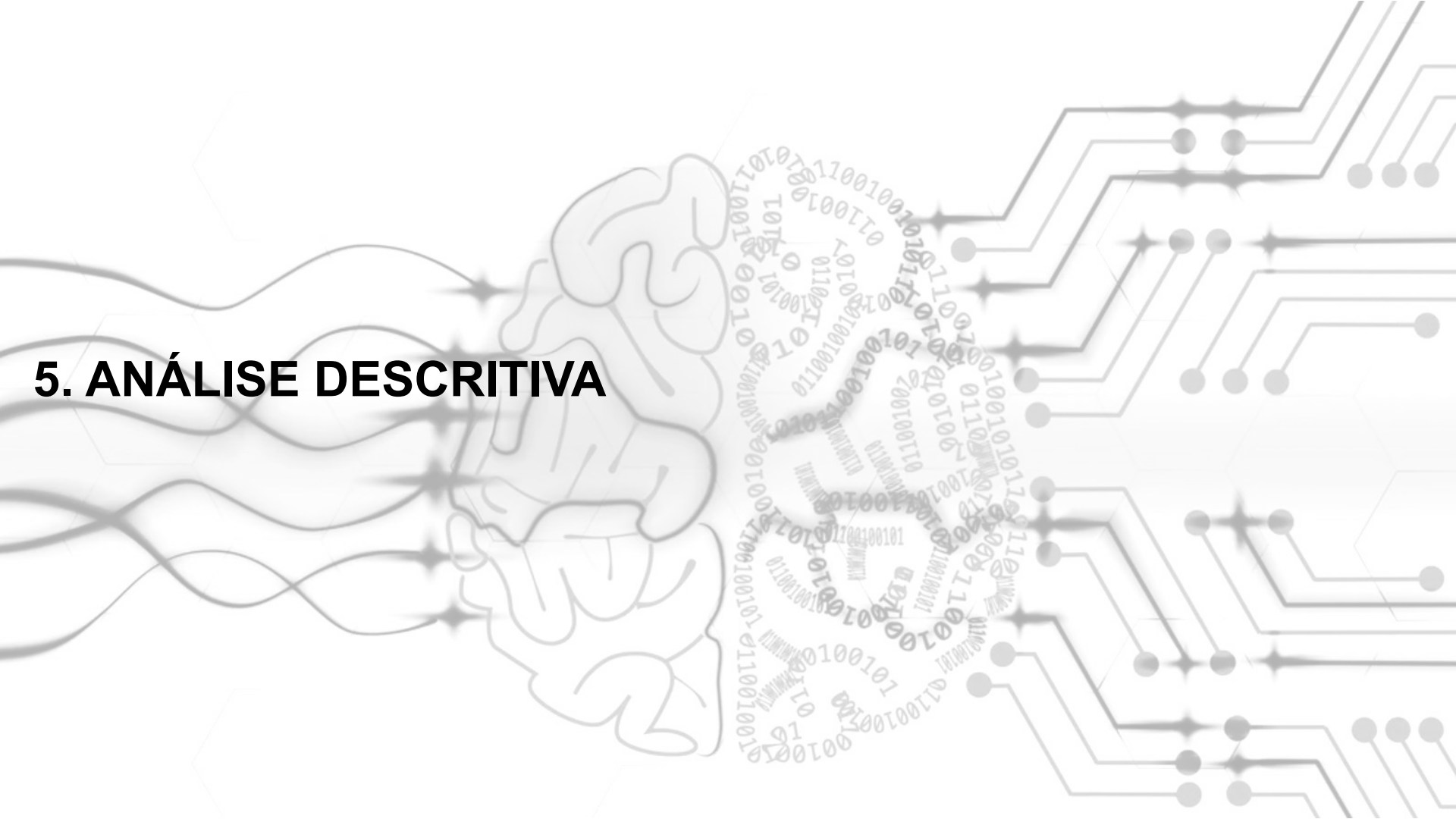


O controle de inconsistências é um passo crucial no processo de pesquisa jurimétrica. Assegurar que os dados são confiáveis e precisos é fundamental para que as análises sejam significativas e os insights gerados sejam de fato relevantes para a tomada de decisões no âmbito jurídico.

Diferenças para análises internas

- Base de dados internas normalmente são mais ricas do que as bases de dados dos tribunais.
 - Informações como valor de condenação, estratégia de acordo, pedidos, datas dos pagamentos etc só estarão estruturadas em uma base de dados interna.
 - Mais informações -> mais inconsistências
- É mais fácil corrigir inconsistências internas do que externas, pois análises externas podem ser utilizadas para corrigir inconsistências em uma base de dados interna.
 - Uma base de dados estruturada interna normalmente é acompanhada dos documentos processuais, isto é, é possível utilizar ferramentas de NLP para corrigir as informações inconsistentes de uma base de dados.
- Muitas inconsistências são geradas quando há mudança e migração de sistema de gerenciamento processual.
 - É possível criar metodologias automáticas de detecção de inconsistências e de padronização/sugestão de informações categóricas (como classe e assuntos)

5. ANÁLISE DESCRITIVA





Obrigado!