

ERRADICAÇÃO
DA POBREZA



FOME ZERO E
AGRICULTURA
SUSTENTÁVEL



SAÚDE E
BEM-ESTAR



EDUCAÇÃO DE
QUALIDADE



IGUALDADE
DE GÊNERO



ÁGUA POTÁVEL
E SANEAMENTO



ENERGIA LIMPA
E ACESSÍVEL



TRABALHO DECENTE
E CRESCIMENTO
ECONÔMICO



INDÚSTRIA,
INOVAÇÃO E
INFRAESTRUTURA



REDUÇÃO DAS
DESIGUALDADES



CIDADES E
COMUNIDADES
SUSTENTÁVEIS



CONSUMO E
PRODUÇÃO
RESPONSÁVEIS



AÇÃO CONTRA A
MUDANÇA GLOBAL
DO CLIMA



VIDA NA
ÁGUA



VIDA
TERRESTRE



PAZ, JUSTIÇA E
INSTITUIÇÕES
EFICAZES



PARCERIAS
E MEIOS DE
IMPLEMENTAÇÃO



COMISSÃO PERMANENTE DE ACOMPANHAMENTO DOS OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL E DA AGENDA 2030

ESTATÍSTICA APLICADA A DADOS

Atos normativos estruturantes



BRASÍLIA/DF
2021

CNU CONSELHO
NACIONAL
DE JUSTIÇA

**CONSELHO NACIONAL DE JUSTIÇA****Presidente**

Ministro Luiz Fux

Corregedora Nacional de Justiça

Ministra Maria Thereza Rocha de Assis Moura

Conselheiros

Ministro Emmanoel Pereira
Luiz Fernando Tomasi Keppen
Mário Augusto Figueiredo de Lacerda Guerreiro
Rubens de Mendonça Canuto Neto
Candice Lavocat Galvão Jobim
Tânia Regina Silva Reckziegel
Flávia Moreira Guimarães Pessoa
Ivana Farina Navarrete Pena
André Luis Guimarães Godinho
Marcos Vinícius Jardim Rodrigues
Maria Tereza Uille Gomes
Luiz Fernando Bandeira de Mello Filho

Secretário-Geral

Valter Shuenquener de Araujo

**Secretário Especial de Programas,
Pesquisas e Gestão Estratégica**

Marcus Livio Gomes

Diretor-Geral

Johaness Eck

SECRETARIA DE COMUNICAÇÃO SOCIAL**Secretária de Comunicação Social**

Juliana Neiva

Projeto gráfico

Virgínia Gomes

Diagramação

Vinícius de Medeiros

Revisão

Carmem Menezes

2021

CONSELHO NACIONAL DE JUSTIÇA
SAF SUL Quadra 2 Lotes 5/6 - CEP: 70070-600
Endereço eletrônico: www.cnj.jus.br

**COMISSÃO PERMANENTE DE ACOMPANHAMENTO
DOS OBJETIVOS DE DESENVOLVIMENTO
SUSTENTÁVEL E DA AGENDA 2030**

ESTATÍSTICA APLICADA A DADOS
Atos normativos estruturantes

BRASÍLIA/DF
2021



CONSELHO NACIONAL DE JUSTIÇA

EXPEDIENTE INTERNO

COMISSÃO PERMANENTE DE ACOMPANHAMENTO DOS OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL E A AGENDA 2030

Conselheira Maria Tereza Uille Gomes (Presidente)

Conselheira Flávia Moreira Guimarães Pessoa

Conselheiro Henrique de Almeida Ávila

LABORATÓRIO DE INOVAÇÃO, INTELIGÊNCIA E OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL (LIODS)

Coordenadora

Conselheira Maria Tereza Uille Gomes

INTEGRANTES LIODS – PORTARIA Nº 124/2019

Secretaria Especial de Programas, Pesquisas e Gestão Estratégica – SEP

Representante da Corregedoria Nacional de Justiça

Departamento de Monitoramento e Fiscalização do Sistema Carcerário

e do Sistema de Execução de Medidas Socioeducativas – DMF

Departamento de Pesquisas Judiciárias – DPJ

Diretora do Departamento de Gestão Estratégica – DGE

Departamento de Tecnologia da Informação e Comunicação – DTI

Departamento de Acompanhamento Orçamentário – DAO

Centro de Formação e Aperfeiçoamento de Servidores

do Poder Judiciário – CEAJUD

Laboratório de Inovação do Poder Judiciário – iJuspLab

Centro Nacional de Inteligência da Justiça Federal

Conselho Nacional do Ministério Público – CNMP

Gabinete da Coordenação do LIODS

COORDENAÇÃO-GERAL DA PESQUISA E PUBLICAÇÃO

Conselheira Maria Tereza Uille Gomes

REDAÇÃO DO CADERNO

Ana Paula Garutti

Paula Ferro Costa de Sousa

EQUIPE DO GABINETE CONSELHEIRA MARIA TEREZA UILLE GOMES & LIODS

Paula Ferro Costa de Sousa – Assessora-Chefe de Gabinete

Jorge Henrique Mendes – Assessor

Angela Maria dos Santos – Assistente VI

Ana Paula de Melo Soares – Secretária

Ana Paula Garutti – Servidora

Fernando Pinheiro Gomes – Servidor

COOPERAÇÃO CNJ/PNUD

Victor Nabhan – Assistente-técnico

Allan Canuto de Melo – Assistente-técnico

Raquel Lasalvia Correia – Assistente-técnico

Clara Wardi – Assistente-técnico

Gustavo Fecundo Nino – Estatístico

Sumário

APRESENTAÇÃO	7
1 INTRODUÇÃO.....	9
2 ATOS NORMATIVOS DO CNJ – ESTATÍSTICA E TECNOLOGIA.....	11
2.1 Resolução CNJ nº 46, de 18/12/2007	11
2.2 Resolução CNJ nº 49 de 18/12/2007	12
2.3 Resolução CNJ nº 65 de 16/12/2008	13
2.4 Resolução CNJ nº 76 de 12/05/2009	13
2.5 Resolução CNJ nº 215 de 16/12/2015	14
2.6 Provimento CNJ nº 63 de 14/11/2017	14
2.7 Resolução CNJ nº 331 de 20/08/2020:	15
2.8 Resolução CNJ nº 333 de 21/09/2020:	17
3 HISTÓRIA DA ESTATÍSTICA.....	19
4 ESTATÍSTICA.....	21
4.1 Estatística Descritiva e Inferencial.....	21
4.2 Pesquisa e Coleta de Dados	21
4.3 Classificação dos Dados	22
4.4 Medidas de Posição	30
4.5 Medidas de Dispersão	33
4.6 Correlação	34
4.7 Amostragem	35
5 BIBLIOGRAFIA	38

Apresentação

O Poder Judiciário, no Brasil, tem realizado um trabalho inédito de institucionalização da Agenda 2030, das Nações Unidas, no escopo de sua atuação. A Agenda 2030 é um plano de ação que reúne esforços de vários países signatários da Resolução A/RES/72/279, da Organização das Nações Unidas, dentre eles o Brasil, para garantir a sustentabilidade para as pessoas e o planeta.

O plano de ação global possui 17 Objetivos de Desenvolvimento Sustentável (ODS) e 169 metas, que servem de diretrizes para a atuação de todos os órgãos envolvidos nessa missão. O Conselho Nacional de Justiça (CNJ), enquanto órgão de controle da atuação administrativa e financeira do Poder Judiciário e que possui, entre suas competências, a elaboração de relatórios estatísticos pertinentes à atividade jurisdicional (BRASIL, 1988, Art. 103-B; CONSELHO NACIONAL DE JUSTIÇA, 2009, Art. 4º), coordena o projeto de integração das metas e indicadores da Agenda 2030 no âmbito do Poder Judiciário brasileiro.

O CNJ tem construído um caminho profícuo na institucionalização da Agenda 2030 desde a instituição do Comitê Interinstitucional destinado a avaliar a integração das metas do Poder Judiciário às metas e indicadores dos Objetivos de Desenvolvimento Sustentável (ODS), Agenda 2030, e elaborar relatório de trabalho com apoio de todos os Tribunais do País (Portaria nº 133, de 28/09/2018). Nessa esteira, foram instituídos também o Laboratório de Inovação, Inteligência e ODS (LIODS) e a Comissão Permanente de Acompanhamento dos ODS e da Agenda 2030, que o coordena.

A Comissão da Agenda 2030, com o apoio do LIODS, tem desenvolvido parcerias com os laboratórios de inovação e centros de inteligência dos tribunais brasileiros para atuação direta na pauta da Agenda 2030 e na proposição de políticas judiciárias para promoção do desenvolvimento sustentável. Além das parcerias com os tribunais, o LIODS apoia as ações relacionadas aos temas adotados no Observatório Nacional sobre Questões Ambientais, Econômicas e Sociais de Alta Complexidade e Grande Impacto e Repercussão. As ações planejadas podem envolver parcerias externas ao Poder Judiciário também, ampliando o escopo dos debates e soluções propostas.

Para apoiar o desenvolvimento dos projetos que contam com a participação do LIODS, o CNJ assinou Acordo de Cooperação Técnica com Programa das Nações Unidas para o Desenvolvimento - PNUD, cujo objeto é a consolidação dos Laboratórios de Inovação e Inteligência no âmbito do Poder Judiciário e fortalecimento das capacidades do CNJ para a produção de pesquisas em temas relacionados à Agenda 2030.

No escopo desse Acordo, foi realizada a contratação de uma equipe de apoio para organização e registro das atividades, e para o levantamento e tratamento de dados utilizados no escopo de alguns projetos. Contratamos, também, o desenvolvimento de um curso para formação de magistrados e servidores que atuam na rede de laboratórios e centros de inteligência do Poder Judiciário, promovendo o conhecimento institucional, a inovação e a cooperação.

Esta série de Relatórios Temáticos registram os principais projetos desenvolvidos pela Comissão da Agenda 2030, e têm como objetivos, dentre outros, o de estimular a inovação, o desenvolvimento de parcerias e do trabalho cooperativo na identificação e na busca de possíveis soluções para os problemas estudados. Os Relatórios mencionam os nomes dos laboratórios e dos colaboradores que participaram do processo, descrevendo a dinâmica

de organização do trabalho e os resultados obtidos. Desta forma, não é uma publicação que tem como objetivo o rigor metodológico na análise de dados quantitativos ou qualitativos, e não é uma publicação científica.

Dentre os principais temas publicados destacamos a atuação na temática indígena, na pandemia do novo coronavírus, nos problemas ocorridos com barragens de mineração e no processo de institucionalização da Agenda 2030 no Poder Judiciário por meio da Meta Nacional 9.



Maria Tereza Uille Gomes

Presidente da Comissão Permanente de Acompanhamento dos ODS e da Agenda 2030

1 Introdução

A Estatística é uma ciência cujo campo de aplicação estende-se a diversas áreas do conhecimento humano. Ela está presente em várias etapas de uma pesquisa para compreensão de algum problema ou fenômeno. Está presente na coleta, apresentação, interpretação dos dados, sejam eles qualitativos ou quantitativos. Seu objetivo é apresentar informações sobre dados de uma pesquisa ou um levantamento para que se tenha maior compreensão dos fatos. Qualquer hipótese, qualquer ideia, pode ser provada por meio da estatística, bastando termos o dado de forma correta para analisá-lo.

Para se obter bons resultados numa análise estatística, além dos métodos aplicados, também é necessário ter clareza nos conceitos utilizados. A seguir são apresentados alguns desses conceitos, mas, mais importante que isto, veremos a relevância da coleta dos dados, do hábito de se controlar de maneira estruturada ou tabulada as informações que produzimos, bem como, onde e como podemos ter acesso aos dados do Poder Judiciário.

Neste caderno, antes de entrarmos nestas questões de estatística, iremos tratar da evolução dos atos normativos relativos às estatísticas no Poder Judiciário, dentre eles os relacionados às Tabelas Processuais Unificadas do Poder Judiciário (TPU), núcleos de estatística e gestão estratégica no Poder Judiciário e o Número Único do Processo (NUP).

2 Atos normativos do CNJ – Estatística e Tecnologia

Neste estudo iremos tratar de sete Resoluções, uma Portaria e um Provimento, todos do CNJ, que tratam do tema estatísticas e tecnologias e que são importantes para a tomada de decisões corretas, sempre baseadas em evidências.

2.1 Resolução CNJ nº 46, de 18/12/2007

Cria as Tabelas Processuais Unificadas do Poder Judiciário e dá outras providências.

TPU do Poder Judiciário

O Direito é dividido em vários ramos e existe uma taxonomia própria que o classifica em Direito Público, Direito Privado, Direito Canônico e Direito Romano. Dentro do Direito Público temos: Direito Internacional Público, Direito Constitucional, Direito Administrativo, Direito Processual, Direito Penal, Direito Previdenciário, Direito Militar e Direito Aéreo. Já o Direito Privado abrange Direito Civil, Direito Comercial, Direito Internacional Privado, Direito do Consumidor, Direito do Trabalho¹.

No âmbito do Poder Judiciário, as Tabelas Processuais Unificadas foram instituídas pela Resolução CNJ nº 46, de 18/12/2007, e observada por mais de 90 tribunais.

Elas podem ser vistas no site do Sistema de Gestão das Tabelas Processuais Unificadas², onde também está disponibilizado o Manual correspondente.

A TPU é organizada em grandes eixos:

- Classes
- Movimentos
- Assuntos

Um dos eixos diz respeito à escolha dos Assuntos. Neste aspecto, a TPU se organiza por meio de níveis. O nível 1 simboliza a escolha pelo ramo do direito, são eles:

- Direito Administrativo e Outras Matérias de Direito Público;
- Direito Assistencial; Direito Civil;
- Direito da Criança e do Adolescente;
- Direito da Saúde;

¹ <http://www4.planalto.gov.br/centrodeestudos/assuntos/classificacao-decimal-de-direito/classif-decimal.pdf>

² https://www.cnj.jus.br/sgt/consulta_publica_classes.php

- Direito do Consumidor;
- Direito do Trabalho;
- Direito Eleitoral;
- Direito Eleitoral e Processo Eleitoral do STF;
- Direito Marítimo;
- Direito Internacional;
- Direito Penal; Direito Penal Militar;
- Direito Previdenciário;
- Direito Processual Civil e do Trabalho;
- Direito Processual Penal;
- Direito Processual Penal Militar;
- Direito Tributário;
- Questões de Alta Complexidade, Grande Impacto e Repercussão; e,
- Registros Públicos.

Do nível 2 ao 6 são feitas as desagregações dentro do mesmo ramo do direito.

São mais de 3.500 assuntos considerando do nível 1 ao 6.

Cada um destes assuntos foi vinculado com o Objetivo do Desenvolvimento Sustentável da Agenda 2030. Essa referência pode ser acessada no Caderno Acompanhando a Agenda 2030 no Poder Judiciário³.

Dada essa correlação, recentemente, a Comissão Permanente da Agenda 2030 encaminhou solicitação à Secretaria Especial de Pesquisas, Programas e Gestão Estratégica para incluir um ponto de consulta sobre os ODS no Sistema de Gestão das Tabelas Processuais Unificadas, para que seja possível identificar os Assuntos relacionados a determinado ODS. Essa medida está alinhada ao processo de institucionalização da Agenda 2030 no Judiciário, bem como auxiliará os Tribunais no cumprimento da Meta 9.

2.2 Resolução CNJ nº 49 de 18/12/2007

Dispõe sobre a organização de Núcleo de Estatística e Gestão Estratégica nos órgãos do Poder Judiciário relacionados no Art. 92, incisos II ao VII, da Constituição da República Federativa do Brasil.

³ <https://www.cnj.jus.br/wp-content/uploads/2020/07/Acompanhando-a-Agenda-2030-3-1.pdf>

NÚCLEOS DE ESTATÍSTICA E GESTÃO ESTRATÉGICA

Esta Resolução surgiu devido à necessidade de estabelecer núcleos de estatística e gestão estratégica que cooperem para o pleno funcionamento do Sistema de Estatística do Poder Judiciário.

Ela prevê a necessidade de os tribunais criarem estrutura administrativa, em caráter permanente, para elaboração das estatísticas e plano de gestão estratégica.

2.3 Resolução CNJ nº 65 de 16/12/2008

Dispõe sobre a uniformização do número dos processos nos órgãos do Poder Judiciário e dá outras providências.

Número único do processo - NUP

Esta Resolução, com o objetivo de melhorar a administração da justiça e a prestação jurisdicional, definiu padrões de interoperabilidade, entre eles a padronização do número dos processos, que possui a seguinte estrutura:

NNNNNNN-DD.AAAA.J.TR.0000

- O campo (NNNNNNN), com 7 (sete) dígitos, identifica o número sequencial do processo por unidade de origem (0000), a ser reiniciado a cada ano;
- O campo (DD), com 2 (dois) dígitos, identifica o dígito verificador;
- O campo (AAAA), com 4 (quatro) dígitos, identifica o ano do ajuizamento do processo;
- O campo (J), identifica o órgão ou segmento do Poder Judiciário;
- O campo (TR), com 2 (dois) dígitos, identifica o tribunal do respectivo segmento do Poder Judiciário e, na Justiça Militar da União, a Circunscrição Judiciária;
- O campo (0000), com 4 (quatro) dígitos, identifica a unidade de origem do processo.

2.4 Resolução CNJ nº 76 de 12/05/2009

Dispõe sobre os princípios do Sistema de Estatística do Poder Judiciário (SIESPJ), estabelece seus indicadores, fixa prazos, determina penalidades e dá outras providências.

Sistema de Estatística do Poder Judiciário (SIESPJ)

Considerando a importância das estatísticas para fundamentar decisões em matéria de políticas públicas, o SIESPJ foi criado e é regido pelos princípios da publicidade, eficiência, transparência, obrigatoriedade de informação dos dados estatísticos e presunção de veracidade dos dados estatísticos.

A padronização das estatísticas é de extrema importância, mas já vimos que com a Resolução CNJ nº 46/2007 este passo foi iniciado.

A forma de coleta dos dados estatísticos do SIESPJ está prevista na Resolução CNJ nº 331/2020 – DataJud, que veremos mais adiante.

Nesta Resolução também temos a menção da Comissão Permanente de Gestão Estratégica, Estatística e Orçamento que tem como um de seus objetivos o exercício das funções de orientação e monitoramento do SIESPJ.

2.5 Resolução CNJ nº 215 de 16/12/2015

Dispõe, no âmbito do Poder Judiciário, sobre o acesso à informação e a aplicação da Lei 12.527, de 18 de novembro de 2011.

ACESSO À INFORMAÇÃO

Nesta Resolução temos a regulamentação do Poder Judiciário frente à LAI (Lei de Acesso à Informação - Lei nº 12.527/2011). A publicidade é um dos princípios fundamentais regentes da administração pública, compreendendo a transparência, a acessibilidade, a integralidade e a integridade das informações referentes à gestão da coisa pública, e a Resolução CNJ nº 215/2015 vem ao encontro dessa temática.

2.6 Provimento CNJ nº 63 de 14/11/2017

Institui modelos únicos de certidão de nascimento, de casamento e de óbito, a serem adotadas pelos órgãos de registro civil das pessoas naturais, e dispõe sobre o reconhecimento voluntário e a averbação da paternidade e maternidade socioafetiva no Livro “A” e sobre o registro de nascimento e emissão da respectiva certidão dos filhos havidos por reprodução assistida.

REGISTRO CIVIL

Com a fixação de um modelo único dos registros civis, a capacidade de transformação em um banco de dados, de forma automática, com inteligência artificial, passa a ser possível. E, a partir do momento que temos os dados organizados em um banco, é possível vários cruzamentos serem feitos.

Além disso, essa norma traz a obrigatoriedade de inclusão do CPF nos registros.

2.7 Resolução CNJ nº 331 de 20/08/2020:

Institui a Base Nacional de Dados do Poder Judiciário – DataJud como fonte primária de dados do Sistema de Estatística do Poder Judiciário – SIESPJ para os tribunais indicados nos incisos II a VII do art. 92 da Constituição Federal.

DATAJUD

O DataJud é uma Base de Dados alimentada com dados e metadados processuais relativos a todos os processos físicos ou eletrônicos, públicos ou sigilosos dos tribunais indicados nos incisos II a VII do art. 92 da Constituição Federal. Estes dados seguem o padrão estabelecido na Resolução CNJ nº 331/2020, e a gestão das informações e atualização dos dados cabe ao Departamento de Tecnologia da Informação do CNJ.

O acesso público aos dados do DataJud é previsto no art. 11 da referida norma:

Art. 11. Ato da Presidência disporá sobre as informações que serão disponibilizadas por meio de API pública para consulta aos metadados do DataJud, resguardados o sigilo e a confidencialidade das informações, nos termos da legislação processual e da Lei Geral de Proteção de Dados.

Parágrafo único. O fornecimento de dados além do estabelecido no ato da Presidência dependerá de requerimento do ente público ou instituição de pesquisa interessada e de termo específico a ser firmado com o CNJ, que conterà cláusula de sigilo e confidencialidade.

Pela Portaria CNJ nº 160/2020, o acesso público deve ocorrer por meio de API que será disponibilizada a partir de agosto de 2021.

PORTARIA Nº 160, DE 9 DE SETEMBRO DE 2020.

Estabelece o cronograma de saneamento da Base Nacional de Dados do Poder Judiciário - DataJud e regulamenta o acesso público aos dados do DataJud por meio de API – Application Programming Interface.

O PRESIDENTE DO CONSELHO NACIONAL DE JUSTIÇA, no uso de suas atribuições legais e regimentais, CONSIDERANDO o disposto nos artigos 11 e 12 da Resolução CNJ nº 331, de 20 de outubro de 2020,

RESOLVE:

Art. 1º Tornar público o cronograma para correção e saneamento de dados constantes do DataJud e definir as informações que serão disponibilizadas por meio de API - Application Programming Interface.

Art. 2º Os tribunais deverão envidar os esforços necessários para correção e saneamento dos dados constantes no DataJud, de acordo com o seguinte cronograma:

I – até 18 de dezembro de 2020, para elaboração de “de-para” ou método similar, de forma que todos os movimentos inseridos no DataJud que são utilizados para o cálculo das variáveis e indicadores constantes dos glosários do Justiça em Números e do Módulo de Produtividade, segundo os Anexos I e II da Resolução CNJ no 76, de 12 de maio de 2009, estejam em conformidade com as Tabelas Processuais Unificadas (TPUs) do CNJ, instituídas pela Resolução CNJ no 46, de 18 de dezembro de 2007;

II – até 3 de fevereiro de 2021, para carga de teste do DataJud referente à correção de que trata o inciso I deste artigo;

III – até 3 de fevereiro de 2021, para carga corretiva, em ambiente de produção, da numeração de processos, em consonância com a Resolução CNJ no 65, de 16 de dezembro de 2008, e dos dados cadastrais de partes que estejam incompletos ou inconsistentes, nos termos apontados em painel a ser disponibilizado pelo CNJ;

IV – até 3 de maio de 2021, para elaboração de “de-para” ou método similar, de forma que todos os assuntos inseridos no DataJud estejam em consonância com as TPUs, classificados em assuntos de último nível e em conformidade com as regras negociais e com painel a ser disponibilizado pelo CNJ;

V – até 3 de junho de 2021, para elaboração de “de-para” ou método similar, de forma que todos os movimentos estejam em consonância com as TPUs, classificados em movimentos de último nível nacional e acompanhados dos complementos vinculados, quando aplicáveis;

VI – até 7 de julho de 2021, para carga de teste do DataJud referente às correções de que tratam os incisos IV e V deste artigo;

VII – até 31 de julho de 2021, para carga completa no DataJud, com todas as correções efetuadas.

§ 1º A carga de teste será realizada em ambiente de homologação, a ser disponibilizado pelo CNJ, e abrangerá todos os processos movimentados no período de janeiro a junho de 2015 e de janeiro a junho de 2019.

§ 2º A carga completa será realizada em ambiente de produção, contendo todos os processos em tramitação e os que tenham sido baixados a partir de 1º de janeiro de 2015.

Art. 3º Os tribunais deverão observar a integridade e a validação dos dados, conforme Modelo de Transmissão de Dados (MTD) em vigor, sem prejuízo da inclusão de outras etapas de saneamento não previstas nesta Portaria.

Art. 4º a API pública conterá os seguintes dados, segundo o MTD:

I – número do processo;

II – sigla do tribunal atual;

III – grau de jurisdição atual;

IV – órgão julgador do processo atual;

V – classe processual atual;

VI – assuntos processuais das tabelas nacionais e assuntos locais atuais;

VII – prioridade;

VIII – procEL – tramitação em sistema eletrônico;

IX – sistema em que tramita;

X – movimentos nacionais e movimentos locais;

XI – complementos dos movimentos nacionais, resguardados os dados das partes;

XII – órgão julgador atrelado ao movimento.

Parágrafo único. A API não conterá os processos que tramitam em segredo de justiça.

Art. 5º A API pública será desenvolvida em até 30 dias, a contar do término do cronograma de saneamento.

Art. 6º Esta Portaria entra em vigor na data de sua publicação.

Ministro DIAS TOFFOLI

2.8 Resolução CNJ nº 333 de 21/09/2020:

Determina a inclusão de campo/espço denominado Estatística na página principal dos sítios eletrônicos dos órgãos do Poder Judiciário indicados nos incisos I-A a IV, VI e VII do art. 92 da Constituição Federal e dá outras providências.

CAMPO ESTATÍSTICA NOS SITES DOS TRIBUNAIS

Além do necessário acesso à informação por parte do cidadão às atividades dos órgãos públicos, a facilidade de encontrá-los é um item muito importante. Para isso surge a Resolução CNJ nº 333/2020.

Ela prevê a reunião de dados abertos, painéis de business intelligence e relatórios estatísticos em um campo só no site dos tribunais.

Diante destas informações de onde encontramos dados sobre a atuação do Poder Judiciário, vamos agora falar um pouco de como e o que fazer com estes dados.

3 História da estatística

História da estatística e seus precursores

A palavra “Estatística” vem de “Status” (Estado em latim). Desde o começo foi utilizada como uma ferramenta administrativa pelos estadistas. Começou a ser utilizada por Gottfried Achenwall, acadêmico alemão, em meados do século XVIII, aparecendo pela primeira vez como verbete em uma enciclopédia apenas em 1797 na Inglaterra.

Apesar da origem da palavra ser do século XVIII, há vestígios de que a utilização da estatística vem de 3.000 anos A.C., quando Babilônios, Chineses e Egípcios realizavam censos pensando em tarifas e impostos.

Na própria Bíblia, no livro Números do velho testamento, há orientações para Moises levantar a quantidade de homens de Israel aptos à guerra.

Em 1085, Guilherme, O Conquistador, com interesses tributários, ordenou a realização de um levantamento na Inglaterra de informações sobre proprietários de terras, empregados, animais e uso da terra.

No século XVII, também na Inglaterra, destacou-se a Aritmética Política de John Graunt, com estatísticas de nascimentos e mortes, dando origem às tábuas de mortalidade usadas até hoje por companhias de seguros entre outros.

No final do século XVII, a contribuição veio com Jacob Bernoulli com cálculos de probabilidade.

No século XVIII, Thomas Bayes estabeleceu parâmetros para a Inferência Probabilística (meios de se calcular a probabilidade de um evento passado acontecer novamente no futuro). No mesmo século, Karl Friedrich Gauss descobriu leis importantes, como por exemplo, a média aritmética e os mínimos quadrados.

No século XIX, Karl Person, aprimora teorias tais como: provas de significância, correlação, iniciando a Teoria de Inferência Estatística. Destaque também para teorias de probabilidade de Siméon-Denis Poisson. Florence Nightingale passa a utilizar a estatística nas Ciências Sociais, reduzindo a taxa de mortalidade por meio de controles estatísticos de higiene, consumo de medicamentos etc, criando o polígono de frequências e o histograma.

Também no século XIX, Francis Galton, primo de Charles Darwin, modela a Análise de Regressão.

Já no século XX, William Sealey Gosset, com o pseudônimo Student, teoriza sobre a probabilidade do erro médio, sobre o erro padrão e sobre o coeficiente de correlação.

Sir Ronald Aylmer Fisher colaborou com Técnicas de Amostragem e Testes de Hipóteses e Jerzy Neyman com a Teoria de estimação e Intervalos de Confiança. Em 1988 com a publicação do livro Company-Wide Total Quality Control de Shigeru Mizuno, que trouxe o conceito de Gestão da Qualidade Total que o controle de dados passa a ser levado a sério, passando posteriormente a ser considerado em diversas áreas do conhecimento.

Com o advento da internet e a velocidade em se obter informações, surge a era do Data Science. Passa-se a ter computadores robustos, novas técnicas de extração de dados (mineração de texto) e consequentemente o conceito de Big-Data.

4 Estatística

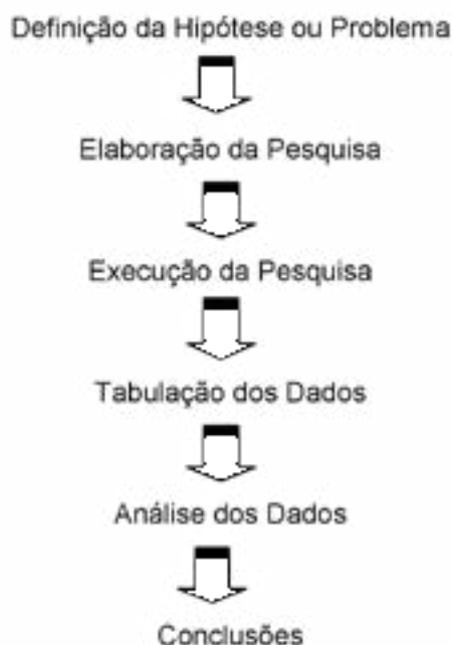
A Estatística é uma Ciência voltada para a obtenção, organização e análise de dados. Porém, para chegarmos a conclusões válidas todo este processo de coleta, organização e análise de dados também devem ser legítimos. Para tanto, necessário se faz conhecer algumas técnicas básicas.

4.1 Estatística Descritiva e Inferencial

O ato de tratar o dado, organizando-o e apresentando-o por meio de tabelas e gráficos, possibilitando assim compreender e tirar conclusões, é chamado de Estatística Descritiva. A Estatística Inferencial tem como objetivo tirar conclusões sobre uma população por meio de uma amostra, como por exemplo, as pesquisas eleitorais, onde parte dos eleitores são questionados, e com uma margem de erro, pode-se afirmar o resultado final. A Estatística Inferencial engloba diversos tipos de testes estatísticos, bem como análises mono ou multivariadas.

4.2 Pesquisa e Coleta de Dados

O primeiro passo para uma pesquisa é a definição do problema a ser pesquisado ou hipótese a ser comprovada. Por exemplo, digamos que queremos estudar alguns aspectos sobre a satisfação dos servidores do Poder Judiciário Brasileiro. Para tanto necessitamos realizar um interrogatório a alguns servidores do Poder Judiciário, obtendo assim alguns dados específicos que devem ser tabulados para se chegar a conclusões sobre tal.



A coleta destes dados pode ser feita de maneira primária ou secundária. A primária é quando não temos fontes prontas com o dado que necessitamos para análise, ou seja, teremos que fazer um levantamento destes dados por meio de observação ou por meio de um questionário, este segundo também chamado de survey. A coleta secundária é quando utilizamos dados já coletados e registrados por fontes diversas, como anuários ou como os dados que vimos nas resoluções acima.

4.3 Classificação dos Dados

Dado é o elemento principal para a formação de um juízo. Dependendo do tipo de dado é possível ou não se fazer determinadas estatísticas, por exemplo, quando nosso dado não é numérico, não há como estabelecermos sua média.

Os dados podem ser divididos em dois tipos: quantitativos e qualitativos.

Os quantitativos são dados numéricos e podem ser discretos ou contínuos. Os primeiros resultam de contagens e são sempre números inteiros, ou seja, não possuem casas decimais, exemplo, quantidade de filhos, total de processos de roubos etc. Os dados contínuos podem ser qualquer valor, como por exemplo, estatura dos servidores, tempo de tramitação de um processo etc.

Os dados qualitativos são observações não numéricas expressas por meio de termos ou palavras. Eles podem ser ordinais ou nominais, sendo os primeiros aqueles que possuem uma ordenação ou hierarquia, como por exemplo, 'bom, regular e ruim'. Os nominais não possuem esta hierarquia, como por exemplo, 'sexo do entrevistado', 'estado civil' etc.

Apresentação dos Dados

Após a definição do problema a ser pesquisado, da coleta destes dados e de sua correta tabulação, passamos para a fase de apresentação destes dados. Teremos aqui instruções de como apresentar tabelas, quadros e gráficos.

Tabelas

Uma forma eficiente de apresentar os dados coletados é agrupando-os conforme a frequência de repetição das variáveis, formando assim tabelas.

Base de dados:

A _C ^R situacao	A _C ^R estado	A _C ^R numero_processo	A _C ^R grau	A _C ^R intervencao_mp
Ativo	SP	68426120148260240	G2	NÃO
Baixado	SP	49343720128260050	G2	NÃO
Ativo	SP	423926220178260000	G2	NÃO
Ativo	SP	15005695720198260000	G1	NÃO
Ativo	SP	274463420178260400	G1	NÃO
Ativo	SP	10092069420198260000	G1	NÃO
Ativo	SP	70391220178260570	G1	NÃO
Ativo	SP	2406920188260224	G1	NÃO
Ativo	SP	15201224320198260000	G1	NÃO
Baixado	SP	5701720188260111	G1	NÃO
Ativo	SP	16743620148260062	G2	NÃO
Ativo	SP	18562320184036120	G1	SIM
Ativo	SP	419881820178260200	G1	NÃO
Ativo	SP	203047620178260130	G1	NÃO
Ativo	SP	77174120054036130	G2	NÃO
Ativo	SP	160686120128260030	G1	NÃO
Ativo	SP	68662720148260000	G1	NÃO
Ativo	SP	3262020178260242	G1	NÃO
Ativo	SP	5835320188260616	G1	NÃO
Ativo	SP	15000413820198260000	G1	NÃO

Tabela representando os dados do Banco de Dados acima:

Quantidade de processo sobre agrotóxico do ano de 2019 por UF e Situação

UF	Ativo	Baixado	Total
ES	267	62	327
PR	178	64	240
RS	139	71	209
MT	122	37	157
SP	107	15	121
SC	63	23	85
MS	69	2	70
DF	40	19	59
MA	22	11	30
PA	21	3	24
AL	10	11	21
CE	17	3	20
SE	19		19
PE	10	2	12
RO	7	5	10
PI	8	1	9
RR	6	3	9
AP	4	3	7
AM	2	2	4
AC	1	1	2
Total	1111	338	1434

Fonte: DATAJUD

Uma tabela é formada por título, cabeçalho e corpo.

Título: localizado no topo da tabela, ele deve responder ‘o que’, ‘quando’ e ‘onde’ sobre os dados apresentados;

Cabeçalho: indica o conteúdo das colunas;

Corpo: é o conjunto de linhas e colunas que possuem os dados apresentados.

Outro ponto essencial em uma tabela é a fonte de origem daqueles dados.

TÍTULO → Crimes Ambientais - Ano 2020 - Quantidade de Processos por TJ

Tribunal	Quantidade
TJSP	4821
TJMG	2486
TJRO	1430
TJSC	1062
TJPR	794
TJMT	731
TJRR	667
TJDA	129
Total	15493

CABEÇALHO ←

CORPO {

RODAPÉ → Fonte: Datajud

Toda tabela deve ser autoexplicativa. Ela deve ter bordas horizontais que delimitam seu cabeçalho e outra que faz seu fechamento em baixo. Outras bordas horizontais ou verticais somente devem ser utilizadas para facilitar a visualização dos dados. Nas laterais das tabelas, nunca utilizar estas bordas. Estas normas são do Instituto Brasileiro de Geografia e Estatística (IBGE) e são adotadas pela Associação Brasileira de Normas Técnicas (ABNT).

Quadro

Os quadros se diferenciam das tabelas por não apresentarem dados numéricos, ou seja, seus dados são exclusivamente qualitativos.

Status de CPFs de Aux. Emergencial por Tribunal - Ano 2020

Tribunal	Status
TRF1	ELEGIVEL
TRF1	ELEGIVEL JUDICIAL
TRF1	Inconclusivo
TRF1	INELEGIVEL
TRF1	Registro em processamento
TRF1	Registro em Reanalise
TRF1	Registro na base de retidos
TRF1	Registro nao encontrado
TRF2	ELEGIVEL
TRF2	ELEGIVEL JUDICIAL
TRF2	FALSE
TRF2	Inconclusivo
TRF2	INELEGIVEL
TRF2	Registro em processamento
TRF2	Registro nao encontrado
TRF3	ELEGIVEL
TRF3	ELEGIVEL JUDICIAL
TRF3	FALSE
TRF3	Inconclusivo
TRF3	INELEGIVEL
TRF3	Registro em processamento

Fonte: Tabelas dos TRFs e DATAPREV (caixa GIS do CNJ)

Gráficos

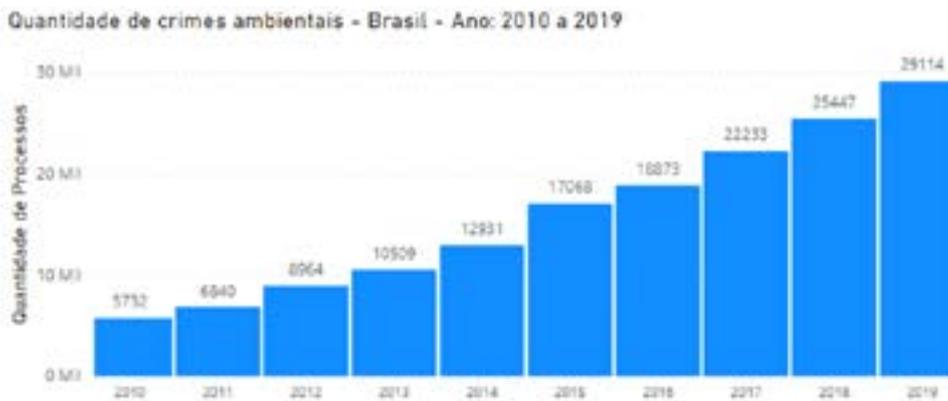
Os gráficos são em geral mais comunicativos que as tabelas. Veremos seis tipos de gráficos.

- Gráficos de colunas

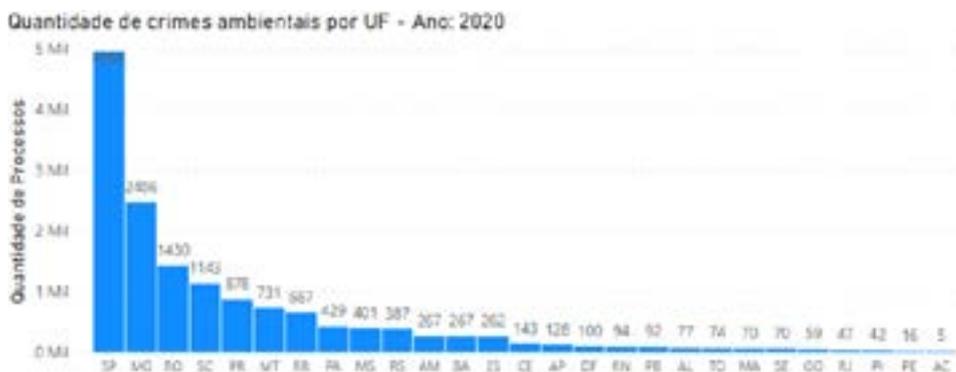
Tendo uma tabela com dados organizados em suas linhas e colunas, podemos representá-los em um gráfico de colunas. Eles são úteis para mostrar mudanças ao longo do tempo ou comparar dados de diferentes localidades, períodos etc.

Exemplos:

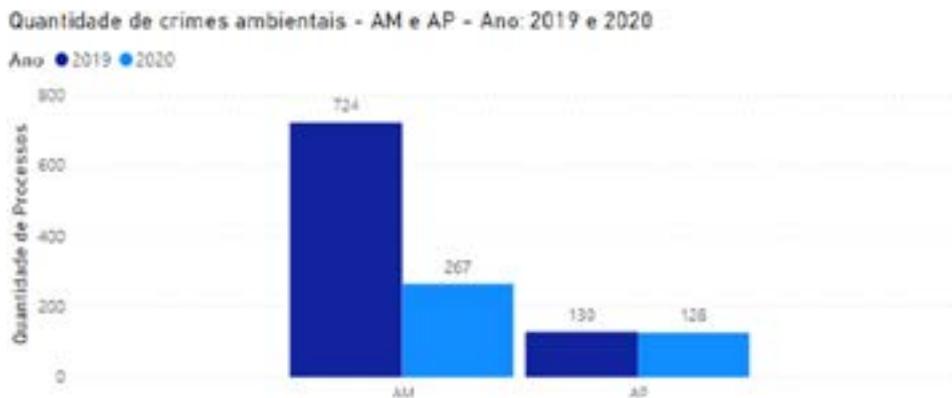
1) Mesma localidade, mesma variável, alterando os anos:



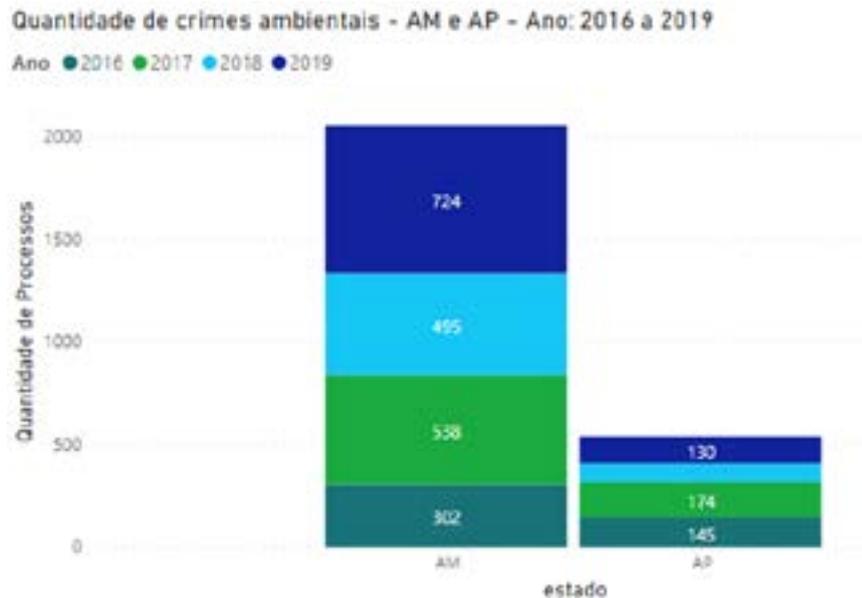
2) Mesmo período, mesmo crime e alterando a localidade:



3) Gráfico com colunas agrupadas: mesma variável, variando período e localidade:



4) Gráfico com colunas empilhadas: Mesma variável, variando período e localidade:



• Gráficos de barras

A principal diferença dos gráficos de colunas para os de barras é que, nos de colunas, os valores são colocados nas colunas (eixo vertical ou y) e nos de barras, eles são colocados nas linhas (eixo horizontal ou x).

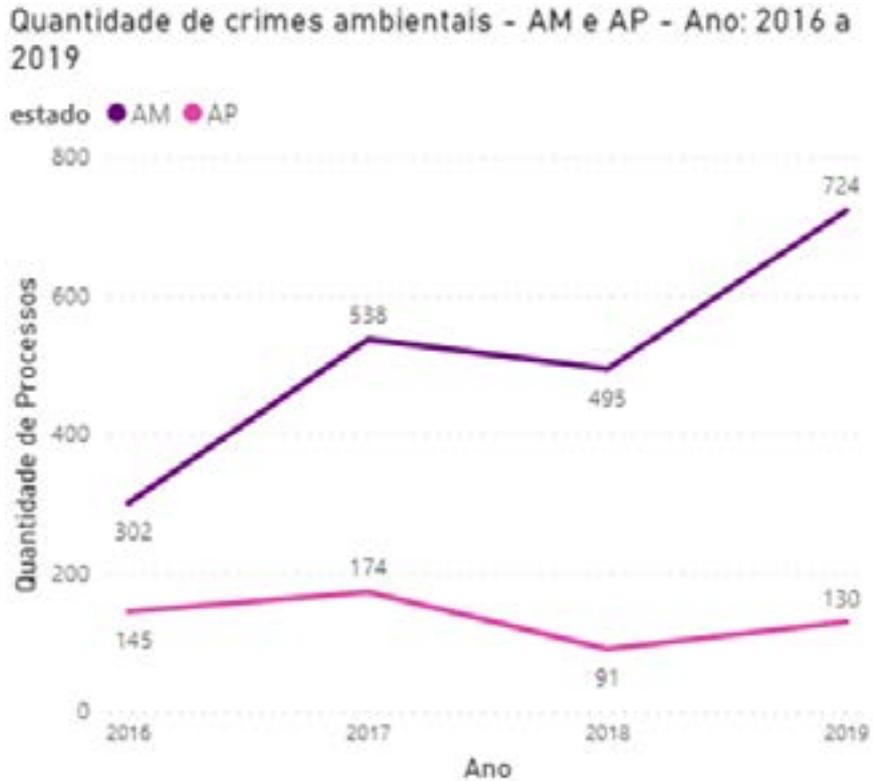
Quantidade de Processos por assunto - Ano 2020



• Gráficos de linhas

Gráficos de linhas são ideais para exibir dados contínuos ao longo do tempo, mostrando tendência e sazonalidade, tendo assim no eixo horizontal dias, meses, trimestres etc. Podem conter uma ou mais linhas, representando itens diferentes ou localidades diferentes.

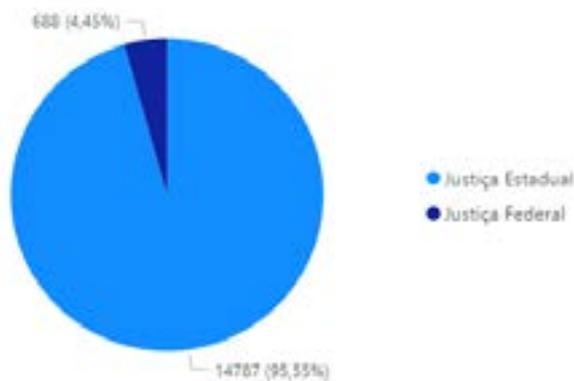
1) Mesma variável, alterando período e localidade:



• Gráficos de pizza

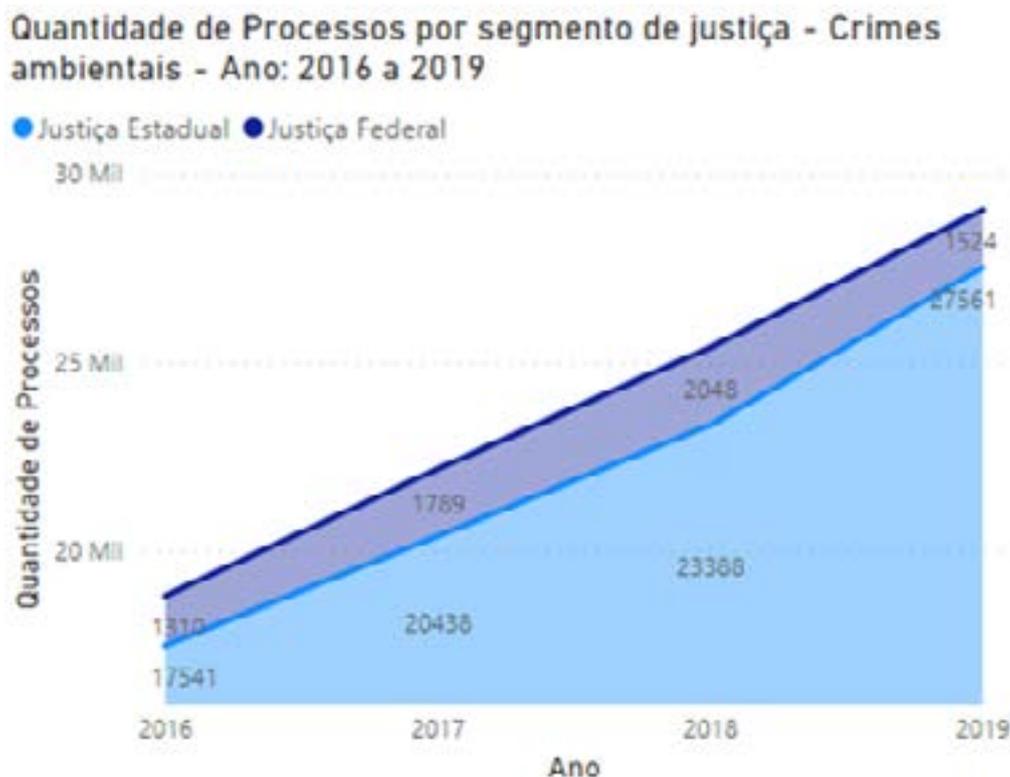
Mostram o tamanho de cada grupo de dados dentro do todo existente na nossa tabela, como por exemplo, quantidade de processos de crimes ambientais distribuídos por Segmento da Justiça. Cada grupo possui uma cor diferente, e seu tamanho proporcional ao que ele representa. Eles não são indicados quando temos valores muito baixos em vários grupos, algum valor negativo ou tenha que ser dividido em mais de cinco partes.

Quantidade de Processos por segmento de justiça - Crimes ambientais - Ano: 2020



• Gráficos de área

Assim como os gráficos de linha, os de área mostram a mudança no decorrer do tempo, enfatizando também a diferença de valores existentes entre os grupos, mostrando a relação das partes para o todo.



Coeficientes, índices e taxas

Quando comparamos um determinado grupo que foi retirado de um grupo maior, estamos calculando um coeficiente, por exemplo, de uma sala de 30 alunos, 10 são reprovados, ou seja, $10/30=0,33$, temos que 33% dos alunos reprovaram. 0,33 é o Coeficiente de Reprovação.

Quando temos essa comparação, porém com dados de grupos distintos, temos um índice. Exemplo: em um município de 2.500 habitantes e uma quantidade de 10 servidores do Poder Judiciário, temos um índice de 250 habitantes por servidor.

Uma outra medida muito utilizada é a taxa. Geralmente para se comparar localidades entre si, como por exemplo municípios entre si. Sempre que multiplicamos um coeficiente ou um índice por 10,100, 1.000 etc, o resultado é uma taxa.

Ano	HD	Pop Estado	Tx 100 mil hab
1995	9814	33848251	28.99
1996	10447	34451927	30.32
1997	10567	35062867	30.14
1998	11861	35698511	33.23
1999	12818	36346903	35.27
2000	12638	36974378	34.18
2001	12475	37542521	33.23
2002	11854	38123695	31.09
2003	10953	38718301	28.29
2004	8753	39326776	22.26
2005	7076	39949487	17.71
2006	6057	40484029	14.96
2007	4877	41029414	11.89
2008	4426	41139672	10.76
2009	4563	41633802	10.96
2010	4320	42136277	10.25

Fonte: Pop: SEADE E IBGE

Dado Criminal: SSP/CAP

4.4 Medidas de Posição

As medidas de tendência central ou medidas de posição são valores que resumem nosso grupo de dados, elas demonstram onde a maioria de nossos dados se concentram. As medidas de posição mais importantes são a média aritmética, a mediana, a moda e os quartis.

Média Aritmética

A Média Aritmética é calculada pela soma de todos os valores dividido pela quantidade de valores, por exemplo, na tabela abaixo temos a quantidade de terras indígenas por TRF. A média de terras indígenas nos TRFs será a soma da quantidade de terras dividido por cinco.

Qtd de Terras Indígenas por TRF	
TRF	Qtd
TRF1	343
TRF3	46
TRF4	36
TRF5	13
TRF2	6
Total	444

Média = 88.8

Fonte: LIODS/CNJ

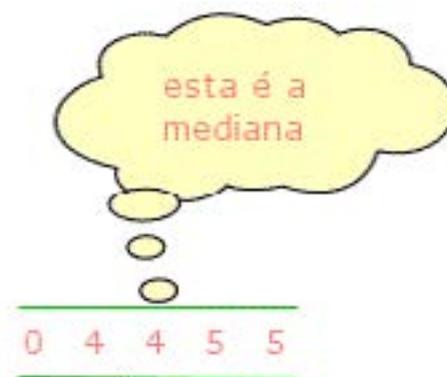
Assim, chegamos à conclusão de que existem, em média, 88,8 terras indígenas nos TRFs. Isso significa que a maioria dos dados estão em torno do valor 89, uns para mais, outros para menos.

O símbolo da média é um 'x' com uma barra em cima: \bar{x}

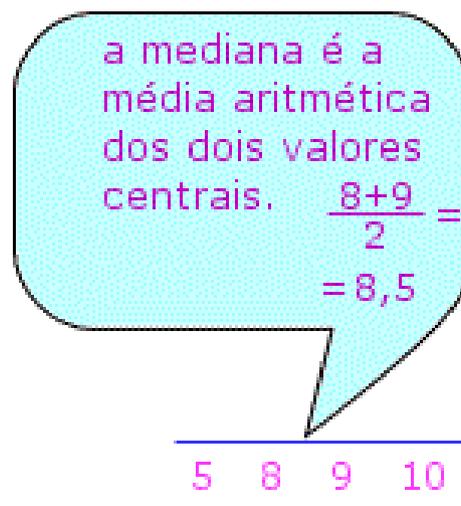
Um problema da média é que seu valor é muito influenciado por valores muito grandes ou muito pequenos, podendo apresentar uma imagem distorcida dos dados, sendo nestes casos mais indicado a utilização da Mediana. Veremos um pouco mais sobre isso na parte de Medidas de Dispersão.

Mediana

A mediana é exatamente o valor do meio do nosso grupo de dados. Para encontrá-la primeiramente devemos colocar nossos dados em ordem crescente e localizar o que ocupa a posição do meio. Em um rol de dados que possua uma quantidade ímpar de casos, fica fácil encontrar a posição do meio, por exemplo, em um rol (já em ordem crescente) de cinco posições, identificaremos a mediana na terceira posição, ficando dois valores para a esquerda e dois para a direita.



Já um rol de dados com quantidade de posições pares, devemos localizar os dois valores do meio e extrair a média destes dois.



A mediana não é influenciada por valores muito baixos ou muito altos como a média, pois ela depende da posição e não dos valores dos elementos do grupo como acontece com a média. No caso anterior das terras indígenas por TRF, temos que a mediana é de 36.

Qtd de Terras Indígenas por TRF

TRF	Qtd
TRF1	343
TRF3	46
TRF4	36
TRF5	13
TRF2	6
Total	444

Mediana = 36

Fonte: LIODS/CNJ

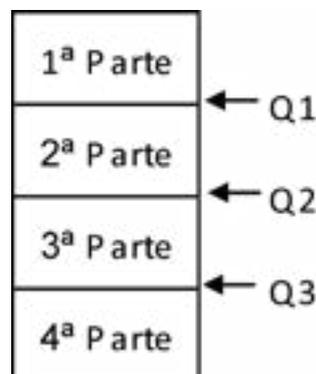
Moda

A moda é o valor que aparece com maior frequência na série de dados. É útil para casos em que temos apenas dados qualitativos, não podendo haver o cálculo da média ou mediana. A série pode ser amodal (não ter moda), ou seja, nenhum item se repete, ou bimodal, quando mais de um item se repete na mesma quantidade.

Resumindo: a moda é o valor com ocorrência mais frequente, a mediana é o valor do meio e a média é o valor médio.

Quartis

Os quartis são pontos numéricos que dividem os valores de um grupo de dados em quatro partes iguais, ou seja, são três valores, identificados por Q1, Q2 e Q3, dividindo o grupo em quatro partes.



Como os quartis dividem os valores do grupo, e não a quantidade de elementos do grupo, podemos ter as quatro partes com quantidade diferente de elementos.

Os quartis são muito úteis para verificarmos a existência de dados muito discrepantes, os outliers, que veremos mais adiante.

4.5 Medidas de Dispersão

Além do conhecimento da média ou mediana do grupo analisado, é necessário termos conhecimento dos desvios dos nossos dados em relação a esta medida central, veremos alguns deles.

Variância, Desvio Padrão e Coeficiente de Variação

Considere que a média é uma medida que represente os valores de um conjunto de dados que temos, porém notamos que há valores dentro deste grupo bem diferentes do apresentado pela média. Para verificar esta variação, esta diferença entre o que cada valor realmente é e o que a média mostra, podemos recorrer ao Desvio Padrão. Ele mede este espalhamento dos valores em torno da média, portanto, quanto menor seu valor, mais cada valor do grupo se aproxima da média.

Exemplo: os dados a seguir são das notas de cinco alunos de duas Turmas diferentes, podendo essas notas serem de 0 a 10. Verificamos que as duas Turmas possuem a mesma média, porém possuem desvios diferentes:

Notas Turma A	Notas Turma B		
3	5		
4	7		
5	6		
10	7		
8	5		
		Turma A	Turma B
		Média	6
		Desvio Padrão	2.61
			6
			0.89

Podemos afirmar que na Turma A, a maioria das notas estão entre 3,39 e 8,61, ou seja, 2,61 (que é o Desvio Padrão) para mais e para menos da média. Na turma B este intervalo é do 5,11 ao 6,89, ou seja, apesar das duas turmas terem a mesma média, os alunos da turma B tiveram um melhor desempenho.

Porém há casos em que os grupos são diferentes, como por exemplo, um grupo onde a média é igual a 1.000 e outro onde a média é 10. Um desvio padrão de 5 será pequeno no primeiro grupo e grande para o segundo grupo. Nestes casos podemos utilizar também o Coeficiente de Variação (CV). Ele é a divisão do desvio padrão pela média, e o grupo que tiver menor CV tem menor dispersão ou variabilidade. Em alguns casos ele é apresentado em porcentagem, multiplicando assim seu resultado por 100. Fala-se em até 20%, o grupo possui uma dispersão pequena.

A variância, assim como o desvio padrão, avalia a dispersão do conjunto de dados, sendo a variância o desvio padrão ao quadrado. Ela é muito utilizada na estatística inferencial que veremos em breve.

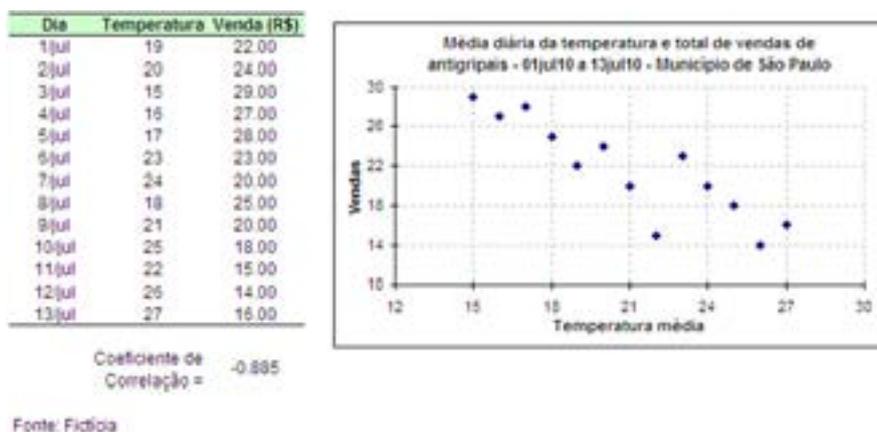
Dados Discrepantes

Por meio dos quartis podemos chegar a dados discrepantes, mais conhecidos como outlier. A identificação destes dados pode nos mostrar que houve um erro na digitação, o que pode ser arrumado de forma fácil, ou demonstra que realmente aquele elemento é discrepante. Além dos 'Outliers' existe também os 'Outliers Extremos', que são valores extremamente discrepantes.

4.6 Correlação

A correlação é uma análise que nos auxilia a verificar se uma variável se relaciona com outra variável, como por exemplo, a queda da temperatura e o aumento da venda de antigripais. Este caso é denominado de correlação indireta ou negativa, pois enquanto o valor de uma variável sobe, o da outra cai. Quando o valor das duas variáveis aumenta juntamente chamamos de correlação direta ou positiva.

O gráfico que demonstra esta relação entre variáveis é o Gráfico de Dispersão.



Quanto mais próximo de +1 ou -1 o valor do coeficiente de correlação, maior a dependência das variáveis. Para avaliarmos se seu valor é alto ou baixo, podemos recorrer ao Teste de Rugg, considerando a seguinte tabela:

Tabela de Rugg	
$0 < r < 0,14$	desprezível
$0,15 < r < 0,29$	baixo
$0,30 < r < 0,49$	apreciável
$r > 0,50$	acentuado

Isso também vale para os valores negativos.

Para provar a significância do Coeficiente de Correlação podemos utilizar o Erro Provável. Ele é um valor gerado por meio do coeficiente de correlação, do número de elementos (pares) e da constante 0,6745 conforme a seguinte fórmula:

$$E.P. = 0,6745 \times \frac{1 - r^2}{\sqrt{n}}$$

Onde:

n = número de elementos

r = coeficiente de correlação

0,6745 = constante

Para considerarmos o coeficiente de correlação estatisticamente significativo, seu valor deve ser, em módulo, pelo menos três vezes maior que o valor do E.P. (erro provável), ou seja, no nosso exemplo acima sobre temperatura temos $r = 0,88$ e $E.P. = 0,146$.

$n = 13$ $r = -0.885$ $\text{constante} = 0.6745$ $E.P. = \sqrt{F_{44} \cdot (F_{43}^2) / RAIZ(F_{42})}$	$n = 13$ $r = -0.885$ $\text{constante} = 0.6745$ $E.P. = 0.14652$ $0.885 > 3 \cdot 0.146$ $0.885 > 0.440$
--	--

Três vezes 0,146 é igual 0,44 que é menor que nosso 'r', ou seja, este coeficiente de correlação é significativo.

Importante ressaltar que apesar de haver correlação entre duas variáveis, isto não comprova que exista relação de causa-efeito, ou seja, não podemos afirmar que variações em uma delas cause variação na outra. Isto somente poderá ser afirmado com outras análises estatísticas.

4.7 Amostragem

Um levantamento por amostragem é utilizado seja por economia de tempo, ou de recursos, ou por inviabilidade de se coletar o dado de toda a população analisada.

Em alguns casos não é recomendado a utilização de uma amostra, mas sim a extração de dados de todos os elementos da população, ou seja, é necessário um censo. São eles:

- Quando a população for muito pequena (menor que 30 elementos);
- Quando o dado a ser coletado está fácil de se obter, como por exemplo, uma população de funcionários que trabalham em um mesmo local, sala etc.
- Quando necessitamos de uma alta precisão para o dado coletado.

Chamamos de população o grupo de interesse de coleta de dados para o pesquisador. Por exemplo, digamos que queremos analisar a quantidade de maconha apreendida com réus condenados por Tráfico de Drogas no Brasil em um determinado período. Para isso não necessitamos ler todas as sentenças, mas apenas uma amostra delas, podendo extrair mais dados em menos tempo do que se fossemos analisar todos os casos, porém para que não incorramos em erros, estes processos devem ser escolhidos de maneira adequada, o que veremos adiante.

As amostras podem ser probabilísticas ou não probabilísticas. Os casos onde cada elemento da população possui a mesma probabilidade de ser selecionado para a amostra, são 'probabilísticas'. Já os casos em que, em um conjunto da população, indagamos sobre voluntários para responderem um questionário sobre um deter-

minado assunto, ou enviamos por e-mail um questionário e temos como amostra os respondentes, estamos selecionando uma amostra não probabilística. Neste último exemplo a coleta acaba sendo viesada, pois os respondentes provavelmente se interessam pelo assunto, por isso responderam, não obtendo assim uma representação correta da população. A amostragem não probabilística deve ser utilizada somente para casos em que uma amostra aleatória é muito difícil ou impossível de ser coletada (geralmente por não se saber o total populacional), sendo assim coletados apenas os elementos que estão à disposição.

Tamanho da amostra

Existem várias formas de se chegar ao tamanho da amostra, aqui vamos tratar apenas da amostragem aleatória simples.

Antes de determinar o tamanho da amostra, o pesquisador deve traçar o erro que tolera se submeter. O erro amostral mostra a diferença entre o valor apresentado pela análise da amostra e o real valor da população, por exemplo, quando uma pesquisa eleitoral afirma que o candidato 'A' tem 60% dos votos com uma margem de erro de dois pontos percentuais para mais ou para menos, quer dizer que o erro é de 2%, podendo a preferência da população para o candidato 'A' estar entre 58% e 62%.

Na área de Ciências Biológicas as análises são feitas sempre com um erro bem baixo, geralmente em torno de 1%. Já na área de Ciências Humanas esta margem geralmente gira em torno dos 5%.

O cálculo da amostra simples leva em consideração as seguintes variáveis:

- N = quantidade de elementos da população (tamanho da população);
- n_0 = primeira aproximação do tamanho da amostra;
- n = quantidade de elementos da amostra (tamanho da amostra);
- E_0 = erro amostral tolerado pelo pesquisador.

O cálculo de n_0 é feito mesmo sem o conhecimento do tamanho da população, chegando-se a um primeiro tamanho da amostra:

$$n_0 = \frac{1}{E_0^2}$$

Tendo o valor de N, conseguimos corrigir o valor identificado acima, por meio da fórmula:

$$n = \frac{N \times n_0}{N + n_0}$$

No nosso exemplo acima, da quantidade de maconha em condenações por tráfico de drogas, imaginando que no período escolhido temos 4320 sentenças, tolerando um erro de 5% e aplicando a fórmula da amostra simples temos:

Com erro de 5%

População	Amostra
4320	=B4*\$C\$7/(B4+\$C\$7)

erro = 0.05
=1/(C6^2)

Com erro de 5%

População	Amostra
4320	366

erro = 0.05
400

Com um erro de 5%, chegamos ao total de 366 processos a serem analisados para tirarmos conclusões. Digamos que se chegarmos com a leitura desses processos que a média em kg de maconha apreendida é dois, poderemos afirmar que do total de condenações no período escolhido, no Brasil, a média de maconha apreendida está entre 1,9 kg e 2,1 kg.

Agora vamos ver as formas possíveis e corretas de escolha destes processos da amostra.

Tipos de amostragem

O processo de amostragem pode ser feito de várias formas, veremos aqui quatro tipos, são elas:

- amostragem estratificada: quando nossa população possui características diferentes, ou seja, ela é heterogênea, temos que selecionar a amostra de acordo com estes estratos, por exemplo, uma pesquisa aplicada aos funcionários do CNJ, onde o tipo de trabalho exercido pelo funcionário influencia na forma de se comportar, sendo assim teríamos dois estratos, magistrados e servidores administrativos;
- amostragem por conglomerados: quando temos vários grupos homogêneos, ao invés de selecionarmos alguns elementos de cada grupo, podemos escolher alguns grupos e realizar a pesquisa em cima de todos os elementos destes grupos selecionados;
- amostragem sistemática: neste tipo de seleção dos elementos para a amostra, devemos ter os dados organizados de forma aleatória, tendo o valor total da amostra, selecionamos os elementos de acordo com a razão entre a população e a amostra, ou seja, N/n . No exemplo das sentenças de processos de tráfico de drogas, digamos que temos eles organizados em uma planilha de Excel em ordem crescente por data, são 4.320 casos para extração de uma amostra de 366, ou seja, $4.320/366 = 12$, ou seja, iremos selecionar 366 sentenças de 12 em 12, a primeira delas deve ser um valor sorteado entre 1 e 12, digamos que o valor 10 seja o sorteado, sendo lido a sentença que se encontra na décima linha e em seguida o da vigésima segunda, posterior o da trigésima quarta etc.
- amostragem simples: quando temos uma população homogênea, podemos fazer a escolha dos elementos da amostra de forma aleatória, utilizando uma tabela de números aleatórios ou a função do Excel “=ALEATÓRIOE-NTRE(valor inferior;valor superior)”.

5 Bibliografia

BARBETTA, Pedro Alberto. Estatística aplicada às Ciências Sociais. 5 ed. rev. Florianópolis: Editora da UFSC, 2005.

COSTA, Sérgio Francisco. Introdução Ilustrada à Estatística. São Paulo: Harbra, 2005.

CNJ. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/167>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/159>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/119>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/110>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/2236>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/2525>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/3428>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/3453>>. Acesso em: 23 de fev. de 2021.

_____. Disponível em <<https://atos.cnj.jus.br/atos/detalhar/3488>>. Acesso em: 23 de fev. de 2021.

FAWCETT, T., & PROVOST, F. Data Science para Negócios. Rio de Janeiro: Alta Books, 2016.

LAPPONI, Juan Carlos. Estatística usando Excel. 4 ed. Rio de Janeiro: Elsevier, 2005.

RODRIGUES, Pedro Carvalho. Bioestatística. 3 ed. Niterói: EdUFF, 2002.

VIRGILLITO, Salvatore Benito. Estatística Aplicada. São Paulo: Ed. Alfa Ômega, 2004.

ERRADICAÇÃO
DA POBREZA



FOME ZERO E
AGRICULTURA
SUSTENTÁVEL



SAÚDE E
BEM-ESTAR



EDUCAÇÃO DE
QUALIDADE



IGUALDADE
DE GÊNERO



ÁGUA POTÁVEL
E SANEAMENTO



ENERGIA LIMPA
E ACESSÍVEL



TRABALHO DECENTE
E CRESCIMENTO
ECONÔMICO



INDÚSTRIA,
INOVAÇÃO E
INFRAESTRUTURA



REDUÇÃO DAS
DESIGUALDADES



CIDADES E
COMUNIDADES
SUSTENTÁVEIS



CONSUMO E
PRODUÇÃO
RESPONSÁVEIS



AÇÃO CONTRA A
MUDANÇA GLOBAL
DO CLIMA



VIDA NA
ÁGUA



VIDA
TERRESTRE



PAZ, JUSTIÇA E
INSTITUIÇÕES
EFICAZES



PARCERIAS
E MEIOS DE
IMPLEMENTAÇÃO



CNU