

Conferência Nacional dos Agentes Produtores e Usuários de Dados

Governança do Sistema

GT5: Qualidade institucional dos processos e produtos

30 de Julho de 2024

Carlos Torres Freire

Diretor do Centro de Ciência de Dados para Estatísticas Públicas

Sumário

- 1) Código de Boas Práticas das Estatísticas: Princípios de qualidade de processos e produtos estatísticos**
- 2) Era digital e desafios na produção estatística**
- 3) Casos da produção recente do SEADE:**
 - Painel COVID-19: novos sistemas de ingestão e tratamento de dados
 - Pesquisa com URA (“robô”): novo modo de coleta
 - PIB+30: uso de registros fiscais digitais
 - Estatísticas Vitais: mudança de processos e ferramentas computacionais para imagem
 - Investimentos anunciados: modelos de classificação de texto
- 4) Centro de Ciência de Dados para Estatísticas Públicas (CCCDEP)**
- 5) Autoridade Estatística: Sistema, Articulação, Coordenação e Participação**

Princípios de qualidade de processos e produtos estatísticos

PRODUTOS ESTATÍSTICOS

- **Relevância:** importância para o usuário, demanda da sociedade
- **Precisão:** descrição correta do fenômeno, validade
- **Atualidade:** distância da data de referência
- **Coerência:** comparabilidade, convenções, padrão conceitual, universo
- **Acessibilidade:** disponibilidade e transparência
- **Interpretabilidade:** informações adicionais, como metadados, metodologias, classificações

PROCESSOS ESTATÍSTICOS

- **Metodologia sólida:** conceitos, classificações, convenções nacionais e internacionais
- **Adequação** de instrumentos, técnicas e sistemas tecnológicos
- **Solicitação não excessiva** em coletas e uso de registros administrativos
- **Relação entre custo e eficiência**

Era digital e desafios na produção estatística

Tradicionalmente

- Estatísticas com coleta primária de dados – censos e *surveys* e registros administrativos

Na era digital

- **dados em grande volume, velocidade e variedade**
- **novos métodos e tecnologias** para coletar, armazenar, tratar e analisar dados para produção de estatísticas públicas
- novos modos de regular essa produção

Datafication

Registros gerados digitalmente, passivamente ou ativamente, por ações humanas e interações sociais, capturadas automaticamente.

Registros administrativos digitais

Cadastros sociais: educação, saúde e assistência social

informações fiscais: pagamentos, notas fiscais de transações

Ocorrências de segurança pública

Declarações de nascimentos e óbitos.

Subprodutos de ações em equipamentos

Uso de celular

Transações financeiras

Cartões de transporte público

Visitas a websites e buscas

Publicação de notícias e posts em redes sociais

Sensores

Presença física: tráfego de veículos e iluminação;

Remotos: satélites e veículos aéreos

Painel Coronavírus

Produção de dados para o Governo de SP e informações abertas para a sociedade.

Ingestão dados Sivep-Gripe e Censo Covid (ESP).

Tratamento de dados, consistência, produção de indicadores, visualização e repositório de dados.

www.seade.gov.br/coronavirus

ATUALIDADE

RELEVÂNCIA

ADEQUAÇÃO TÉCNICA

PRECISÃO

Estado de São Paulo

6.862.684 Casos
183.602 Óbitos
2,7 Letalidade

Brasil

38.824.288 Casos
712.380 Óbitos
1,8 Letalidade

Mundo

775.615.736 Casos
7.051.323 Óbitos
0,9 Letalidade

Dep. Regional de Saúde

Todos

Município

Todos

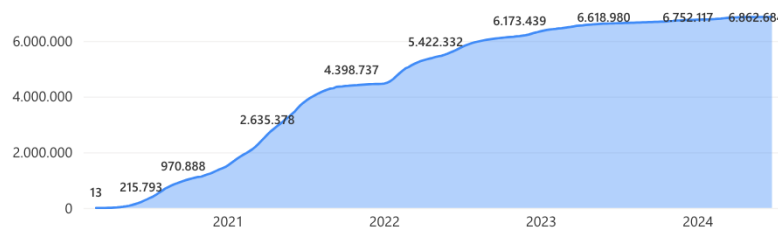
Casos 6.862.684 Total
21 Variação semanal
4,7 Últimos 14 dias/100 mil hab.

Date
01/03/2020 23/06/2024

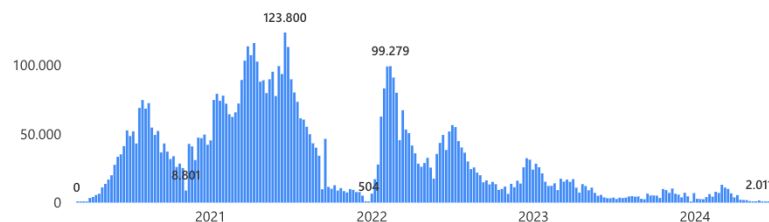
Casos

Óbitos

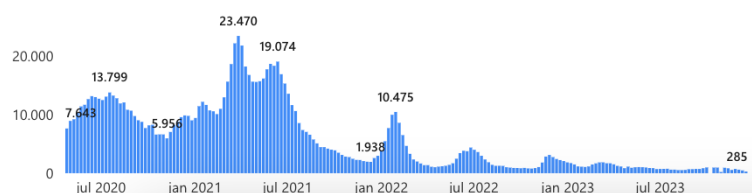
Total de casos



Casos por semana



Internações por semana, por Dep. Regional de Saúde

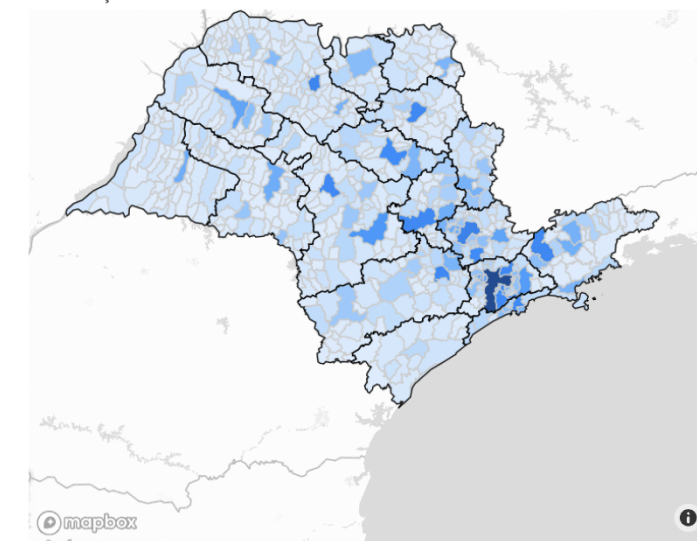


Óbitos 183.602 Total
Infinito Variação semanal
0,2 Últimos 14 dias/100 mil hab.

Mapa

Tabela

Distribuição dos casos



Internações por Dep. Regional de Saúde

(casos suspeitos e confirmados)

Novas na semana 285 Variação semanal (%) -0,74

Ocupação Leitos Covid na semana, por Dep. Regional de Saúde

(casos suspeitos e confirmados)

Enfermaria (%) 15,4 UTI (%) 21,8

Monitor Covid-19

ATUALIDADE

RELEVÂNCIA

CUSTO / BENEFÍCIO

COERÊNCIA e COMPARABILIDADE

Pesquisa diária sobre temas relacionados à pandemia

- Acompanhamento de sintomas; distanciamento social; atividade econômica e consumo; trabalho e renda; educação; vacina; expectativas da população.
- De 23/Março/2020 a Setembro/2021.
- **Coleta de dados por telefone**, utilizando Unidade de Resposta Audível (URA).
- **800 entrevistas diárias**, para constituir um painel móvel de 4.800 entrevistas
- **População do Estado de São Paulo**, com amostra representativa por local de residência, **sexo e idade**
- Desagregação por Estado, RMSP, município SP, demais municípios da RMSP e Interior; e **17 Departamentos Regionais de Saúde**.

Experiência levou a pesquisas de percepção:

- **Cultura**
- **Atividade física**
- **Turismo**
- **Endividamento**
- **Segurança Alimentar**
- **Cuidados,**
- **Hábitos Consumo App e Internet**

Modelos de **aprendizado de máquina** para **classificação automática de textos** a partir de **áudios captados** nas pesquisas de percepção com a população

PIB+30

RELEVÂNCIA

ATUALIDADE

ADEQUAÇÃO TÉCNICA

COERÊNCIA e COMPARABILIDADE

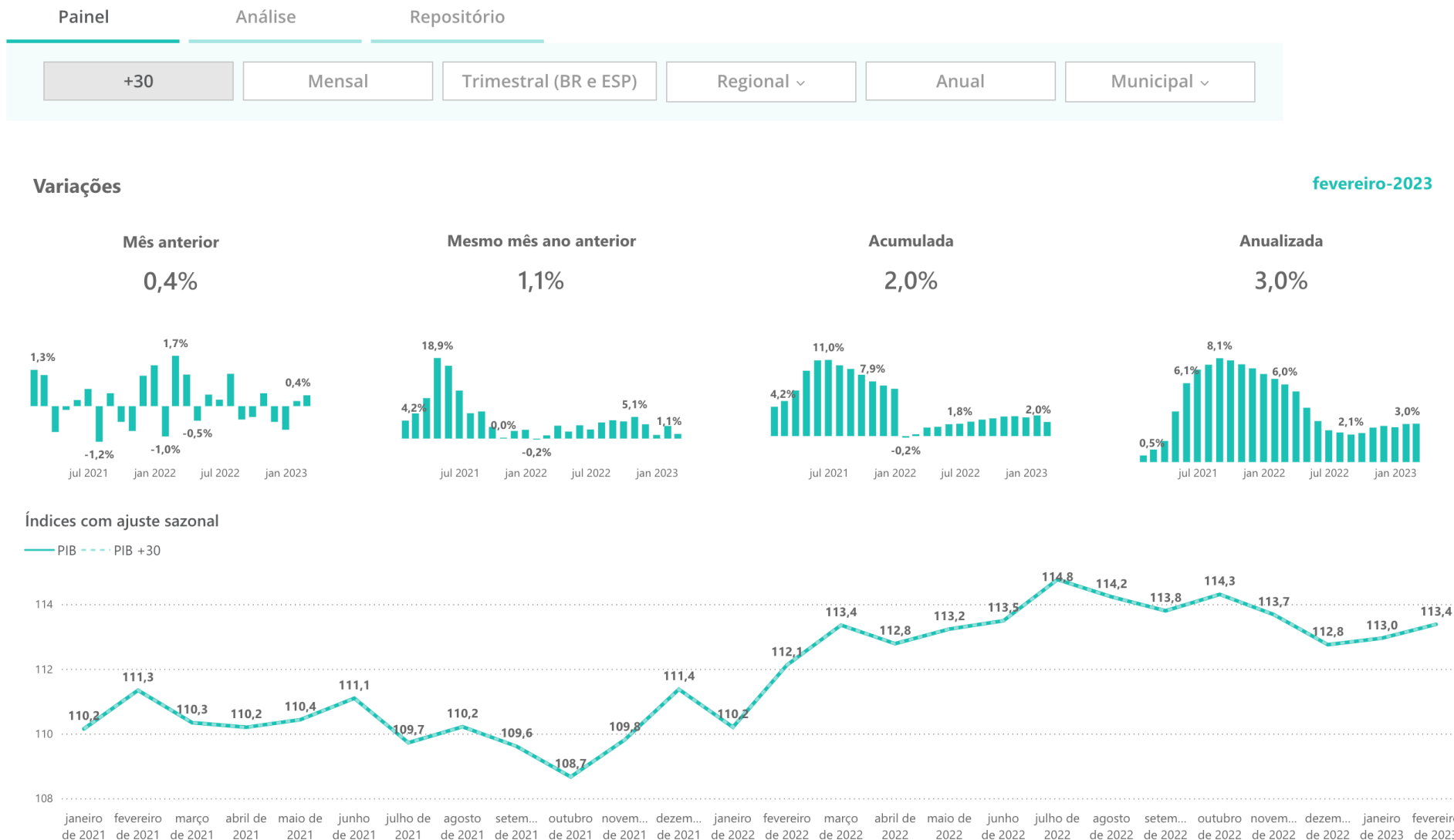
Indicador mensal sobre atividade econômica do Estado de SP

Diminui defasagem entre mês de publicação e data de referência dos dados de 60 dias para 30

Representa 97% do PIB estadual

Usa estatísticas de alta frequência e 200 variáveis (NFe, documentos fiscais)

Complementa produção: PIB Mensal, PIB Trimestral, PIB Anual, Projeções 2021 e 2022



Fonte: Fundação Seade.

Nota: Devido às mudanças implementadas pelo IBGE em abril de 2023 em suas pesquisas conjunturais (PMS, PMC e PIM_PF), a Fundação Seade suspendeu temporariamente a publicação do PIB+30 para realizar a revisão metodológica do indicador.

SEADE ESTATÍSTICAS VITAIS

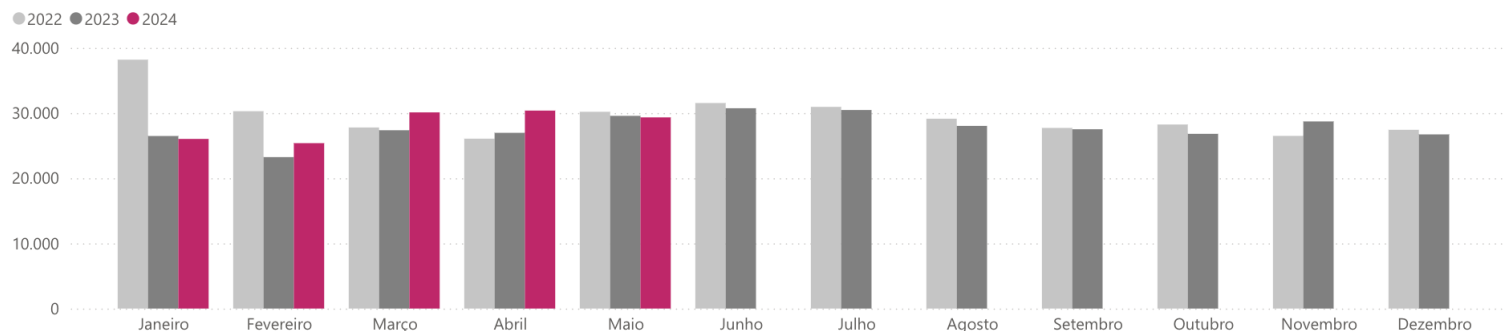
- Indicadores mensais de nascimentos, óbitos e casamentos.
- Mudanças nos processos de gestão da coleta e de produção para diminuir tempo de defasagem para divulgação.
- Modelos de aprendizado de máquina para leitura automatizada de declarações em imagem e arquivos de dados dos cartórios de Registro Civil.

141.211
Óbitos em 2024

ATUALIDADE

PRECISÃO

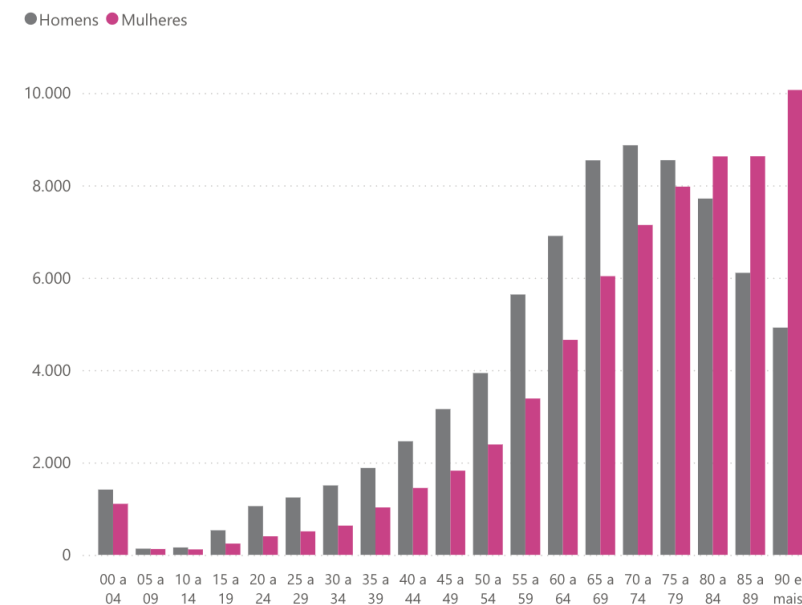
Óbitos por mês de ocorrência



Óbitos por sexo - 2024

Município	Homens	Mulheres	Total (1)
São Paulo	18.296	17.618	35.914
Guarulhos	2.261	1.973	4.234
Campinas	1.877	1.790	3.668
Santo André	1.246	1.221	2.467
Sorocaba	1.199	1.058	2.257
Osasco	1.188	1.045	2.233
São Bernardo do Campo	1.120	1.063	2.184
São José dos Campos	1.117	964	2.081
Ribeirão Preto	1.030	966	1.996
Santos	908	987	1.895
Mogi das Cruzes	787	731	1.518
São José do Rio Preto	752	696	1.449
Jundiaí	744	683	1.427
Piracicaba	747	647	1.394
Praia Grande	658	595	1.253
Bauru	602	609	1.211
São Vicente	606	549	1.155
Mauá	638	511	1.149
Franca	602	513	1.115
Barueri	566	457	1.023
Guarujá	570	448	1.018
Taubaté	551	459	1.010
Diadema	521	466	987
Limeira	520	458	978
Itaquaquecetuba	483	419	902
Total	74.829	66.372	141.211

Óbitos por sexo e idade - 2024



Fonte: Fundação Seade. Sistema de Estatísticas do Registro Civil.

Nota: O Seade produz estatísticas de mortalidade a partir dos registros de óbitos enviados mensalmente pelos Cartórios de Registro Civil de todos os municípios paulistas. As informações contemplam as mortes de residentes no Estado de São Paulo ocorridas no ano/mês de referência. O menor número de eventos em certos meses da série pode ser devido à incompletude no envio dos registros pelos Cartórios, principalmente no último mês de divulgação. Dados de municípios inferiores a 50 mil habitantes estão mais sujeitos a oscilações. Os dados de óbitos de 2023 foram atualizados em 14/04/2024 e os de 2024 em 15/07/2024.

SEADE INVESTIMENTOS

- Monitoramento sistemático e automatizado dos investimentos de empresas privadas e públicas no Estado
- Captação diária de anúncios na imprensa
- Modelos de aprendizado de máquina faz triagem de textos sobre investimento
- Equipe de coleta confirma com empresas e completa informações como tipo, período, valor e local do investimento.



Centro de Ciência de Dados para Estatísticas Públicas

Chamada FAPESP: Estimular pesquisas em parceria com governos para aperfeiçoar políticas públicas.

Contexto: IA, novas fontes de dados digitais e ferramentas computacionais e desafios na produção de informação.

Objetivos:

- 1 Desenvolver novos métodos e ferramentas para produção de estatísticas para políticas públicas.
- 2 Introduzir capacidades técnicas para produção e análise de dados no Seade e na gestão pública.
- 3 Disseminar conhecimento e indicadores no Gov. ESP.

Problema de pesquisa: Como usar dados de alta frequência, em grande volume e não estruturados para indicadores de políticas públicas?

Parcerias: órgãos públicos (Sec. Fazenda, Metro, CPTM), instituições de pesquisa (USP, FGV, Unesp).

Linhas de pesquisa:

- 1 **Atividade econômica:** Análises setoriais e regionais sobre economia de SP com indicadores baseados em registros fiscais digitais e dados de serviços públicos.
- 2 **Mercado de Trabalho:** Novos modos de coleta para indicadores sobre condição de atividade e formação e inserção no mercado de trabalho.
- 3 **Mobilidade Urbana:** Análises sobre dinâmica urbana e uso de transporte público baseadas em dados de bilhetagem eletrônica, localização e infraestrutura.
- 4 **Segurança pública:** Indicadores de eventos criminais com base em textos de registros de ocorrência e cruzamento com sistemas da Segurança e da Saúde.
- 5 **Tecnologia e engenharia de software:** Desenvolver aplicações para coleta, tratamento e processamento de dados para apoiar as demais linhas de pesquisa.

Tópicos para discussão sobre o sistema nacional de estatísticas

- Autoridade estatística, sistema, articulação, coordenação e participação
- Regulação, normatização e curadoria
- Composição da autoridade, quem regula e quem produz
- Princípios de qualidade na era digital
- Restrições materiais de órgãos produtores e estímulo a fortalecimento institucional

Obrigado!

Carlos Torres Freire | carlosfreire@seade.gov.br

SEADE
Fundação Sistema Estadual
de Análise de Dados

