

Análise de dados: uma leitura crítica das informações

Visão Sistêmica Sobre a Análise de Dados

3
Módulo

Fundação Escola Nacional de Administração Pública

Diretoria de Desenvolvimento Profissional

Conteudista/s

Ricardo Alexandre Amaral (conteudista, 2022);
Diretoria de Desenvolvimento Profissional.



Enap, 2022

Fundação Escola Nacional de Administração Pública

Diretoria de Desenvolvimento Profissional

SAIS - Área 2-A - 70610-900 — Brasília, DF

Sumário

Unidade 1: Elementos de Tratamento Numérico e Computacional de Dados5

1.1 Tratamento computacional de dados..... 6

1.2 Aplicação de métodos numéricos de ajuste de dados..... 7

Referências 10

Unidade 2: Fundamentar Opinião com Base em Resultados12

2.1 Sumário da análise de dados 13

2.2 Tratamento da informação..... 15

Referências 17

Unidade 3: O Campo de Dados e o Cenário Atual 18

3.1 Discussão e fechamento 18

3.2 Referências de estudo 20

Referências 22

Apresentação e Boas-vindas

Seja bem-vindo e bem-vinda ao curso Análise de Dados: uma Leitura Crítica das Informações.

Neste curso as noções de estatística serão as lentes fundamentais que nortearão o seu estudo. Mas não se preocupe, as ferramentas estatísticas necessárias para o entendimento do conteúdo serão definidas e retomadas ao longo de todo o curso para ajudar você a alcançar o objetivo geral desta capacitação. Ao final de seus estudos é esperado que você saiba utilizar os fundamentos estatísticos para ler os dados de maneira objetiva e crítica. Dessa forma, o objetivo central do curso se volta à necessidade de entender, na atual era da informação, o conceito de dados e de como tecer uma leitura crítica e objetiva destes.

A estatística é entendida como a ciência dos dados, visto que tem ligação direta com o tratamento desses dados em vários campos do saber. Especialmente neste curso, você poderá analisar o alinhamento da estatística com o desenvolvimento da sociedade, via produção de pesquisa social e por meio de indicadores que permitem mensurar desde características regionais de um país, até suas demandas de desenvolvimento socioeconômico.

O propósito é que este curso sirva para aprimorar um olhar crítico sobre o tema análise de dados e sobre outras áreas de atuação adjacentes ao tema como, por exemplo: gestão, mercado, políticas públicas e a cultura do cidadão pleno. Dessa forma, uma vez que o indivíduo estabeleça vínculo com as mídias informativas espera-se que ele esteja apto à leitura e interpretação de dados.

O conteúdo ao longo do curso está distribuído em três módulos. No primeiro módulo você irá reconhecer qual a importância da análise de dados, irá compreender a conexão entre estatística e tratamento de dados, além de examinar sistemas sob o prisma estatístico.

No segundo módulo você irá reconhecer os principais formatos de representação de dados, analisar correlações não-lineares, além de examinar os ajustes linear e quadrático.

Por fim, no terceiro módulo, irá identificar a contribuição computacional associada ao tratamento de dados, reconhecerá as conexões entre análise de dados e pensamento estratégico, além de reconhecer a importância da análise de dados para a leitura crítica das informações.

Então é hora de começar!

2 Visão Sistêmica Sobre a Análise de Dados

O presente módulo do curso de análise de dados traz uma visão sistêmica sobre este tema e que tem por objetivos identificar a contribuição computacional associada ao tratamento de dados; evidenciar conceitos fundamentais sobre pensamento estratégico e, finalmente, reiterar a emergência da análise de dados para uma leitura crítica das informações.

É sabido que programas de computador de análise de dados como SciDAVis facilitam em muito a determinação de fits, que modelam problemas de vários campos, e são massivamente utilizados no exame de curvas de dados não-lineares.

A ideia aqui é mostrar como obter a análise de dados e como fazer esses fits, que vão sendo internalizados e contribuem para o desenvolvimento da sua percepção e criticidade. Logo, o olhar clínico do examinador é imprescindível, uma vez que é ele que faz as escolhas quanto aos ajustes, do mais simples para o mais complexo: reta, quadrático etc. Sites como o Planet Calc fazem regressões lineares, quadráticas e cúbicas de modo simples, direto e sem a necessidade de instalação de nenhum programa.



SAIBA MAIS

■ Para saber mais sobre o Planet Calc clique [aqui](#) e acesse o site

Unidade 1: Elementos de Tratamento Numérico e Computacional de Dados

Objetivo de aprendizagem

Ao final desta unidade você será capaz de identificar a contribuição computacional associada ao tratamento de dados, ou seja, a prática de ajuste, seu significado e desdobramentos, via utilização de métodos computacionais simples, diretos e didáticos para tratamento de dados.

1.1 Tratamento computacional de dados

O desenvolvimento dos computadores e outros dispositivos eletrônicos equivalentes, alavancou a utilização de métodos numéricos em basicamente todas as áreas do conhecimento via uso e popularização de algoritmos de aproximação e simulação de problemas. Aqui serão designados conjuntos de técnicas estatísticas desenvolvidas com uso desse tratamento computacional de dados.



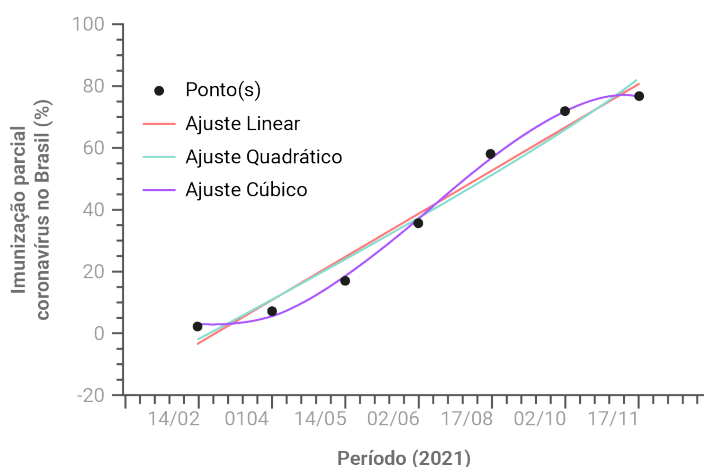
Tratamento computacional de dados.

Fonte: Freepik. Elaboração: CEPED/UFSC (2022).

É importante entender que tais dispositivos são essencialmente máquinas de calcular, e que resolvem operações de modo automático e muito rápido, logo é excelente para desempenharem trabalhos com o perfil numérico, também denominados não-analíticos. Contudo o foco aqui é analisar a representação gráfica das informações à luz dos conceitos estatísticos abordados.

1.2 Aplicação de métodos numéricos de ajuste de dados

Para seguir com o conteúdo será usada novamente a análise da curva de imunização parcial contra coronavírus no Brasil e seus ajustes. É importante lembrar que o *fit* em vermelho é linear e não passa por todos os pontos, assim tendo uma correlação menor que a dos demais ajustes, representados em verde claro e roxo, que são ajustes de graus 2 e 3 respectivamente.



Dados e esboço da curva do percentual de imunização parcial com a primeira dose da vacina contra o coronavírus no Brasil utilizando o SciDAVis.

Fonte: Our World in Data (2021). Elaboração: CEPED/UFSC (2022).

O resultado computacional gerado possibilita ajustar a curva por diferentes funções escritas, veja:

- $y = A \cdot x + B$ (grau-1, reta, linear);
- $y = a_0 + a_1 \cdot x + a_2 \cdot x^2$ (grau-2);
- $y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3$ (grau-3).

A figura a seguir demonstra os parâmetros de ajuste e correlação dos pontos obtidos, isto é, os termos a_0 , a_1 e a_2 .

O **coeficiente de correlação** é denominado " R^2 ", o qual estima o grau de correlação de cada ajuste, portanto é a "confiabilidade do ajuste".

Idealmente $R^2 = 1 = 100\%$, seria uma reta perfeita, porém não existe para dados reais, uma vez que dados tem uma incerteza intrínseca, portanto quanto mais baixo

o R^2 , entendido como R ao quadrado, pior é o ajuste executado. Em especial, os coeficientes de correlação associados aos ajustes dos dados sobre a imunização parcial contra covid no Brasil são:

- $R^2 = 0,968854555231625$ para o *fit* linear;
- $R^2 = 0,970304504278698$ para o *fit* quadrático;
- $R^2 = 0,998294079690944$ para o *fit* cúbico.

Consequentemente os coeficientes de correlação são muito próximos, 96,8%; 97%; e 99,8% para os ajustes linear, quadrático e cúbico.

Ajuste Linear de Dados
usando uma função $y=A*x+B$;

B = -18,07

A = 14,07

 $R^2 = 0,96$

Ajuste Polinomial de Dados
usando uma função $y=a_0+a_1*x+a_2*x^2$ (grau-2);

$a_0 = -14,3$

$a_1 = 11,55$

$a_2 = 0,31$

 $R^2 = 0,97$

Ajuste Polinomial de Dados
usando uma função $y=a_0+a_1*x+a_2*x^2+a_3*x^3$ (grau-3);

$a_0 = 16,70$

$a_1 = -23,74$

$a_2 = 10,64$

$a_3 = -0,86$

 $R^2 = 0,99$

Representação dos parâmetros de ajuste computacional de curvas usando o SciDAVis dos dados acerca da imunização com a primeira dose da vacina contra o coronavírus no Brasil.

Fonte: Our World in Data (2021). Elaboração: CEPED/UFSC (2022).

Neste caso, pela análise dos coeficientes de correlação dos ajustes temos uma boa aproximação que indica que todos os ajustes para quadro sobre a imunização parcial no Brasil são aceitáveis. E deste sucesso da parametrização computacional é possível ter um bom modelo aproximativo com qualquer um dos três ajustes, inclusive o mais simples.

Você chegou ao fim desta unidade. Caso surjam dúvidas, faça uma releitura dos tópicos de interesse. Bons estudos!

Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

CARVALHO, Marília Sá; SOUZA-SANTOS, Reinaldo. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 21, p. 361-378, 2005.

COMARELA, Giovanni et al. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. SCHAEFFER FILHO, Alberto Egon; CORDEIRO, Weverton Luis da Costa; CAMPISTA, Miguel Elias Mitre (ed.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, Porto Alegre, Sociedade Brasileira de Computação, p. 246-295, 2019.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. **Big Data para leigos**. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

LEITE, Angela; FUITA, Hiroco. **Aplicações da matemática: administração, economia e ciências contábeis**. Boston, Cengage Learning, 2008.

OUR WORLD IN DATA. **Brazil: coronavirus pandemic country profile**. Coronavirus Pandemic Country Profile. 2021. Disponível em: <https://ourworldindata.org/coronavirus/country/brazil>. Acesso em: 09 mar. 2022.

PINTO, José Carlos; SCHWAAB, Marcio. **Análise de Dados Experimentais: I. Fundamentos de Estatística e Estimação de Parâmetros**. Rio de Janeiro, Editora E-papers, 2007.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

SOUZA, Emanuel Fernando Maia de; PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. **Software Livre R: aplicação estatística**. 2014. Universidade Federal da Paraíba.

VITALI, Marieli Mezari. Estatística sem matemática para psicologia. **Revista Brasileira de Psicodrama**, São Paulo, v. 27, n. 1, p. 139-144, 2019.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.

Unidade 2: Fundamentar Opinião com Base em Resultados

Objetivo de aprendizagem

Ao final desta unidade você será capaz de reconhecer a análise de dados e o pensamento estratégico, assim atrelando as fits e seus modelos ao olhar crítico, o qual deve fundamentar o discurso e exame dos dados.

A sociedade está imersa na cultura da informação. Jornais, revistas e a internet promovem intenso contato com dados, que caracterizam o tempo presente e que também permitem ingênuos prognósticos do porvir.

Ocorre que a ideia por trás de um ajuste é obter uma função que modela os dados, que os recria, e deste modo que permite fazer inferências, mesmo que tímidas, sobre a evolução de um sistema, ou seja, um futuro caracterizado com base na tendência determinada por um rol de dados.

Como por exemplo: modelos com relação à população brasileira, expectativa de vida, percentual de trabalhadores ativos, são cruciais para o futuro da previdência.



DESTAQUE

Modelos com base em dados (medidas) são sempre aproximados e imprecisos, visto que todo dado tem um erro intrínseco, porque são fruto de medições e não há instrumentos de medida e experimentadores perfeitos. Todavia, é interessante perceber que os modelos fenomenológicos são entendidos como nossa melhor resposta para diversos problemas e respondem bem a vários casos onde não há grande volatilidade ou onde não ocorreram crises.

Deste modo os modelos fenomenológicos são amplamente utilizados na descrição de tendências, isto é, demonstram como será a evolução dos dados sem o acontecimento de crises. Como exemplos de resultados experimentais, tecidos via ajuste de dados, é possível extrair diagnósticos em várias áreas de conhecimento. Conheça algumas áreas em que se pode utilizar a análise de dados para a definição de tendências:

- **Evolução populacional de pessoas de um país ou região:** estes comumente apresentam um padrão de crescimento ou decréscimo, sem a consideração de eventos como: guerras, recessão, pandemias etc;
- **Taxa de crescimento de cultura de bactérias:** as bactérias apresentam uma alta velocidade de crescimento, comumente se duplicam a cada intervalo de tempo, assim a descrição fenomenológica do ajuste do crescimento livre da população de bactérias é possível, considerando a não intervenção de agentes externos, como temperatura ou medicamentos;
- **Comportamento de índices econômicos não muito voláteis:** neste caso poderiam ser considerados índices não tão sensíveis ao mercado, como a inflação de países desenvolvidos, ou de modo mais simples, índices que exibam alguma regularidade.

2.1 Sumário da análise de dados

Motivado pelo ajuste linear de boa qualidade (ótima correlação), é possível avançar na busca de extrair resultados sobre os dados da curva de imunização parcial contra coronavírus no Brasil. Logo a ideia é listar os passos para a construção de informações a partir dos dados e a interpretação do contexto.

É interessante tratar um pouco mais tais dados sobre a imunização, uma vez que é possível determinar o eixo temporal em dias, calculando a diferença entre datas. Um exemplo pode ser visto no período de tempo entre 14/02 e 01/04 de 2021, que pode ser reescrito, buscando uma padronização quanto ao tempo. A tabela a seguir mostra tal raciocínio e resultados.

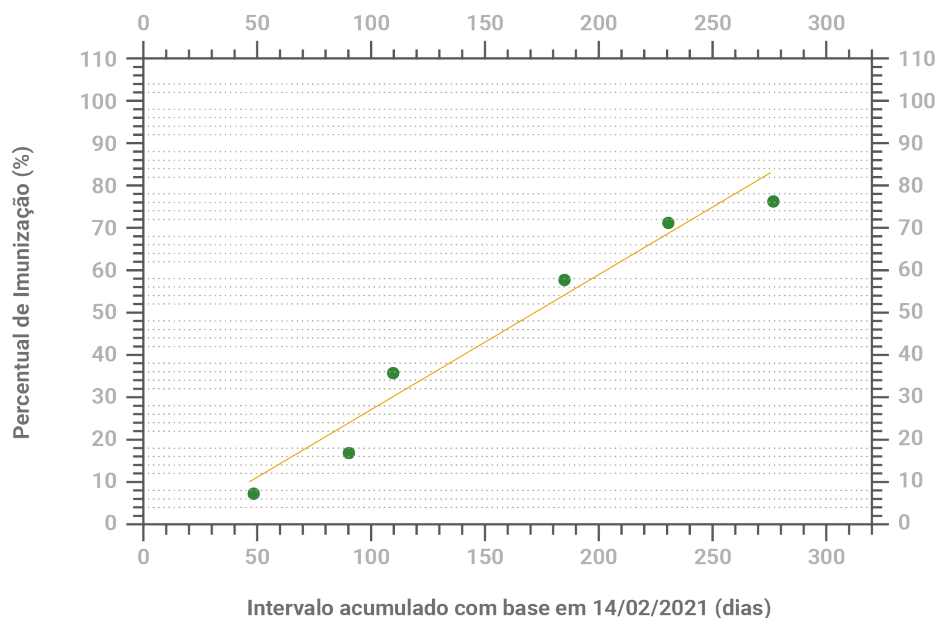
Data	Intervalo (dias)	Intervalo Acumulado (dias)	Percentual de Imunização
12/02/2021	-		2
01/04/2021	46	46	7
14/05/2021	43	89	16,8
02/06/2021	19	108	35,7
17/08/2021	76	184	57,8
02/10/2021	46	230	71,6
17/11/2021	46	276	76,6

Organização e tratamento dos dados do percentual de imunização contra o coronavírus no Brasil.

Fonte: Our World in Data (2021). Elaboração: CEPED/UFSC (2022).

A partir dos dados da tabela é possível construir um gráfico que exprime o comportamento da evolução da imunização parcial contra coronavírus no Brasil.

A figura a seguir, via análise do gráfico e da reta de ajuste, permite prever que se chegará a 100% da população por volta de 340 dias, ou seja, tratar dados e promover argumentação com base nos mesmos.



Percentual de Imunização contra o coronavírus no Brasil, utilizando tratamento simples dos dados temporais e executando um *fit* linear.

Fonte: Our World in Data (2021). Elaboração: CEPED/UFSC (2022).

É possível ainda, tendo em vista a evolução da imunização parcial contra coronavírus, examinar os parâmetros do *fit* mostrados na figura acima e que permitem uma ótima equação de reta para o problema descrito pela equação:

$$y = (0,319389909358684)x + (-5,4151309052754)$$

De modo que, substituindo $y = 100$, ou seja, 100% da imunização, infere quanto a prever o tempo (x) necessário para tal tarefa:

$$100 = (0,319389909358684)x + (-5,4151309052754)$$
$$x = 340 \text{ dias como esperado!}$$

**Ajuste Linear de Dados
usando uma função $y=A*x+B$;**

$$B = -5,415$$

$$A = 0,319$$

 $R^2 = 0,96$

**Parâmetros de Ajuste do Percentual de Imunização contra o coronavírus no Brasil,
utilizando tratamento simples dos dados temporais e executando um *fit* linear.**

Fonte: Our World in Data (2021). Elaboração: CEPED/UFSC (2022).

2.2 Tratamento da informação

Como descrito no tópico anterior, é possível modelar o comportamento da evolução da imunização de modo simples e, principalmente, tecer uma previsão plausível sobre o tema.

Note ainda que a função de ajuste, escrita de modo simplificado como: $y = (0,319)x + (-5,415)$ se comporta como uma regra, mas não é a regra, pois é possível, por exemplo, que se admita percentuais negativos.

Exemplo, para $x = 0$ resulta-se em $y = -5,414\%$, o que é matematicamente correto, mas absurdo pelo contexto do problema. O fato é que a expertise humana é sempre bem-vinda para análise de dados.

Deste modo a construção de ajustes nos leva a desenvolver modelos mentais com relação a dados e assim inferir resultados intuitivamente com relação à informação. É deste modo que a aplicação de conceitos associados a médias, representações, gráficos, ajustes, entre outros, molda diversas fenomenologias, seja em ciências naturais, mercados não demasiadamente voláteis, comportamento de características socioeconômicas etc.

Você chegou ao fim desta unidade. Parabéns! Caso exista dúvida sobre o tema, reveja o conteúdo para fixar seus estudos.

Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. **Big Data para leigos**. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

LEITE, Angela; FUITA, Hiroco. **Aplicações da matemática: administração, economia e ciências contábeis**. Boston, Cengage Learning, 2008.

OUR WORLD IN DATA. Brazil: coronavirus pandemic country profile. Coronavirus Pandemic Country Profile. 2021. Disponível em: <https://ourworldindata.org/coronavirus/country/brazil>. Acesso em: 09 mar. 2022.

PINTO, José Carlos; SCHWAAB, Marcio. **Análise de Dados Experimentais: I. Fundamentos de Estatística e Estimação de Parâmetros**. Rio de Janeiro, Editora E-papers, 2007.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

SOUZA, Emanuel Fernando Maia de; PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. Software Livre R: aplicação estatística. 2014. Universidade Federal da Paraíba.

VITALI, Marieli Mezari. Estatística sem matemática para psicologia. **Revista Brasileira de Psicodrama**, São Paulo, v. 27, n. 1, p. 139-144, 2019.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.

Unidade 3: O Campo de Dados e o Cenário Atual

Objetivo de aprendizagem

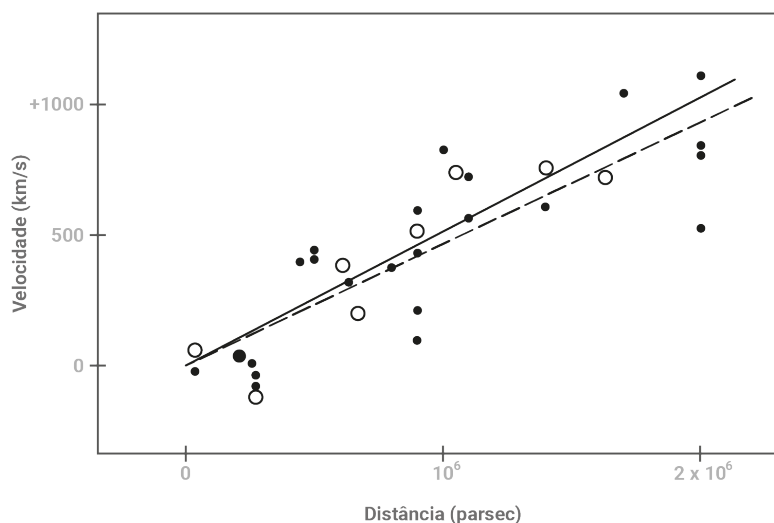
Ao final desta unidade você será capaz de reconhecer a emergência da análise de dados para uma leitura crítica de informações.

3.1 Discussão e fechamento

A seguir você verá a imagem de um gráfico de importância histórica no mundo da ciência, que toca na compreensão que a humanidade tem do Universo e ilustra as semelhanças operacionais com que os cientistas das ciências da natureza estabelecem sua relação ao tratamento de dados.

A referida imagem mostra a relação da velocidade em função da distância de estrelas muito distantes e cujo ajuste, inicialmente feito rapidamente à mão, usando somente a intuição, evidenciou a Lei de Hubble, ou seja, desvendou uma das linguagens da natureza.

O gráfico mostra o aumento da velocidade com a distância, em outras palavras, os objetos luminosos mais distantes do universo estão com maior velocidade em relação a nós, logo não estão fixos, mas sim se movimentando, de modo que se conclui que o universo está inflando.



A velocidade em função da distância de objetos astronômicos muito distantes.

Fonte: UFRGS. Elaboração: CEPED/UFSC (2020).

O gráfico acerca da Lei de Hubble desvenda uma das linguagens da natureza (lei natural), visto que o gráfico mostra o aumento da velocidade com a distância de nebulosas, isto é, nuvens de poeira cósmica sendo que:

- os pontos representados por círculos cheios são as nebulosas individuais;
- a linha contínua é o ajuste relacionado às nebulosas individuais;
- os pontos representados por círculos remetem a grupos de nebulosas;
- a linha tracejada é o ajuste quanto ao grupo de nebulosas.

O resultado da Lei de Hubble demonstra que os objetos luminosos distantes do universo, como as nebulosas, estão em movimento em relação a nós (movimento relativo), uma vez que mostram um aumento da velocidade relativa (velocidade em relação a um referencial, no caso, nós) através do aumento da distância. O fato é que a Lei de Hubble evidencia que as nebulosas ou grupos de nebulosas não estão fixas, mas sim se movimentando, aumentando a distância relativa, traduzindo: o universo está inflando como um balão.

Note que as discussões que derivam da Lei de Hubble recriam um histórico de como são tecidos pareceres científicos a partir de um ajuste linear de dados, ou, em outras palavras, como uma metodologia baseada na análise de dados se aplica a um certo nicho específico de conhecimento humano, nesse caso a física e a cosmologia.

A figura a seguir mostra um gráfico de ações em tempo real, retirado do site Investing.com. Nesta figura está relacionado o volume de ações da Petrobrás, empresa cuja atuação influencia nos custos de vida de todo brasileiro. O investidor usa sua intuição, principalmente para perceber se o gráfico não descreve um sistema muito latente.



Volume de ações da Petrobrás comercializadas em tempo real.

Fonte: Investing. Elaboração: CEPED/UFSC (2022).

Caso o sistema acima não seja muito volátil, ou de outro modo apresente alguma regularidade ao longo de um período, então é estratégico pensar em qual tipo de estrutura matemática descreve esse instantâneo do mercado, sendo que fica claro: não é uma reta.

3.2 Referências de estudo

Nos atuais tempos da internet há muito material consolidado e livre acerca do tratamento de dados e sua grande área correlacionada, por isso foram citados nichos de conhecimento diretamente atrelados ao contexto de análise de dados como: cursos de estatística aplicados, tratamento de dados experimentais muitas vezes disponibilizados gratuitamente no site de várias universidades, USP, INPA, UNIFESP, MIT etc.

Ao longo do estudo foi evidenciada a potencialidade do tratamento da informação e como a sociedade está imersa no domínio de planilhas e *softwares* combinados a dados estatísticos.



Potencialidade do tratamento da informação.

Fonte: Freepik. Elaboração: CEPED/UFSC (2022).

Lógico que para esses estudos um conhecimento de linguagens de programação é muito interessante, no entanto é importante inicialmente ter o domínio de planilhas para a grande maioria dos propósitos. Contudo a procura de documentação acerca de estatística aplicada a planilhas, análise de dados usando SciDAVis, R ou Python é bastante recomendada para aporte prático do tema.



DESTAQUE

Atente-se que nem sempre a leitura está subjugada a ferramentas computacionais, uma vez que a intuição acerca de um gráfico é humana em vários casos, sendo o analisador que vai inferir e guiar a análise numérica, respondendo, por exemplo, se o conjunto de pontos tem comportamento de reta, ou parábola.

O fato é que a análise de dados é uma área altamente interdisciplinar que se compõe com a estatística, matemática, programação, resolução de problemas acadêmicos e científicos, desenvolvimento de técnicas numéricas etc. Dado que vários estudos, mesmo com o uso de computadores são muito demorados, simplificações são muito bem-vindas. Um exemplo disso é utilizar um fit de menor grau, quando a correlação é suficientemente satisfatória, comumente acima de 0,9.

Ao chegar até aqui você finalizou o estudo desta unidade! Acesse a atividade avaliativa no ambiente virtual de aprendizagem para refletir sobre o que foi apresentado até aqui.

Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

CARVALHO, Marília Sá; SOUZA-SANTOS, Reinaldo. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 21, p. 361-378, 2005.

COMARELA, Giovanni et al. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. SCHAEFFER FILHO, Alberto Egon; CORDEIRO, Weverton Luis da Costa; CAMPISTA, Miguel Elias Mitre (ed.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, Porto Alegre, Sociedade Brasileira de Computação, p. 246-295, 2019.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. **Big Data para leigos**. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

LEITE, Angela; FUITA, Hiroco. **Aplicações da matemática: administração, economia e ciências contábeis**. Boston, Cengage Learning, 2008.

PINTO, José Carlos; SCHWAAB, Marcio. **Análise de Dados Experimentais: I. Fundamentos de Estatística e Estimação de Parâmetros**. Rio de Janeiro, Editora E-papers, 2007.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

SOUZA, Emanuel Fernando Maia de; PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. **Software Livre R: aplicação estatística**. 2014. Universidade Federal da Paraíba.

VITALI, Marieli Mezari. Estatística sem matemática para psicologia. **Revista Brasileira de Psicodrama**, São Paulo, v. 27, n. 1, p. 139-144, 2019.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.