Universidade de Brasília

EXPERIMENTAÇÃO 5 – DEPOIS DO EXPERIMENTO

RESUMO

Como decidir qual análise vai usar

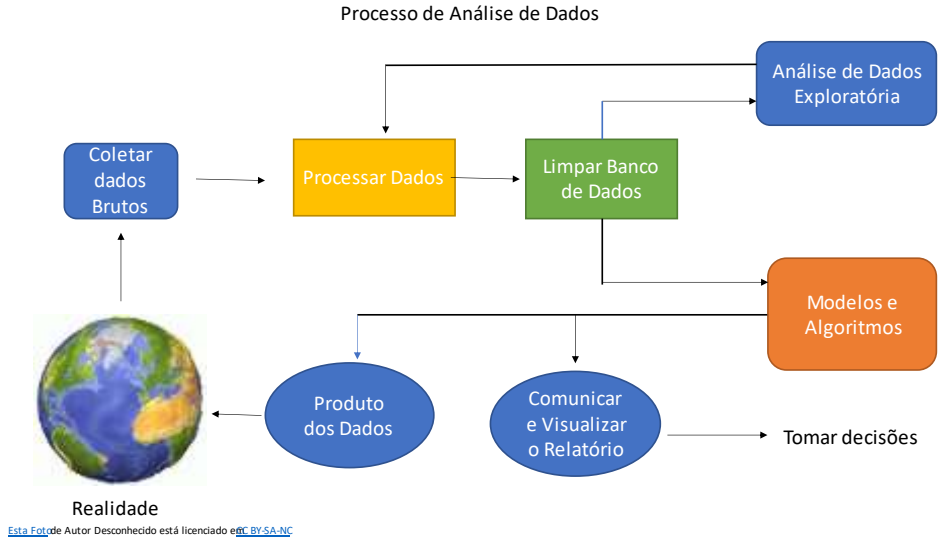
User

Concepta McManus, José Jivago Rolo, Andrea Queiroz Maranhão, Sonia Nair Báo, Felipe Pimentel, Daniel Pimentel

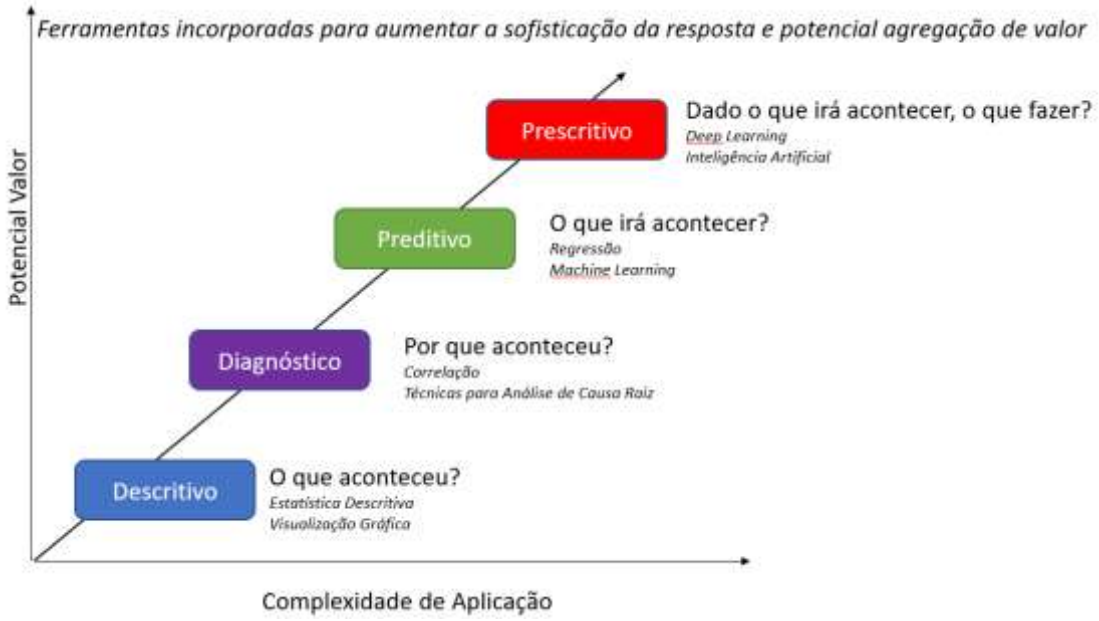
2023

Sumário

Depois do Experimento



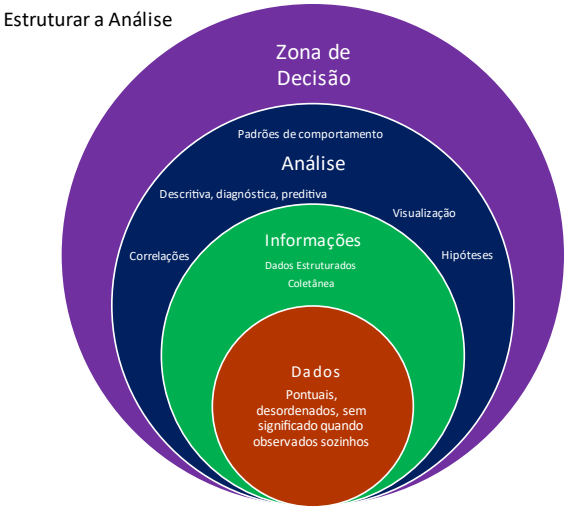
Análise dos dados



- *ANÁLISE DESCRITIVA*
    - *Organização dos dados por meio de classificação, contagem ou mensuração*
    - *Dados apresentados de forma clara por meio de tabelas, gráficos e medidas resumo (posição e variabilidade),*

- *Não tem conclusões analíticas.*

- *ANÁLISE INFERENCIAL*
  - *Permite realizar inferências (conclusões e analíticas) pela aplicação de testes de hipóteses e/ou construção de intervalos de confiança.*
  - *utilizando-se amostras para inferir os dados reais da população (parâmetros), existindo uma margem de erro.*
  - *A exceção é o censo, quando toda a população é pesquisada.*



Quais são os métodos de estimação de mínimos quadrados e de máxima verossimilhança?

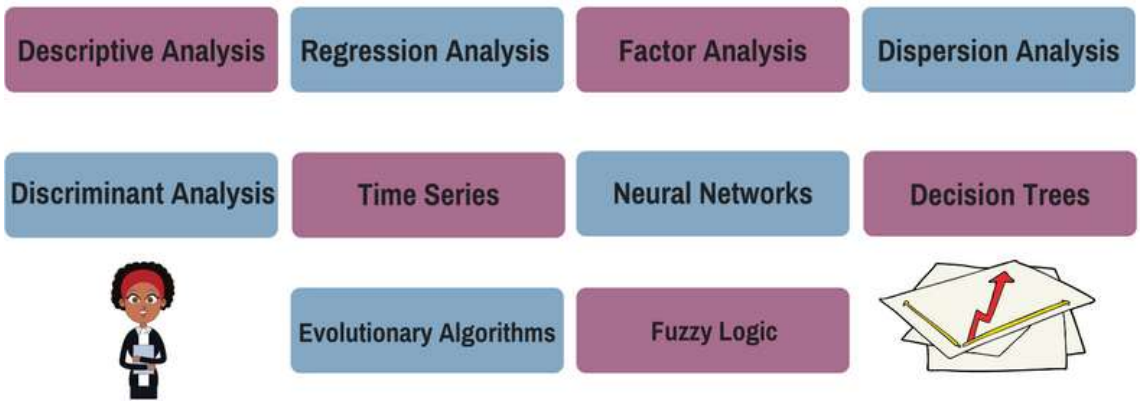
Duas abordagens comumente usadas para estimar parâmetros populacionais a partir de uma amostra aleatória são o método de estimação de máxima verossimilhança (padrão) e o método de estimação de mínimos quadrados.

Método de estimativa de máxima verossimilhança (MLE)

A função de verossimilhança indica a probabilidade da amostra observada em função dos valores de parâmetros possíveis. Portanto, maximizar a função de verossimilhança determina os parâmetros com maior probabilidade de produzir os dados observados. Do ponto de vista estatístico, o MLE é geralmente recomendado para grandes amostras porque é versátil, aplicável à maioria dos modelos e diferentes tipos de dados e produz as estimativas mais precisas.

Método de estimativa de mínimos quadrados (LSE)

As estimativas de mínimos quadrados são calculadas ajustando uma linha de regressão aos pontos de um conjunto de dados que tem a soma mínima dos desvios ao quadrado (erro de mínimos quadrados). Na análise de confiabilidade, a linha e os dados são plotados em um gráfico de probabilidade.



Análise de dados em Pesquisas Qualitativas

A análise de dados e a pesquisa de dados qualitativos funcionam um pouco diferente dos dados numéricos, pois os dados de qualidade são compostos por palavras, descrições, imagens, objetos e, às vezes, símbolos. Obter uma visão dessas informações complicadas é um processo complicado. Portanto, é normalmente usado para pesquisa exploratória e análise de dados.

Encontrando padrões nos dados qualitativos

- Método baseado em **palavras** é a técnica global mais confiável e amplamente usada para pesquisa e análise de dados.
    - O processo de análise de dados em pesquisas qualitativas é manual.
    - Os pesquisadores costumam ler os dados disponíveis e encontrar palavras repetitivas ou comumente usadas.
    - Por exemplo, ao estudar dados coletados em países africanos para entender os problemas mais urgentes que as pessoas enfrentam, os pesquisadores podem descobrir que "comida" e "fome" são as palavras mais comumente usadas e irão destacá-las para análise posterior.
- O contexto de **palavras-chave**
    - O pesquisador tenta compreender o conceito analisando o contexto em que os participantes usam uma determinada palavra-chave.
    - Por exemplo, os pesquisadores que conduzem pesquisas e análises de dados para estudar o conceito de 'diabetes' entre os entrevistados podem analisar o contexto de quando e como o entrevistado usou ou se referiu à palavra 'diabetes'.
- A técnica baseada em **escrutínio**
    - Comparar e contrastar é o método amplamente usado nessa técnica para diferenciar como um texto específico é semelhante ou diferente um do outro.
    - Por exemplo: Para saber a "importância do médico residente em uma empresa", os dados coletados são divididos em pessoas que acham

necessário contratar um médico residente e aquelas que acham desnecessário.

- Comparar e contrastar é o melhor método que pode ser usado para analisar as enquetes com tipos de perguntas de resposta única.

- As **metáforas**
  - podem ser usadas para reduzir a pilha de dados e encontrar padrões nela para que seja mais fácil conectar os dados com a teoria.
- **O particionamento de variáveis**
  - dividir variáveis para que os pesquisadores possam encontrar descrições e explicações mais coerentes a partir dos enormes dados.

Métodos usados para análise de dados em pesquisas qualitativas

- **Análise de Conteúdo**
  - Amplamente aceita e a técnica mais frequentemente empregada para análise de dados na metodologia de pesquisa.
  - Pode ser usado para analisar as informações documentadas de texto, imagens e, às vezes, de itens físicos.
  - Depende das questões de pesquisa para prever quando e onde usar este método.
- **Análise narrativa**:
  - Analisar o conteúdo coletado de várias fontes, como entrevistas pessoais, observação de campo e pesquisas.
  - Na maioria das vezes, as histórias ou opiniões compartilhadas pelas pessoas se concentram em encontrar respostas para as perguntas da pesquisa.
- **Análise do discurso**:
  - Semelhante à análise narrativa, é usada para analisar as interações com as pessoas.
  - Este método particular considera o contexto social sob o qual ou dentro do qual a comunicação entre o pesquisador e o entrevistado ocorre.
  - Se concentra no estilo de vida e no ambiente do dia a dia, ao mesmo tempo em que obtém qualquer conclusão.
- **Teoria fundamentada:**
  - Quando você deseja explicar por que um fenômeno específico aconteceu, o melhor recurso é usar a teoria fundamentada para analisar dados de qualidade.
  - A teoria fundamentada é aplicada para estudar dados sobre uma série de casos semelhantes ocorrendo em diferentes ambientes.
  - Quando os pesquisadores estão usando esse método, eles podem alterar explicações ou produzir novas até chegarem a alguma conclusão.

Análise de dados em Pesquisa Quantitativa

Fase I: Validação de Dados

A validação de dados é feita para entender se a amostra de dados coletados está de acordo com os padrões predefinidos ou se é uma amostra de dados tendenciosa novamente dividida em quatro estágios diferentes

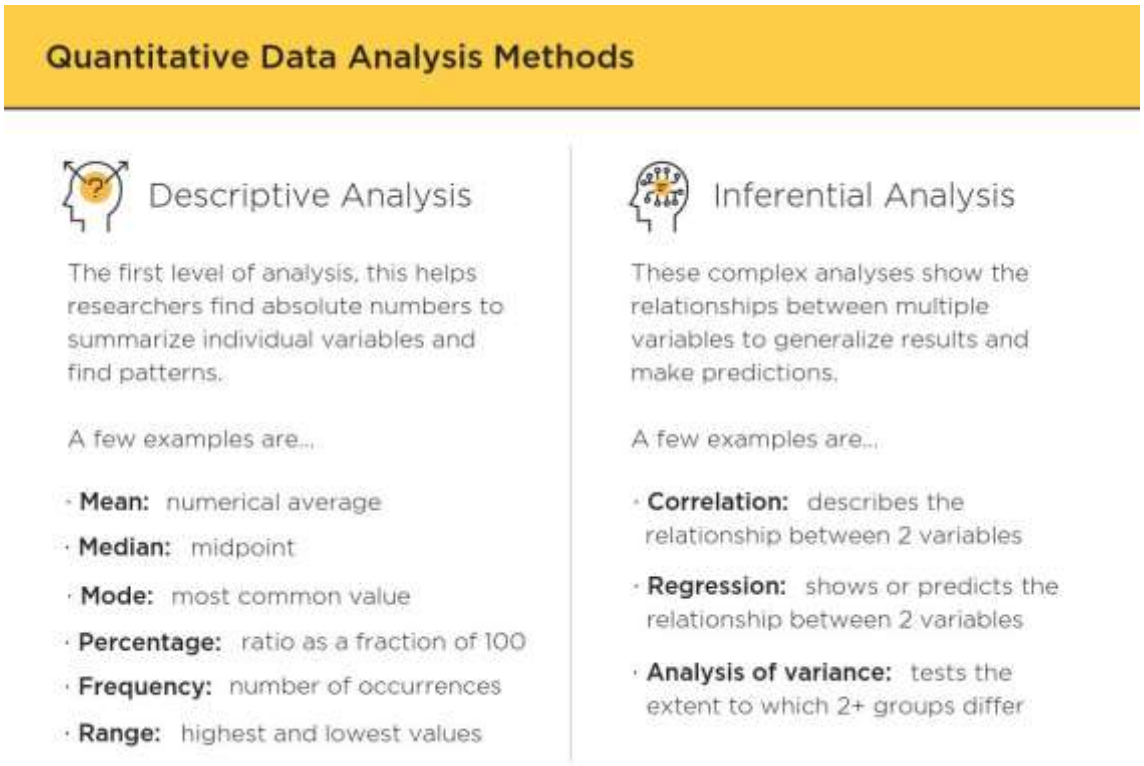
- Fraude: Para garantir que um ser humano real registre cada resposta à pesquisa ou ao questionário
- Triagem: Para garantir que cada participante ou entrevistado seja selecionado ou escolhido de acordo com os critérios da pesquisa
- Procedimento: Para garantir que os padrões éticos foram mantidos durante a coleta da amostra de dados
- Completude: Para garantir que o entrevistado respondeu a todas as perguntas em uma pesquisa online. Caso contrário, o entrevistador havia feito todas as perguntas elaboradas no questionário.

Fase II: Edição de dados

- Dados de pesquisa podem ter erros.
- Os respondentes às vezes preenchem alguns campos incorretamente ou às vezes os ignoram acidentalmente.
- A edição de dados é um processo em que os pesquisadores devem confirmar que os dados fornecidos estão livres de tais erros. Eles precisam realizar verificações e verificações de outlier necessárias para editar a edição bruta e torná-la pronta para análise.

Fase III: Codificação de Dados

- Fase mais crítica da preparação de dados associada ao agrupamento e atribuição de valores às respostas da pesquisa.
- Se uma pesquisa for concluída com um tamanho de amostra de 1.000, o pesquisador criará uma faixa etária para distinguir os respondentes com base em sua idade.
- Assim, torna-se mais fácil analisar pequenos depósitos de dados em vez de lidar com uma enorme pilha de dados.



Métodos usados para análise de dados em pesquisas quantitativas

A Análise de Dados é definida como um processo de inspeção, limpeza, transformação e modelagem de dados com o objetivo de descobrir informações úteis, informando conclusões e apoiando a tomada de decisões.

A análise de dados pode ser separada e organizada em 6 tipos, organizados em uma ordem crescente de dificuldade.

- Análise descritiva
- Análise Exploratória
- Análise Inferencial
- Análise Preditiva
- Análise Causal
- Análise Mecanicista

Qualitative Data Preparation and Analysis

Get familiar with the data

Start by reading the data several times to get familiar with it and start looking for basic observations or patterns. This also includes transcribing the data.

Revisit research objectives

Revisit the research objective and identify the questions that can be answered through the collected data.

Develop a framework

Identify broad ideas, concepts, behaviors, or phrases and assigns codes to them. This is helpful for structuring and labeling the data.

Identify patterns and connections

Start identifying themes, looking for the most common responses to questions, identifying data or patterns that can answer research questions, and finding areas that can be explored further.

7MAIN TYPES OF STATISTICAL ANALYSIS

DESCRIPTIVE TYPE

As the name suggests, the descriptive statistic is used to describe! It describ
the basic features of data and shows or summarizes data in a rational wa

However, descriptive statistics do not allow making conclusions. You can not g
conclusions and make generalizations that extend beyond the data at han

INFERENTIAL TYPE

Inferential statistics is a result of more complicated mathematical estimations, and allow us to infer trends about a larger population based on samples of "subjects" taken from it.

This type of statistical analysis is used to study the relationships between variables within a sample, and you can make conclusions, generalizations or predictions about a bigger population.

PREDICTIVE ANALYTICS

Predictive analytics uses techniques such as statistical algorithms a
machine learning to define the likelihood of future results, behavio
and trends based on both new and historical da

It is important to note that no statistical method can "predict" the future with 100
surety. Businesses use these statistics to answer the question "What might happen?

PRESCRIPTIVE ANALYTICS

Prescriptive analytics is a study which examines data to answer the question "What should be done?"

Prescriptive analytics aim to find the optimal recommendations for a decision making process. It is all about providing advice.

CAUSAL ANALYSIS

When you would like to understand and identify the reasons wh
things are as they are, causal analysis comes to help. This type
analysis answers the question "Why

The business world is full of events that lead to failure. The causal seeks
identify the reasons why? It is better to find causes and to treat them inste
of treating symptom

EXPLORATORY DATA ANALYSIS (EDA)

EDA is an analysis approach that focuses on identifying general patterns in the data and to find previously unknown relationships

EDA alone should not be used for generalizing or predicting. EDA is used for taking a bird's eye view of the data and trying to make some feeling or sense of it. Commonly, It is the first step in data analysis, performed before other formal statistical techniques.

MECHANISTIC ANALYSIS

The mechanistic analysis is about understanding the exact changes
given variables that lead to changes in other variables. Howeve
mechanistic does not consider external influenc

The assumption is that a given system is affected by the interaction of its ow
components. It is useful on those systems for which there are very clear definition

1. Análise Descritiva

Objetivo - Descrever ou Resumir um Conjunto de Dados

Descrição:

- A primeira análise realizada
- Gera resumos simples sobre amostras e medições
- estatísticas descritivas comuns (medidas de tendência central, variabilidade, frequência, posição, etc)

Exemplo: Veja a página de estatísticas do COVID-19 no google, por exemplo, o gráfico de linha é apenas um puro resumo dos casos / óbitos, uma apresentação e descrição da população de um determinado país infectado pelo vírus

Este método é usado para descrever as características básicas de tipos versáteis de dados em pesquisa. Ele apresenta os dados de uma maneira tão significativa que o padrão nos dados começa a fazer sentido. No entanto, a análise descritiva não vai além de tirar conclusões. As conclusões são novamente baseadas nas hipóteses que os pesquisadores formularam até agora.

Medidas de Frequência

- Contagem, porcentagem, frequência
- É usado para denotar o lar com frequência em que um determinado evento ocorre.
- Os pesquisadores usam quando desejam mostrar a frequência com que uma resposta é dada.

Medidas de tendência central

- Média, mediana, modo
- O método é amplamente utilizado para demonstrar a distribuição por vários pontos.
- Os pesquisadores usam esse método quando desejam mostrar a resposta indicada de maneira mais comum ou mediana.

Medidas de dispersão ou variação

- Faixa, variância, desvio padrão

- Aqui, o campo é igual a pontos altos / baixos.
- Desvio padrão da variância = diferença entre a pontuação observada e a média
- É usado para identificar a dispersão das pontuações, declarando intervalos.
- Os pesquisadores usam esse método para mostrar os dados espalhados. Isso os ajuda a identificar a profundidade até a qual os dados são espalhados, o que afeta diretamente a média.

Medidas de posição

- Classificações de percentis, classificações de quartil
- Ele se baseia em pontuações padronizadas que ajudam os pesquisadores a identificar a relação entre as diferentes pontuações.
- Geralmente é usado quando os pesquisadores desejam comparar as pontuações com a contagem média.

Resumo: A Análise Descritiva é a primeira etapa da análise em que você resume e descreve os dados que possui usando estatísticas descritivas e seu resultado é uma apresentação simples de seus dados.

2. Análise Exploratória (EDA)

Objetivo - examinar ou explorar dados e encontrar relações entre variáveis que eram anteriormente desconhecidas

Descrição:

- EDA ajuda a descobrir relações entre medidas em seus dados, que não são evidências da existência da correlação, conforme denotado pela frase (correlação não implica causalidade)
- útil para descobrir novas conexões - formando hipóteses e direciona o planejamento do projeto e a coleta de dados

Exemplo: A mudança climática é um tópico cada vez mais importante à medida que a temperatura global está aumentando gradualmente ao longo dos anos. Um exemplo de EDA sobre mudança climática é pegar o aumento da temperatura ao longo dos anos, digamos 1950 a 2020, por exemplo, e o aumento das atividades humanas e industrialização, e formar relacionamentos a partir dos dados, por exemplo, aumento do número de fábricas, carros na estrada e voos de avião aumentam correlatos.

Resumo: A EDA explora dados para encontrar relações entre medidas que nos dizem que existem, sem causa. Eles podem ser usados para formular hipóteses.

3. Análise inferencial (IA)

Objetivo - As estatísticas inferenciais são usadas para fazer previsões sobre uma população maior após pesquisa e análise de dados da amostra coletada da população representativa. Por exemplo, você pode perguntar a cerca de 100 públicos estranhos em um cinema se eles gostam do filme que estão assistindo. Os pesquisadores então usam estatísticas inferenciais na amostra coletada para raciocinar que cerca de 80-90% das pessoas gostam do filme.

Descrição:

- Usando dados estimados que valorizam em população e dão uma medida de incerteza (desvio padrão) em sua estimativa
- A precisão da inferência depende muito do esquema de amostragem; se a amostra não for representativa da população, a generalização será imprecisa (consulte o Teorema do Limite Central)

Exemplo: A ideia de inferir sobre a população em geral com uma amostra menor é bastante intuitiva, muitas estatísticas que você vê na mídia e na internet são inferenciais, uma previsão de um evento a partir de uma pequena amostra. Para dar um exemplo, um estudo de psicologia para os benefícios do sono, um total de 500 pessoas envolvidas no estudo, quando acompanhadas com os candidatos, relataram ter melhor atenção geral e bem-estar com 7 a 9 horas de sono, enquanto aqueles com menos sono e mais sono sofreram com redução da atenção e energia. Este relatório de 500 pessoas era apenas uma pequena porção de 7 bilhões de pessoas no mundo, portanto, uma inferência de uma população maior.

São duas áreas da estatística inferencial.

1. Parâmetros de estimativa: obtém estatísticas dos dados de pesquisa da amostra e demonstra algo sobre o parâmetro da população.
2. Teste de hipótese: trata-se de amostrar dados de pesquisa para responder às perguntas de pesquisa da pesquisa. Por exemplo, os pesquisadores podem estar interessados em entender se a nova tonalidade de batom

lançada recentemente é boa ou não, ou se as cápsulas multivitamínicas ajudam as crianças a ter um melhor desempenho nos jogos.

Esses são métodos de análise sofisticados usados para mostrar a relação entre diferentes variáveis em vez de descrever uma única variável. Geralmente é usado quando os pesquisadores querem algo além dos números absolutos para entender a relação entre as variáveis.

Métodos comumente usados para análise de dados em pesquisas.

Correlação: quando os pesquisadores não estão conduzindo uma pesquisa experimental na qual os pesquisadores estão interessados em entender a relação entre duas ou mais variáveis, eles optam por métodos de pesquisa correlacional.

Tabulação cruzada: também chamada de tabelas de contingência, a tabulação cruzada é usada para analisar a relação entre variáveis múltiplas. Suponha que os dados fornecidos tenham categorias de idade e sexo apresentadas em linhas e colunas. Uma tabulação cruzada bidimensional ajuda a uma análise e pesquisa de dados contínua, mostrando o número de homens e mulheres em cada categoria de idade.

Análise de regressão: para compreender a forte relação entre duas variáveis, os pesquisadores não olham além do método de análise de regressão primário e comumente usados, que também é um tipo de análise preditiva usada. Neste método, você tem um fator essencial chamado variável dependente. Você também tem várias variáveis independentes na análise de regressão. Você empreende esforços para descobrir o impacto das variáveis independentes na variável dependente. Os valores das variáveis independentes e dependentes são assumidos como sendo verificados de uma maneira aleatória sem erros.

Tabelas de frequência: o procedimento estatístico é usado para testar o grau em que dois ou mais variam ou diferem em um experimento. Um grau considerável de variação significa que os resultados da pesquisa foram significativos. Em muitos contextos, o teste ANOVA e a análise de variância são semelhantes.

Análise de variância: o procedimento estatístico é usado para testar o grau em que dois ou mais variam ou diferem em um experimento. Um grau considerável de variação significa que os resultados da pesquisa foram significativos. Em muitos contextos, o teste ANOVA e a análise de variância são semelhantes.

Análises multivariadas: métodos que estudam simultaneamente três ou mais variáveis (características), incluindo Análises de correspondência (associação entre variáveis categóricas gerando tabelas de contingência), componentes principais (análise dos dados de forma reduzida, eliminando as sobreposições e escolhendo a forma mais representativa dos dados a partir de combinações lineares das variáveis originais), fatorial (descrever a variabilidade original de um conjunto de variáveis em um número reduzido de variáveis latentes), cluster (agrupar os elementos selecionados em grupos com características similares entre si de maneira que os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características), discriminante (discriminar e classificar objetos), regressão múltipla (modelos que descrevam de maneira razoável relações entre várias variáveis explicativas de um determinado processo) e Modelagem de equações estruturais (combina aspectos de regressão múltipla e de análise fatorial, com o objetivo de estimar simultaneamente uma série de relações de dependência).

Resumo: IA extrapola e generaliza as informações do grupo maior com uma amostra menor para gerar análises e previsões.

4. Análise preditiva (PA)

Objetivo - Usar dados históricos ou atuais para encontrar padrões e fazer previsões sobre o futuro

Descrição:

- A precisão das previsões depende das variáveis de entrada
- A precisão também depende dos tipos de modelos, um modelo linear pode funcionar bem em alguns casos e vice-versa
- Usar uma variável para prever outra não denota relações causais

Exemplo: A eleição de 2020 nos EUA é um tópico popular e muitos modelos de previsão são construídos para prever o candidato vencedor FiveThirtyEight fez uma ótima previsão da eleição para 2016 e está de volta a ela em 2020. A análise de previsão para uma eleição exigiria variáveis de entrada, como dados históricos de pesquisas, tendências e os dados das pesquisas atuais para obter uma boa previsão. Algo tão grande como uma eleição não seria apenas usar um modelo linear, mas um modelo complexo com certas afinações para melhor servir ao seu propósito.

Resumo: PA coleta dados do passado e do presente para fazer previsões sobre o futuro.

5. Análise Causal (AC)

Objetivo - Examina a causa e o efeito das relações entre as variáveis, com foco em encontrar a causa de uma correlação.

Descrição:

- Para encontrar a causa, você deve questionar se as correlações observadas que levam à sua conclusão são válidas, pois apenas olhar para os dados (superfície) não o ajudará a descobrir os mecanismos ocultos subjacentes às correlações
- Aplicado em estudos randomizados com foco na identificação de causalidade
- o padrão ouro na análise de dados, estudos científicos onde a causa do fenômeno deve ser extraída e individualizada, como separar o joio do trigo

Desafios: Bons dados são difíceis de encontrar e requerem pesquisas e estudos caros. Esses estudos são analisados em conjunto (vários grupos), e as relações observadas são apenas efeitos médios (média) de toda a população (o que significa que os resultados podem não se aplicar a todos)

Exemplo: Digamos que você queira testar este novo medicamento que melhora a força e o foco humanos e, para isso, faça testes de controle randomizados para testar o efeito do medicamento. Você compara a amostra de candidatos para seu novo medicamento com os candidatos que recebem controle simulado com alguns testes de força, foco e atenção geral e observa como o medicamento afeta o resultado

Resumo: AC é descobrir a relação causal entre variáveis, mudar uma variável e o que acontece com outra.

6. Análise mecanicista (MA)

Objetivo - compreender as mudanças exatas nas variáveis que levam a outras mudanças em outras variáveis

Descrição:

- Aplicado em ciências físicas ou de engenharia, situações que requerem alta precisão e pouco espaço para erro (apenas ruído nos dados é erro de medição)
- Projetado para compreender um processo biológico ou comportamental, a fisiopatologia de uma doença ou o mecanismo de ação de uma intervenção. (por NIH)

Exemplo: Muitas pesquisas de pós-graduação e tópicos complexos são exemplos adequados, mas, para colocá-lo de uma maneira simples, digamos que um experimento seja feito para simular uma fusão nuclear segura e eficaz para fornecer energia ao mundo, um estudo mecanicista implicaria na necessidade de equilíbrio de análise controlando e manipulando variáveis com medidas altamente precisas de ambas as variáveis e os resultados desejados. É esse modus operandi (estratégia) intrincado e meticuloso em relação a esses grandes tópicos que permite descobertas científicas e o avanço da sociedade.

Resumo: MA é, de certa forma, uma análise preditiva, mas modificada para lidar com estudos que requerem alta precisão e metodologias meticulosas para ciências físicas ou de engenharia.

Séries temporais

As séries temporais existem ao nosso redor – tanto na ciência de dados quanto no mundo cotidiano. Em sua essência, os dados de séries temporais são dados registrados em intervalos regulares ou períodos de tempo. Qualquer valor não estacionário que dependa do tempo pode fazer parte de uma série temporal.

Como analista de dados, você pode usar dados de séries temporais para descobrir tendências subjacentes ou causas de determinados padrões ao longo do tempo.

Se o conceito soa familiar, é porque você provavelmente se referiu a dados de séries temporais em sua vida cotidiana sem saber! Se você usou um smartwatch para rastrear seus passos durante um período de uma semana ou escreveu suas despesas todos os dias durante um mês, você gravou com sucesso os dados da série temporal.

Quais são os componentes dos dados de séries temporais?

Qualquer conjunto de dados de série temporal pode incluir um ou mais dos quatro componentes a seguir:

Tendência:

- Uma tendência refere-se a um movimento ascendente ou descendente consistente e de longo prazo em uma série. Ao contrário da variação sazonal, uma tendência é inesperada e não imediatamente identificável. Uma tendência na qual podemos encontrar a causa é chamada de determinística, enquanto uma tendência inexplicável é chamada de estocástica. Por exemplo, se um novo autor lança um livro e as vendas do livro disparam, essa tendência seria determinística.

Ciclo:

- Um ciclo é um movimento para cima e para baixo que ocorre em torno de uma tendência. Ao contrário da variação sazonal, um ciclo não tem um tempo preciso e igual entre os períodos de tempo e, portanto, não é previsível.

Sazonalidade:

- Ao contrário de uma tendência, a sazonalidade refere-se a variações que ocorrem em uma frequência previsível e fixa. Por exemplo, as vendas de sorvetes aumentam no verão porque o clima está mais quente e mais pessoas desejam um doce doce.

Irregularidade:

- também chamada de ruído, a irregularidade é o que sobra quando você tira a sazonalidade ou as tendências do conjunto de dados. As irregularidades são aleatórias e imprevisíveis. Um excelente exemplo de variações irregulares são as mudanças nos preços das ações.

Ao identificar esses componentes em um conjunto de dados, você pode realizar transformações e ajustes – um dos mais comuns são os ajustes sazonais – que levam a previsões mais precisas (que exploraremos mais adiante).

O que são séries temporais irregulares?

Quando os valores são coletados em intervalos de tempo iguais e consistentes, a série temporal é chamada de regular.

Mas quando as medições são coletadas em intervalos imprevisíveis e irregulares, a série temporal é chamada de irregular. Por exemplo, um sensor em um telefone registra informações apenas quando o dispositivo é retirado ou um caixa eletrônico registra saques à medida que ocorrem. Ambos são exemplos de dados de séries temporais irregulares. Embora desafiador, ainda é possível modelar dados de séries temporais irregulares usando vários métodos, como modelos neurais de equações diferenciais ordinárias ou redes de interpolação.

Quais são alguns exemplos de dados de séries temporais?

Nossas vidas pessoais e profissionais estão repletas de exemplos de dados de séries temporais.

No mundo dos negócios, por exemplo, os empreendedores acompanham as mudanças nos lucros, retornos ou vendas ao longo de um ano. Se você trabalha com investimentos, por outro lado, pode manter um registro do PIB, ganhos ou preços das ações.

Talvez alguns dos exemplos mais comuns de séries temporais hoje sejam: preços diários de ações, temperaturas diárias, o número de pessoas vacinadas em um determinado dia ou a porcentagem de pessoas desempregadas em um determinado mês.

Qual é a diferença entre dados de séries temporais e dados transversais?

É fácil identificar uma série temporal, pois ela consiste em apenas dois elementos: um período igual e claramente definido e um único valor rastreado no final de cada um desses períodos.

Por outro lado, os dados transversais consistem em várias variáveis rastreadas durante um único período fixo de tempo – por exemplo, a renda média de várias cidades durante um único ano ou uma pesquisa de uma população. Nos dados transversais, o foco está mais na comparação de várias entidades e menos na análise de dados ao longo do tempo para fazer previsões.

Claro, também é possível combinar dados de séries temporais e dados transversais. Esse novo conjunto de dados é conhecido como dados de painel. Com dados em painel, você pode, por exemplo, acompanhar o efeito dos benefícios sociais sobre o desemprego durante um período de tempo.

Para que serve a análise de séries temporais?

Geralmente, a análise de séries temporais consiste em observar pontos de dados e todas as suas variações (ou componentes) ao longo de um período de tempo.

Ao observar dados anteriores, os analistas podem tirar conclusões inteligentes sobre o comportamento em todos os setores, incluindo negócios, finanças, imóveis e varejo, e usar essas informações para tomar decisões futuras (também conhecidas como previsão de séries temporais).

A análise de séries temporais pode ser usada para:

Tome decisões sobre valores futuros, com base em valores passados. Por exemplo, você pode definir preços de varejo para trajes de banho com base em variações sazonais em dados de séries temporais

Preveja valores futuros, com base em valores passados. Por exemplo, você pode prever temperaturas gerais com base em décadas de registros meteorológicos

Identifique irregularidades ou ruídos em séries temporais. Por exemplo, você pode detectar atividades financeiras fraudulentas com base no histórico de atividades financeiras.

A análise de séries temporais pode ser especialmente útil para qualquer pessoa que trabalhe em uma função que precise tomar decisões e planejar políticas.

O que é previsão de séries temporais?

A previsão de séries temporais é um tipo de análise de séries temporais que analisa dados históricos para escolher um modelo para prever dados futuros. Quanto mais holísticos os dados, mais precisa será a previsão.

Por exemplo, uma série temporal de carros comprados nos últimos 50 anos pode gerar previsões mais precisas do que uma série temporal de carros comprados nos últimos dois anos. A previsão de séries temporais é um uso central de dados de séries temporais e, muitas vezes, o principal.

E, na previsão, você pode optar por analisar uma única variável para prever valores futuros (conhecida como previsão de série temporal univariada) ou usar várias variáveis para prevê-los (conhecida como previsão de série temporal multivariada).

Principais conclusões

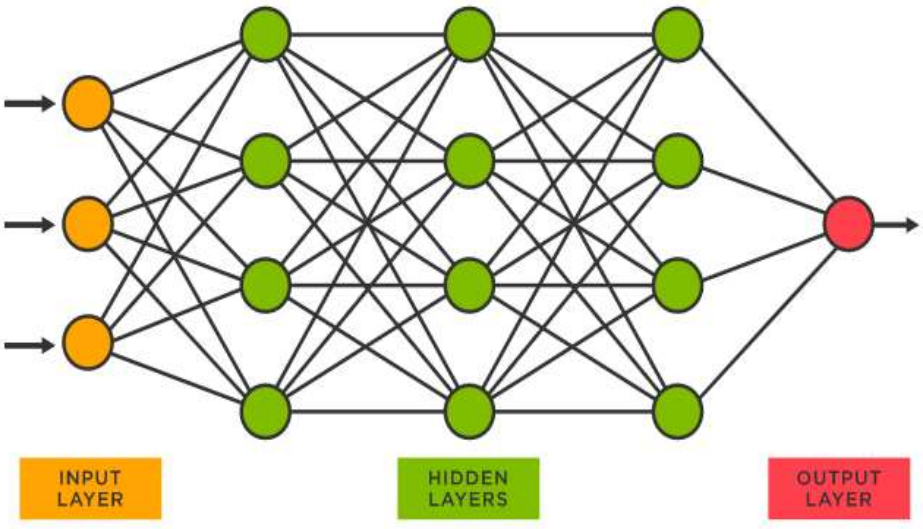
- Dados de série temporal são dados registrados em intervalos regulares ou períodos de tempo
- Um ou mais dos quatro componentes compõem os dados da série temporal: uma tendência, um ciclo, sazonalidade e irregularidades
- A análise de séries temporais pode ser usada para identificar irregularidades, entender resultados passados, tomar decisões sobre valores futuros ou prever valores
- A principal função dos dados de séries temporais é fazer previsões sobre valores futuros, também conhecido como previsão de séries temporais.

Métodos baseados em inteligência artificial, aprendizado de máquina e algoritmos heurísticos

Esses métodos modernos atraem a atenção dos cientistas de dados com seus recursos estendidos e a habilidade de resolver tarefas não tradicionais. Além disso, eles podem ser facilmente e eficientemente implementados e executados por ferramentas e sistemas de software especiais.

Redes Neurais Artificiais

- Um dos novos e modernos tipos de métodos de análise de dados mais populares. De acordo com http://neuralnetworksanddeeplearning.com,"Redes neutras são um paradigma de programação de inspiração biológica que permite que um computador aprenda a partir de dados observacionais"
- Redes Neurais Artificiais (RNA), muitas vezes chamadas apenas de "rede neural", apresentam uma metáfora do cérebro para o processamento de informações.
- Esses modelos são modelos computacionais inspirados biologicamente. Eles consistem em um grupo interconectado de neurônios artificiais e processam informações usando uma abordagem de computação.

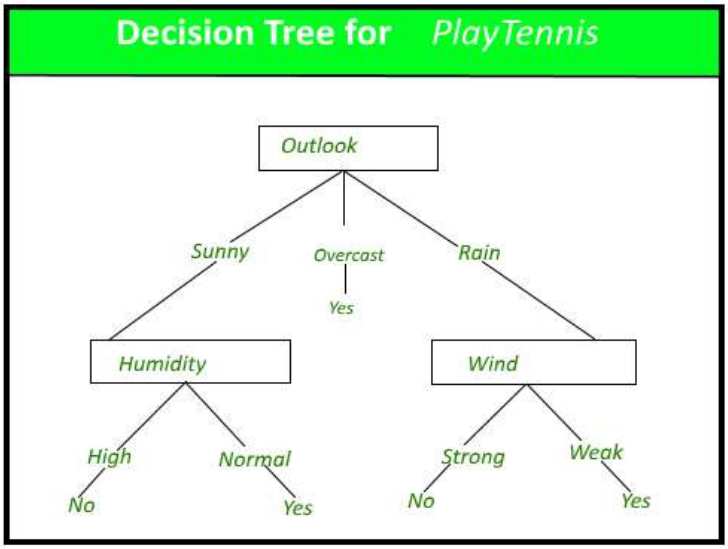


- As soluções avançadas de software ANN são sistemas adaptativos que mudam facilmente sua estrutura com base nas informações que fluem pela rede.
- A aplicação de redes neurais na mineração de dados é muito ampla. Eles têm uma alta capacidade de aceitação de dados ruidosos e alta precisão. A

mineração de dados baseada em redes neurais é pesquisada em detalhes. As redes neurais têm se mostrado sistemas muito promissores em muitos aplicativos de previsão e classificação de negócios.

Árvores de Decisão

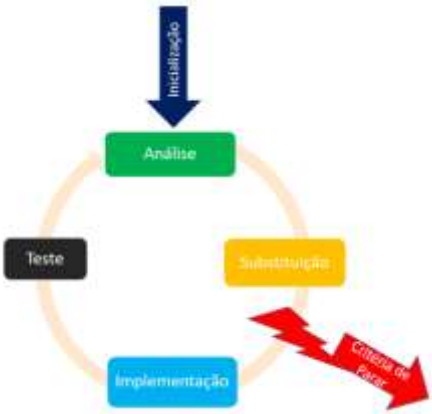
- Este é outro algoritmo de classificação muito popular e moderno em mineração de dados e aprendizado de máquina. A árvore de decisão é um diagrama em forma de árvore que representa um modelo de classificação ou regressão.
- Ele divide um conjunto de dados em subconjuntos cada vez menores (que contêm instâncias com valores semelhantes) enquanto, ao mesmo tempo, uma árvore de decisão relacionada é desenvolvida continuamente. A árvore é construída para mostrar como e por que uma escolha pode levar à outra, com a ajuda dos galhos.
- Entre os benefícios do uso de árvores de decisão estão: o conhecimento do domínio não é necessário; eles são fáceis de compreender; as etapas de classificação de uma árvore de decisão são muito simples e rápidas.



-

Programação Evolutiva

- A programação evolutiva em mineração de dados é um conceito comum que combina muitos tipos diferentes de análise de dados usando algoritmos evolutivos. Os mais populares deles são: algoritmos genéticos, programação genética e algoritmos coevolucionários.
- Na verdade, muitas agências de gerenciamento de dados aplicam algoritmos evolutivos para lidar com alguns dos maiores desafios de big data do mundo.
- Entre os benefícios dos métodos evolutivos estão:
  - são técnicas independentes de domínio
  - têm a capacidade de explorar grandes espaços de pesquisa descobrindo boas soluções
  - são relativamente insensíveis ao ruído
  - pode gerenciar a interação de atributos de uma maneira excelente.



Lógica Fuzzy

- é aplicada para lidar com a incerteza em problemas de mineração de dados. A modelagem em lógica difusa é um dos métodos e técnicas de análise de dados baseados em probabilidade.
- É um campo relativamente novo, mas tem grande potencial para extrair informações valiosas de diferentes conjuntos de dados.
- A lógica difusa é um tipo inovador de lógica de muitos valores em que os valores verdade das variáveis são um número real entre 0 e 1. Nesse termo, o valor verdade pode variar entre totalmente verdadeiro e totalmente falso. A

lógica difusa é aplicável quando o modelo contém parâmetros cujos valores não podem ser determinados com precisão ou esses valores contêm um nível de ruído muito alto.

Resumo de Análise de dados

- A análise descritiva resume os dados disponíveis e apresenta seus dados de uma maneira agradável.
- A análise exploratória de dados ajuda a descobrir correlações e relacionamentos entre variáveis em seus dados.
- A análise inferencial é para generalizar a população maior com um tamanho de amostra menor de dados.
- A análise preditiva ajuda a fazer previsões sobre o futuro com dados.
- A análise causal enfatiza em encontrar a causa de uma correlação entre as variáveis.
- A análise mecanicista serve para medir as mudanças exatas nas variáveis que levam a outras mudanças em outras variáveis.
- As poucas lições importantes acima incluem
- Correlação não implica causalidade
- EDA ajuda a descobrir novas conexões e formar hipóteses
- A precisão da inferência depende do esquema de amostragem
- Uma boa previsão depende das variáveis de entrada corretas
- Um modelo linear simples com dados suficientes geralmente resolve
- Usar uma variável para prever outra não denota relações causais
- Bons dados são difíceis de encontrar e para produzi-los requer pesquisas caras
- Os resultados dos estudos são feitos em conjunto e são efeitos médios e podem não se aplicar a todos

Considerações na análise de dados

- Os pesquisadores devem ter as habilidades necessárias para analisar os dados, sendo treinados para demonstrar um alto padrão de prática de pesquisa. Idealmente, os pesquisadores devem possuir mais do que uma compreensão básica da razão de selecionar um método estatístico em detrimento do outro para obter melhores percepções de dados.
- Normalmente, os métodos de pesquisa e análise de dados diferem por disciplina científica; portanto, obter aconselhamento estatístico no início da análise ajuda a elaborar um questionário de pesquisa, selecionar métodos de coleta de dados e escolher amostras.
- O objetivo principal da pesquisa e análise de dados é obter insights finais imparciais. Qualquer erro em ou manter uma mente tendenciosa para coletar dados, selecionar um método de análise ou escolher uma amostra do público pode fazer uma inferência tendenciosa.
- Irrelevante para a sofisticação usada em dados de pesquisa e análise é o suficiente para retificar as medidas de resultado objetivo mal definidas. Não importa se o design está errado ou as intenções não são claras, mas a falta de clareza pode enganar os leitores, portanto, evite a prática.
- O motivo por trás da análise de dados em pesquisas é apresentar dados precisos e confiáveis. Tanto quanto possível, evite erros estatísticos e encontre uma maneira de lidar com os desafios diários, como outliers, dados ausentes, alteração de dados, mineração de dados ou desenvolvimento de representação gráfica.

10 KEY TYPES OF DATA ANALYSIS METHODS

O que é análise de dados e por que é importante?

A análise de dados é, de forma simples, o processo de descobrir informações úteis por meio da avaliação de dados. Isso é feito por meio de um processo de inspeção, limpeza, transformação e modelagem de dados usando ferramentas analíticas e estatísticas, que exploraremos em detalhes mais adiante neste artigo.

Por que a análise de dados é importante? Analisar dados de forma eficaz ajuda as organizações a tomar decisões de negócios. Atualmente, os dados são coletados pelas empresas constantemente: por meio de pesquisas, rastreamento online, análise de marketing online, dados coletados de assinatura e registro (pense em newsletters), monitoramento de mídias sociais, entre outros métodos.

Formalmente, esses dados aparecerão como estruturas diferentes, incluindo - mas não se limitando a - o seguinte:

Big data

O conceito de big data – dados tão grandes, rápidos ou complexos que são difíceis ou impossíveis de processar usando métodos tradicionais – ganhou força no início dos anos 2000. Então, Doug Laney, um analista do setor, articulou o que hoje é conhecido como a definição convencional de big data como os três Vs: volume, velocidade e variedade.

Volume: Como mencionado anteriormente, as organizações estão coletando dados constantemente. Em um passado não muito distante, armazenar seria um problema real, mas hoje em dia o armazenamento é barato e ocupa pouco espaço.

Velocidade: Os dados recebidos precisam ser tratados em tempo hábil. Com o crescimento da Internet das Coisas, isso pode significar que esses dados estão chegando constantemente e em uma velocidade sem precedentes.

Variedade: os dados coletados e armazenados pelas organizações vêm em muitas formas, desde dados estruturados – ou seja, dados numéricos mais tradicionais – até

dados não estruturados – pense em e-mails, vídeos, áudio e assim por diante. Abordaremos dados estruturados e não estruturados um pouco mais adiante.

Metadados

Esta é uma forma de dados que fornece informações sobre outros dados, como uma imagem. Na vida cotidiana, você encontrará isso, por exemplo, clicando com o botão direito do mouse em um arquivo em uma pasta e selecionando "Obter informações", que mostrará informações como tamanho e tipo do arquivo, data de criação e assim por diante.

Dados em tempo real

São dados que são apresentados assim que são adquiridos. Um bom exemplo disso é o ticket do mercado de ações, que fornece informações sobre as ações mais ativas em tempo real.

Dados da máquina

São dados produzidos inteiramente por máquinas, sem instrução humana. Um exemplo disso pode ser os registros de chamadas gerados automaticamente pelo seu smartphone.

Dados quantitativos e qualitativos

Dados quantitativos – também conhecidos como dados estruturados – podem aparecer como um banco de dados "tradicional" – ou seja, com linhas e colunas. Dados qualitativos – também conhecidos como dados não estruturados – são os outros tipos de dados que não se encaixam em linhas e colunas, que podem incluir texto, imagens, vídeos e muito mais. Discutiremos isso melhor na próxima seção.

Qual é a diferença entre dados quantitativos e qualitativos?

A forma como você analisa seus dados depende do tipo de dados com os quais está lidando – quantitativo ou qualitativo. Então qual é a diferença?

Dados quantitativos são qualquer coisa mensurável, incluindo quantidades e números específicos. Alguns exemplos de dados quantitativos incluem números de vendas, taxas de cliques de e-mail, número de visitantes do site e aumento percentual da receita. As

técnicas de análise de dados quantitativos concentram-se na análise estatística, matemática ou numérica de conjuntos de dados (geralmente grandes). Isso inclui a manipulação de dados estatísticos usando técnicas e algoritmos computacionais. As técnicas de análise quantitativa são frequentemente usadas para explicar certos fenômenos ou fazer previsões.

Os dados qualitativos não podem ser medidos objetivamente e, portanto, estão abertos a interpretações mais subjetivas. Alguns exemplos de dados qualitativos incluem comentários deixados em resposta a uma pergunta de pesquisa, coisas que as pessoas disseram durante entrevistas, tweets e outras postagens de mídia social e o texto incluído nas análises de produtos. Com a análise de dados qualitativos, o foco está em dar sentido a dados não estruturados (como texto escrito ou transcrições de conversas faladas). Muitas vezes, a análise qualitativa organiza os dados em temas – um processo que, felizmente, pode ser automatizado.

Técnicas de análise de dados

Agora que estamos familiarizados com alguns dos diferentes tipos de dados, vamos nos concentrar no tópico em questão: diferentes métodos para analisar dados.

a. Análise de regressão

A análise de regressão é usada para estimar a relação entre um conjunto de variáveis. Ao realizar qualquer tipo de análise de regressão, você procura ver se há uma correlação entre uma variável dependente (essa é a variável ou resultado que você deseja medir ou prever) e qualquer número de variáveis independentes (fatores que podem ter impacto na variável dependente). O objetivo da análise de regressão é estimar como uma ou mais variáveis podem impactar a variável dependente, a fim de identificar tendências e padrões. Isso é especialmente útil para fazer previsões e prever tendências futuras.

Vamos imaginar que você trabalha para uma empresa de comércio eletrônico e deseja examinar a relação entre: (a) quanto dinheiro é gasto em marketing de mídia social e (b) receita de vendas. Nesse caso, a receita de vendas é sua variável dependente – é o fator que você está mais interessado em prever e impulsionar. Os gastos com mídia

social são sua variável independente; você quer determinar se tem ou não impacto nas vendas e, em última análise, se vale a pena aumentar, diminuir ou manter o mesmo. Usando a análise de regressão, você poderá ver se há uma relação entre as duas variáveis. Uma correlação positiva implicaria que quanto mais você gasta em marketing de mídia social, mais receita de vendas você obtém. Nenhuma correlação pode sugerir que o marketing de mídia social não tenha influência em suas vendas. Compreender a relação entre essas duas variáveis o ajudaria a tomar decisões informadas sobre o orçamento de mídia social daqui para frente. No entanto: é importante notar que, por si só, as regressões só podem ser usadas para determinar se há ou não uma relação entre um conjunto de variáveis - elas não dizem nada sobre causa e efeito. Portanto, embora uma correlação positiva entre os gastos com mídia social e a receita de vendas possa sugerir que um afeta o outro, é impossível tirar conclusões definitivas com base apenas nessa análise.

Existem muitos tipos diferentes de análise de regressão, e o modelo que você usa depende do tipo de dados que você tem para a variável dependente. Por exemplo, sua variável dependente pode ser contínua (ou seja, algo que pode ser medido em uma escala contínua, como receita de vendas em USD), nesse caso você usaria um tipo diferente de análise de regressão do que se sua variável dependente fosse categórica em natureza (ou seja, compreendendo valores que podem ser categorizados em vários grupos distintos com base em uma determinada característica, como localização do cliente por continente).

Escolhendo o tipo correto de análise de regressão[1]

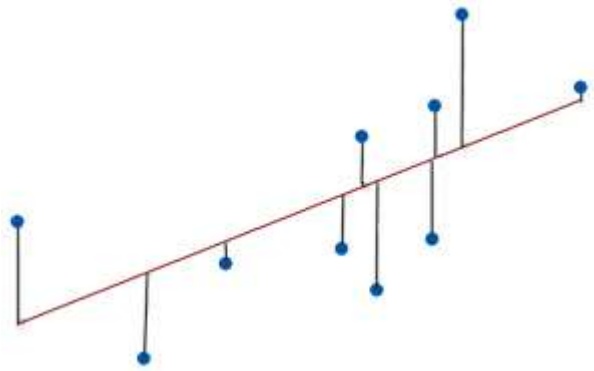
- A análise de regressão descreve matematicamente a relação entre um conjunto de variáveis independentes e uma variável dependente. Existem vários tipos de modelos de regressão que você pode usar. Essa escolha geralmente depende do tipo de dados que você tem para a variável dependente e do tipo de modelo que fornece o melhor ajuste. Neste post, abordo os tipos mais comuns de análises de regressão e como decidir qual é a certa para seus dados.

[1] https://statisticsbyjim.com/regression/choosing-regression-analysis/

- Fornecerei uma visão geral junto com informações para ajudá-lo a escolher. Eu organizo os tipos de regressão pelos diferentes tipos de variável dependente. Se você não tiver certeza de qual procedimento usar, determine qual tipo de variável dependente você possui e, em seguida, concentre-se nessa seção neste post. Este processo deve ajudar a estreitar as escolhas! Abordarei modelos de regressão apropriados para variáveis dependentes que medem dados contínuos, categóricos e de contagem.

Análise de regressão com variáveis dependentes contínuas

- A análise de regressão com uma variável dependente contínua é provavelmente o primeiro tipo que vem à mente. Embora este seja o caso principal, você ainda precisa decidir qual usar.



- OLS produz a linha ajustada que minimiza a soma das diferenças quadradas entre os pontos de dados e a linha.
- Variáveis contínuas são uma medida em uma escala contínua, como peso, tempo e comprimento.

Regressão linear

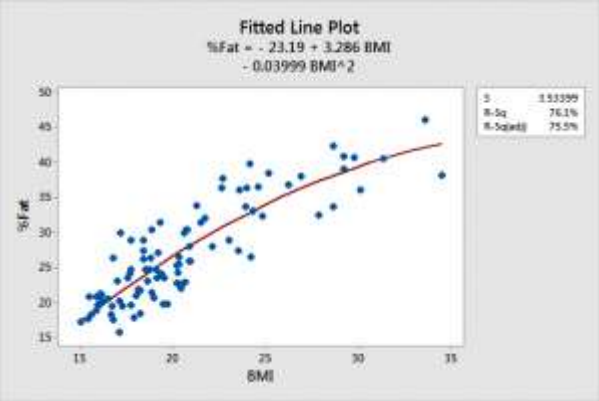
- A regressão linear, também conhecida como mínimos quadrados ordinários (OLS) e mínimos quadrados lineares, é o verdadeiro cavalo de batalha do mundo da regressão. Use a regressão linear para entender a mudança média em uma variável dependente dada uma mudança de uma unidade em cada variável independente. Você também pode usar polinômios para modelar a

curvatura e incluir efeitos de interação. Apesar do termo "modelo linear", esse tipo pode modelar a curvatura.

- Esta análise estima os parâmetros minimizando a soma dos erros ao quadrado (SSE). Os modelos lineares são os mais comuns e fáceis de usar. Se você tem uma variável dependente contínua, a regressão linear é provavelmente o primeiro tipo que você deve considerar.

Existem algumas opções especiais disponíveis para regressão linear.

- Gráfico de linha ajustada para um modelo de regressão linear que se ajusta à relação curva entre o IMC e o percentual de gordura corporal.



Modelo linear que usa um polinômio para modelar a curvatura

- Gráficos de linha ajustada: Se você tiver uma variável independente e a variável dependente, use um gráfico de linha ajustada para exibir os dados junto com a linha de regressão ajustada e a saída de regressão essencial. Esses gráficos tornam a compreensão do modelo mais intuitiva.
- Regressão passo a passo e regressão de melhores subconjuntos: esses métodos automatizados podem ajudar a identificar variáveis candidatas no início do processo de especificação do modelo.

Tipos avançados de regressão linear

- Os modelos lineares são o tipo mais antigo de regressão. Ele foi projetado para que os estatísticos possam fazer os cálculos manualmente. No entanto, o OLS tem vários pontos fracos, incluindo uma sensibilidade a valores discrepantes e

multicolinearidade, e é propenso a overfitting. Para resolver esses problemas, os estatísticos desenvolveram várias variantes avançadas:

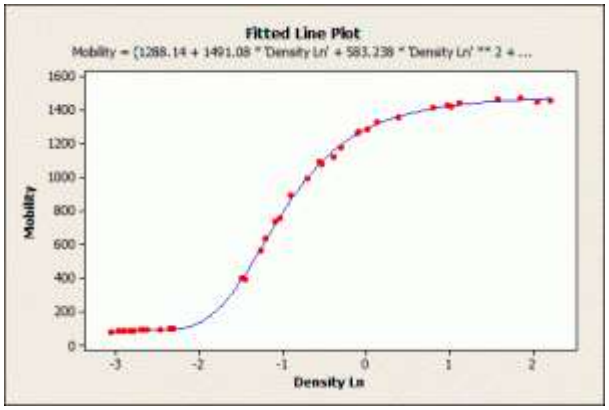
- A regressão de cume permite analisar dados mesmo quando há multicolinearidade severa e ajuda a evitar o overfitting. Esse tipo de modelo reduz a grande e problemática variância que a multicolinearidade causa ao introduzir um leve viés nas estimativas. O procedimento troca grande parte da variância em troca de um pouco de viés, o que produz estimativas de coeficiente mais úteis quando a multicolinearidade está presente.
- A regressão Lasso (operador de seleção e encolhimento mínimo absoluto) realiza a seleção de variáveis que visa aumentar a precisão da previsão, identificando um modelo mais simples. É semelhante à regressão Ridge, mas com seleção de variáveis.
- A regressão de mínimos quadrados parciais (PLS) é útil quando você tem muito poucas observações em comparação com o número de variáveis independentes ou quando suas variáveis independentes são altamente correlacionadas. O PLS diminui as variáveis independentes para um número menor de componentes não correlacionados, semelhante à Análise de Componentes Principais. Em seguida, o procedimento realiza regressão linear nesses componentes em vez dos dados originais. O PLS enfatiza o desenvolvimento de modelos preditivos e não é usado para triagem de variáveis. Ao contrário do OLS, você pode incluir várias variáveis dependentes contínuas. O PLS usa a estrutura de correlação para identificar efeitos menores e modelar padrões multivariados nas variáveis dependentes.

Regressão não linear

- A regressão não linear também requer uma variável dependente contínua, mas oferece maior flexibilidade para ajustar as curvas do que a regressão linear.
- Assim como o OLS, a regressão não linear estima os parâmetros minimizando o SSE. No entanto, os modelos não lineares usam um algoritmo iterativo em vez da abordagem linear de resolvê-los diretamente com equações matriciais. O que isso significa para você é que você precisa se preocupar com qual algoritmo usar, especificando bons valores iniciais e a possibilidade de não

convergir em uma solução ou convergir em um mínimo local em vez de um SSE mínimo global. E isso além de especificar a forma funcional correta!

- A maioria dos modelos não lineares tem uma variável independente contínua, mas é possível ter mais de uma. Quando você tem uma variável independente, pode representar graficamente os resultados usando um gráfico de linha ajustada.



Modelo não linear de mobilidade eletrônica por densidade.

- Meu conselho é ajustar um modelo usando regressão linear primeiro e depois determinar se o modelo linear fornece um ajuste adequado verificando os gráficos de resíduos. Se você não conseguir obter um bom ajuste usando a regressão linear, tente um modelo não linear, pois ele pode se ajustar a uma variedade maior de curvas. Eu sempre recomendo que você experimente o OLS primeiro porque é mais fácil de executar e interpretar.

Análise de regressão com variáveis dependentes categóricas

- Até agora, analisamos modelos que exigem uma variável dependente contínua. Em seguida, vamos passar para variáveis independentes categóricas. Uma variável categórica tem valores que você pode colocar em um número contável de grupos distintos com base em uma característica. A regressão logística transforma a variável dependente e, em seguida, usa a Estimativa de máxima verossimilhança, em vez de mínimos quadrados, para estimar os parâmetros.
- A regressão logística descreve a relação entre um conjunto de variáveis independentes e uma variável dependente categórica. Escolha o tipo de

modelo logístico com base no tipo de variável dependente categórica que você possui.

Regressão Logística Binária

- Use a regressão logística binária para entender como as mudanças nas variáveis independentes estão associadas às mudanças na probabilidade de ocorrência de um evento. Este tipo de modelo requer uma variável dependente binária. Uma variável binária tem apenas dois valores possíveis, como passar e falhar.
- Exemplo: Cientistas políticos avaliam as chances de o atual presidente dos EUA ganhar a reeleição com base no desempenho do mercado de ações.

Regressão Logística Ordinal

- A regressão logística ordinal modela a relação entre um conjunto de preditores e uma variável de resposta ordinal. Uma resposta ordinal tem pelo menos três grupos que têm uma ordem natural, como quente, médio e frio.
- Exemplo: Os analistas de mercado desejam determinar quais variáveis influenciam a decisão de comprar pipoca grande, média ou pequena no cinema.

Regressão Logística Nominal

- A regressão logística nominal, também conhecida como regressão logística multinomial, modela a relação entre um conjunto de variáveis independentes e uma variável dependente nominal. Uma variável nominal possui pelo menos três grupos que não possuem uma ordem natural, como arranhão, amassado e rasgo.
- Exemplo: Um analista de qualidade estuda as variáveis que afetam as chances do tipo de defeito do produto: arranhões, amassados e rasgos.

Análise de regressão com variáveis dependentes de contagem

- Se sua variável dependente for uma contagem de itens, eventos, resultados ou atividades, talvez seja necessário usar um tipo diferente de modelo de regressão. As contagens são inteiras não negativos (0, 1, 2, etc.). Dados de contagem com médias mais altas tendem a ser distribuídos normalmente e você pode usar o OLS com frequência. No entanto, os dados de contagem com médias menores podem ser distorcidos e a regressão linear pode ter dificuldade em ajustar esses dados. Para esses casos, existem vários tipos de modelos que você pode usar.

Regressão de Poisson

- Os dados de contagem frequentemente seguem a distribuição de Poisson, o que torna a Regressão de Poisson uma boa possibilidade. As variáveis de Poisson são uma contagem de algo em uma quantidade constante de tempo, área ou outro período consistente de observação. Com uma variável de Poisson, você pode calcular e avaliar uma taxa de ocorrência. Um exemplo clássico de um conjunto de dados de Poisson é fornecido por Ladislaus Bortkiewicz, um economista russo, que analisou as mortes anuais causadas por coices de cavalo no exército prussiano de 1875-1894.
- Use a regressão de Poisson para modelar como as mudanças nas variáveis independentes estão associadas às mudanças nas contagens. Os modelos de Poisson são semelhantes aos modelos logísticos porque usam a Estimativa de Máxima Verossimilhança e transformam a variável dependente usando o logaritmo natural. Os modelos de Poisson podem ser adequados para dados de taxa, em que a taxa é uma contagem de eventos dividida por uma medida da exposição dessa unidade (uma unidade consistente de observação). Por exemplo, homicídios por mês.
- Exemplo: Um analista usa a regressão de Poisson para modelar o número de chamadas que um call center recebe diariamente.Alternativas à regressão de Poisson para dados de contagem

- Nem todos os dados de contagem seguem a distribuição de Poisson porque essa distribuição tem algumas restrições rigorosas. Felizmente, existem análises alternativas que você pode realizar quando tiver dados de contagem.

Regressão binomial negativa:

- A regressão de Poisson assume que a variância é igual à média. Quando a variância é maior que a média, seu modelo tem superdispersão. Um modelo binomial negativo, também conhecido como NB2, pode ser mais apropriado quando há superdispersão.
- Modelos inflacionados com zeros: seus dados de contagem podem ter muitos zeros para seguir a distribuição de Poisson. Em outras palavras, há mais zeros do que a regressão de Poisson prevê. Modelos inflacionados com zeros assumem que dois processos separados trabalham juntos para produzir os zeros excessivos. Um processo determina se há zero eventos ou mais de zero eventos. O outro é o processo de Poisson que determina quantos eventos ocorrem, alguns dos quais podem ser zero. Um exemplo deixa isso mais claro!

Suponha que os guardas florestais contem o número de peixes capturados por cada visitante do parque quando eles saem do parque. Um modelo inflado com zero pode ser apropriado para este cenário porque existem dois processos para capturar peixe zero:

- Alguns visitantes do parque pegam zero peixe porque não foram pescar.
- Outros visitantes foram pescar, e algumas dessas pessoas não pegaram nenhum peixe.

b. Simulação de Monte Carlo

Ao tomar decisões ou tomar certas ações, há uma série de diferentes resultados possíveis. Se você pegar o ônibus, você pode ficar preso no trânsito. Se você caminhar, poderá ser pego pela chuva ou esbarrar em seu vizinho tagarela, potencialmente atrasando sua jornada. Na vida cotidiana, tendemos a pesar brevemente os prós e os

contras antes de decidir qual ação tomar; no entanto, quando as apostas são altas, é essencial calcular, da forma mais completa e precisa possível, todos os riscos e recompensas potenciais.

A simulação de Monte Carlo, também conhecida como método de Monte Carlo, é uma técnica computadorizada usada para gerar modelos de resultados possíveis e suas distribuições de probabilidade. Essencialmente, considera uma série de resultados possíveis e, em seguida, calcula a probabilidade de que cada resultado específico seja realizado. O método Monte Carlo é usado por analistas de dados para realizar análises de risco avançadas, permitindo que eles prevejam melhor o que pode acontecer no futuro e tomem decisões de acordo.

Então, como funciona a simulação de Monte Carlo e o que ela pode nos dizer? Para executar uma simulação de Monte Carlo, você começará com um modelo matemático de seus dados, como uma planilha. Dentro de sua planilha, você terá uma ou várias saídas de seu interesse; lucro, por exemplo, ou número de vendas. Você também terá várias entradas; essas são variáveis que podem afetar sua variável de saída. Se você estiver analisando o lucro, as entradas relevantes podem incluir o número de vendas, gastos totais com marketing e salários dos funcionários. Se você conhecesse os valores exatos e definitivos de todas as suas variáveis de entrada, seria muito fácil calcular o lucro que teria no final. No entanto, quando esses valores são incertos, uma simulação de Monte Carlo permite calcular todas as opções possíveis e suas probabilidades. Qual será o seu lucro se você fizer 100.000 vendas e contratar cinco novos funcionários com um salário de $ 50.000 cada? Qual é a probabilidade desse resultado? Qual será o seu lucro se você fizer apenas 12.000 vendas e contratar cinco novos funcionários? E assim por diante. Ele faz isso substituindo todos os valores incertos por funções que geram amostras aleatórias de distribuições determinadas por você e, em seguida, executando uma série de cálculos e recálculos para produzir modelos de todos os resultados possíveis e suas distribuições de probabilidade. O método de Monte Carlo é uma das técnicas mais populares para calcular o efeito de variáveis imprevisíveis em uma variável de saída específica, tornando-o ideal para análise de risco.

c. Análise fatorial

A análise fatorial é uma técnica utilizada para reduzir um grande número de variáveis a um número menor de fatores. Ele funciona com base em que várias variáveis observáveis separadas se correlacionam umas com as outras porque estão todas associadas a um construto subjacente. Isso é útil não apenas porque condensa grandes conjuntos de dados em amostras menores e mais gerenciáveis, mas também porque ajuda a descobrir padrões ocultos. Isso permite que você explore conceitos que não podem ser facilmente medidos ou observados, como riqueza, felicidade, condicionamento físico ou, para um exemplo mais relevante para os negócios, fidelidade e satisfação do cliente.

Vamos imaginar que você deseja conhecer melhor seus clientes, então você envia uma pesquisa bastante longa com cem perguntas. Algumas das perguntas estão relacionadas a como eles se sentem em relação à sua empresa e produto; por exemplo, "Você nos recomendaria a um amigo?" e "Como você classificaria a experiência geral do cliente?" Outras perguntas fazem coisas como "Qual é a sua renda familiar anual?" e "Quanto você está disposto a gastar em cuidados com a pele por mês?"

Depois que sua pesquisa for enviada e concluída por muitos clientes, você terá um grande conjunto de dados que basicamente informa cem coisas diferentes sobre cada cliente (supondo que cada cliente dê cem respostas). Em vez de olhar para cada uma dessas respostas (ou variáveis) individualmente, você pode usar a análise fatorial para agrupá-las em fatores que pertencem um ao outro – em outras palavras, relacioná-las a um único construto subjacente. Neste exemplo, a análise fatorial funciona encontrando itens de pesquisa fortemente correlacionados. Isso é conhecido como covariância. Portanto, se houver uma forte correlação positiva entre a renda familiar e o quanto eles estão dispostos a gastar em cuidados com a pele a cada mês (ou seja, à medida que um aumento, o outro também), esses itens podem ser agrupados. Juntamente com outras variáveis (respostas à pesquisa), você pode descobrir que elas podem ser reduzidas a um único fator, como "poder de compra do consumidor". Da mesma forma, se uma classificação de experiência do cliente de 10/10 se correlaciona fortemente com as respostas "sim" em relação à probabilidade de recomendar seu

produto a um amigo, esses itens podem ser reduzidos a um único fator, como "satisfação do cliente".

No final, você tem um número menor de fatores em vez de centenas de variáveis individuais. Esses fatores são então levados adiante para uma análise mais aprofundada, permitindo que você aprenda mais sobre seus clientes (ou qualquer outra área que você esteja interessado em explorar).

d. Análise de coorte

A análise de coorte é definida na Wikipedia da seguinte forma: "A análise de coorte é um subconjunto de análise comportamental que obtém os dados de um determinado conjunto de dados e, em vez de analisar todos os usuários como uma unidade, os divide em grupos relacionados para análise. Esses grupos relacionados, ou coortes, geralmente compartilham características ou experiências comuns dentro de um período de tempo definido".

Então, o que isso significa e por que é útil? Vamos detalhar ainda mais a definição acima. Uma coorte é um grupo de pessoas que compartilham uma característica (ou ação) comum durante um determinado período de tempo. Os alunos que se matricularam na universidade em 2020 podem ser chamados de coorte de 2020. Os clientes que compraram algo em sua loja virtual pelo aplicativo no mês de dezembro também podem ser considerados uma coorte.

Com a análise de coorte, você divide seus clientes ou usuários em grupos e observa como esses grupos se comportam ao longo do tempo. Assim, em vez de olhar para um único e isolado instantâneo de todos os seus clientes em um determinado momento (com cada cliente em um ponto diferente de sua jornada), você está examinando o comportamento de seus clientes no contexto do ciclo de vida do cliente. Como resultado, você pode começar a identificar padrões de comportamento em vários pontos da jornada do cliente – digamos, desde a primeira visita ao seu site, até a inscrição no boletim informativo por e-mail, a primeira compra e assim por diante. Como tal, a análise de coorte é dinâmica, permitindo que você descubra informações valiosas sobre o ciclo de vida do cliente.

Isso é útil porque permite que as empresas personalizem seus serviços para segmentos específicos de clientes (ou coortes). Vamos imaginar que você faça uma campanha de 50% de desconto para atrair novos clientes em potencial para seu site. Depois de atrair um grupo de novos clientes (uma coorte), você desejará acompanhar se eles realmente compram alguma coisa e, se o fizerem, se fazem ou não (e com que frequência) uma compra repetida. Com esses insights, você começará a entender muito melhor quando esse grupo específico pode se beneficiar de outra oferta de desconto ou anúncios de retargeting nas mídias sociais, por exemplo. Em última análise, a análise de coorte permite que as empresas otimizem suas ofertas de serviços (e marketing) para fornecer uma experiência mais direcionada e personalizada.

e. Análise de cluster

A análise de cluster é uma técnica exploratória que busca identificar estruturas dentro de um conjunto de dados. O objetivo da análise de cluster é classificar diferentes pontos de dados em grupos (ou clusters) que são internamente homogêneos e externamente heterogêneos. Isso significa que os pontos de dados em um cluster são semelhantes entre si e diferentes dos pontos de dados em outro cluster. O clustering é usado para obter informações sobre como os dados são distribuídos em um determinado conjunto de dados ou como uma etapa de pré-processamento para outros algoritmos.

Existem muitas aplicações do mundo real de análise de cluster. Em marketing, a análise de cluster é comumente usada para agrupar uma grande base de clientes em segmentos distintos, permitindo uma abordagem mais direcionada à publicidade e comunicação. As seguradoras podem usar a análise de cluster para investigar por que certos locais estão associados a um alto número de sinistros de seguros. Outra aplicação comum é em geologia, onde especialistas usarão análise de cluster para avaliar quais cidades estão em maior risco de terremotos (e, assim, tentar mitigar o risco com medidas de proteção).

É importante observar que, embora a análise de cluster possa revelar estruturas em seus dados, ela não explica por que essas estruturas existem. Com isso em mente, a análise de cluster é um ponto de partida útil para entender seus dados e informar análises adicionais.

f. Análise de séries temporais

A análise de séries temporais é uma técnica estatística usada para identificar tendências e ciclos ao longo do tempo. Os dados de séries temporais são uma sequência de pontos de dados que medem a mesma variável em diferentes momentos (por exemplo, números de vendas semanais ou inscrições mensais de e-mail). Ao observar as tendências relacionadas ao tempo, os analistas podem prever como a variável de interesse pode flutuar no futuro.

Ao realizar análises de séries temporais, os principais padrões que você procurará em seus dados são:

Tendências: aumentos ou diminuições estáveis e lineares durante um período de tempo prolongado.

Sazonalidade: flutuações previsíveis nos dados devido a fatores sazonais em um curto período de tempo. Por exemplo, você pode ver um pico nas vendas de roupas de banho no verão na mesma época todos os anos.

Padrões cíclicos: ciclos imprevisíveis onde os dados flutuam. As tendências cíclicas não se devem à sazonalidade, mas podem ocorrer como resultado de condições econômicas ou relacionadas ao setor.

Como você pode imaginar, a capacidade de fazer previsões informadas sobre o futuro tem imenso valor para os negócios. A análise e a previsão de séries temporais são usadas em vários setores, mais comumente para análise de mercado de ações, previsão econômica e previsão de vendas. Existem diferentes tipos de modelos de séries temporais, dependendo dos dados que você está usando e dos resultados que deseja prever. Esses modelos são normalmente classificados em três tipos amplos: os modelos autorregressivos (AR), os modelos integrados (I) e os modelos de média móvel (MA).

g. Análise de sentido

Quando você pensa em dados, sua mente provavelmente vai automaticamente para números e planilhas. Muitas empresas ignoram o valor dos dados qualitativos, mas, na realidade, há insights incalculáveis a serem obtidos com o que as pessoas

(especialmente os clientes) escrevem e dizem sobre você. Então, como você analisa dados textuais?

Uma técnica qualitativa altamente útil é a análise de sentimento, uma técnica que pertence à categoria mais ampla de análise de texto – o processo (geralmente automatizado) de classificação e compreensão de dados textuais. Com a análise de sentidos, o objetivo é interpretar e classificar as emoções transmitidas nos dados textuais. De uma perspectiva de negócios, isso permite verificar como seus clientes se sentem sobre vários aspectos de sua marca, produto ou serviço. Existem vários tipos diferentes de modelos de análise de sentimentos, cada um com um foco ligeiramente diferente. Os três tipos principais incluem:

Análise de sentido refinada: se você deseja se concentrar na polaridade da opinião (ou seja, positiva, neutra ou negativa) em profundidade, a análise de sentimento refinada permitirá que você faça isso. Por exemplo, se você quiser interpretar as classificações por estrelas dadas pelos clientes, poderá usar uma análise de sentimento refinada para categorizar as várias classificações em uma escala que varia de muito positivo a muito negativo.

Detecção de emoções: esse modelo geralmente usa algoritmos complexos de aprendizado de máquina para selecionar várias emoções de seus dados textuais. Você pode usar um modelo de detecção de emoções para identificar palavras associadas a felicidade, raiva, frustração e entusiasmo, fornecendo informações sobre como seus clientes se sentem ao escrever sobre você ou seu produto em, digamos, um site de avaliação de produtos.

Análise de sentimentos baseada em aspectos: esse tipo de análise permite identificar a quais aspectos específicos as emoções ou opiniões se relacionam, como um determinado recurso do produto ou uma nova campanha publicitária. Se um cliente escreve que "achou o novo anúncio do Instagram tão irritante", seu modelo deve detectar não apenas um sentimento negativo, mas também o objeto para o qual ele é direcionado.

Em poucas palavras, a análise de sentimentos usa vários sistemas e algoritmos de *Processamento de Linguagem Natural* (NLP) que são treinados para associar certas entradas (por exemplo, certas palavras) a determinadas saídas. Por exemplo, a entrada "irritante" seria reconhecida e marcada como "negativa". A análise de sentimentos é crucial para entender como seus clientes se sentem sobre você e seus produtos, para identificar áreas de melhoria e até mesmo para evitar desastres de relações públicas em tempo real!

O processo de análise de dados

Para obter insights significativos dos dados, os analistas de dados realizarão um rigoroso processo passo a passo. Examinamos isso em detalhes em nosso guia passo a passo para o processo de análise de dados - mas, para resumir brevemente, o processo de análise de dados geralmente consiste nas seguintes fases:

Definindo a pergunta

O primeiro passo para qualquer analista de dados será definir o objetivo da análise, às vezes chamado de "declaração do problema". Essencialmente, você está fazendo uma pergunta em relação a um problema de negócios que está tentando resolver. Depois de definir isso, você precisará determinar quais fontes de dados o ajudarão a responder a essa pergunta.

Coletando os dados

Agora que você definiu seu objetivo, o próximo passo será definir uma estratégia para coletar e agregar os dados apropriados. Você usará dados quantitativos (numéricos) ou qualitativos (descritivos)? Esses dados se encaixam em dados primários, secundários ou de terceiros?

Limpando os dados

Infelizmente, seus dados coletados não estão automaticamente prontos para análise – você terá que limpá-los primeiro. Como analista de dados, essa fase do processo será a

que mais demorará. Durante o processo de limpeza de dados, você provavelmente será:

- Remoção de erros importantes, duplicatas e discrepâncias
- Removendo pontos de dados indesejados
- Estruturar os dados, ou seja, corrigir erros de digitação, problemas de layout etc.
- Preenchendo as principais lacunas nos dados

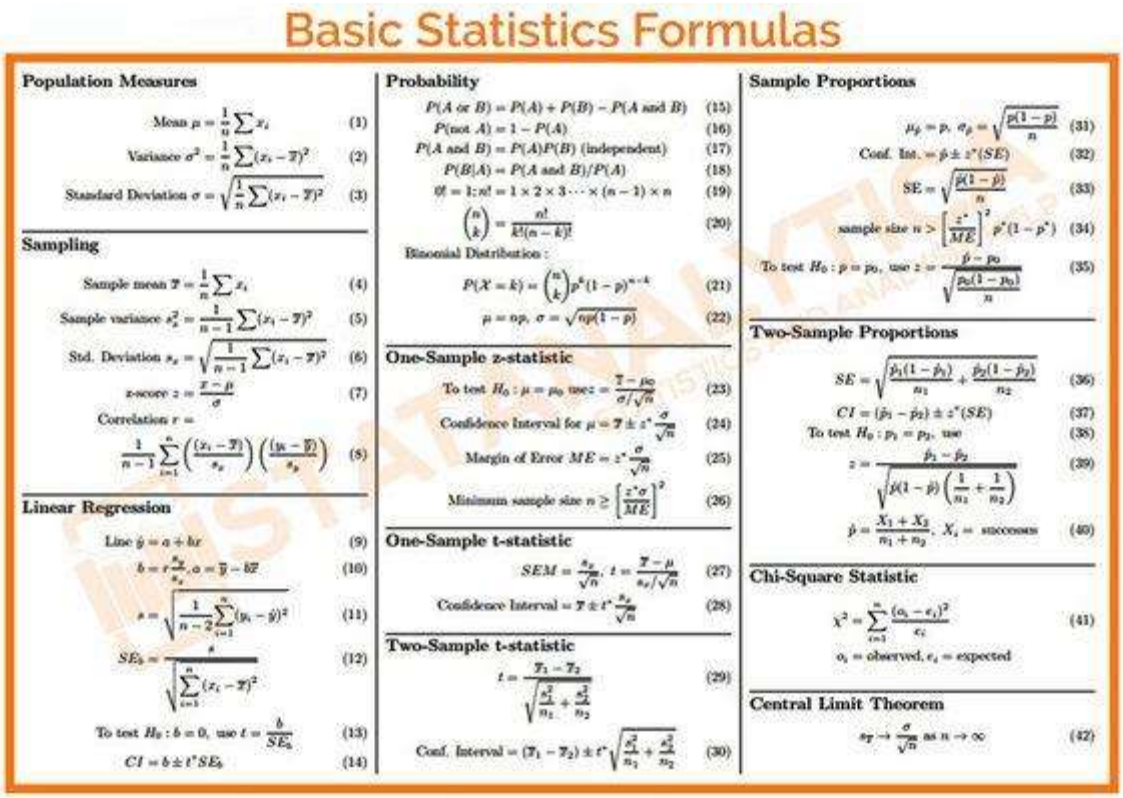
Limpar conjuntos de dados manualmente, especialmente os grandes, pode ser assustador. Felizmente, existem muitas ferramentas disponíveis para agilizar o processo. Ferramentas de código aberto, como o OpenRefine[2], são excelentes para limpeza básica de dados, bem como para exploração de alto nível. No entanto, as ferramentas gratuitas oferecem funcionalidade limitada para conjuntos de dados muito grandes. Bibliotecas Python (por exemplo, Pandas) e alguns pacotes R são mais adequados para limpeza pesada de dados. É claro que você precisará estar familiarizado com os idiomas. Alternativamente, ferramentas empresariais também estão disponíveis. Por exemplo, Data Ladder[3], que é uma das ferramentas de correspondência de dados mais bem avaliadas do setor.

Analisando os dados

Agora que terminamos de limpar os dados, é hora de analisá-los! Muitos métodos de análise já foram descritos neste artigo, e cabe a você decidir qual deles se adequa melhor ao objetivo atribuído. Pode se enquadrar em uma das seguintes categorias:

- Análise descritiva, que identifica o que já aconteceu
- Análise de diagnóstico, que se concentra em entender por que algo aconteceu
- Análise preditiva, que identifica tendências futuras com base em dados históricos
- Análise prescritiva, que permite fazer recomendações para o futuro

[2] https://openrefine.org/
[3] https://dataladder.com/



Visualizando e compartilhando suas descobertas

Análises foram feitas, insights foram colhidos – tudo o que resta a ser feito é compartilhar essas informações com outras pessoas. Isso geralmente é feito com uma ferramenta de visualização de dados.

Existem toneladas de ferramentas de visualização de dados disponíveis, adequadas para diferentes níveis de experiência. Ferramentas populares que exigem pouca ou nenhuma habilidade de codificação incluem Google Charts[4], Tableau[5], Datawrapper[6] e Infogram[7]. Se você estiver familiarizado com Python e R, também existem muitas bibliotecas e pacotes de visualização de dados disponíveis. Por exemplo, confira as bibliotecas Python Plotly[8], Seaborn[9] e Matplotlib[10]. Independentemente das ferramentas de visualização de dados que você usa, certifique-se de aprimorar suas

[4] https://developers.google.com/chart
[5] https://public.tableau.com/en-us/s/
[6] https://www.datawrapper.de/
[7] https://infogram.com/
[8] https://plotly.com/
[9] https://seaborn.pydata.org/
[10] https://matplotlib.org/

habilidades de apresentação também. Lembre-se: a visualização é ótima, mas a comunicação é fundamental!

Apresentação dos Dados

- Descritiva
    - fica em condições de dizer o que de fato acontece, tomando como referência dados reais
- Prescritiva
    - indicada para desenhar simulações e prever comportamentos
- Preditiva
    - busca criar um padrão que explique um dado fenômeno e ajude a prever efeitos
- Diagnóstica
    - detectar as causas de um certo fenômeno ou comportamento

As melhores ferramentas para análise de dados

Como você pode imaginar, cada fase do processo de análise de dados exige que o analista de dados tenha uma variedade de ferramentas que auxiliam na obtenção de informações valiosas dos dados. Essas ferramentas com mais detalhes neste artigo[11], mas, em resumo, aqui está nossa lista dos melhores, com links para cada produto:

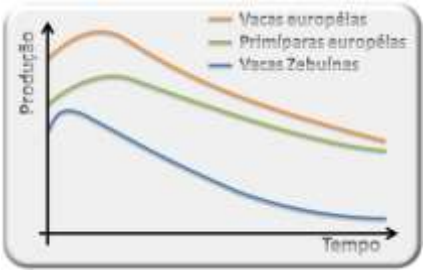
As principais ferramentas para analistas de dados

Apache Spark[12]
KNIME[13]
Microsoft Excel[14]
Microsoft PowerBI[15]

[11] https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/
[12] https://spark.apache.org/
[13] https://www.knime.com/
[14] https://www.microsoft.com/en-us/microsoft-365/excel
[15] https://powerbi.microsoft.com/en-us/what-is-power-bi/

Minitab[16]
Notebook Jupyter[17]
Python[18]
R[19]
SAS[20]
SPSS[21]
Tableau[22]

Entender o comportamento da característica



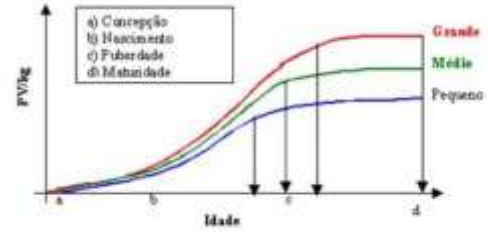
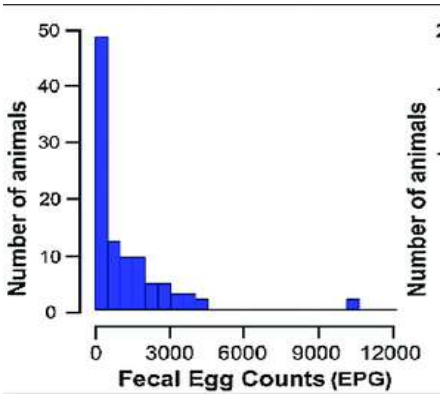
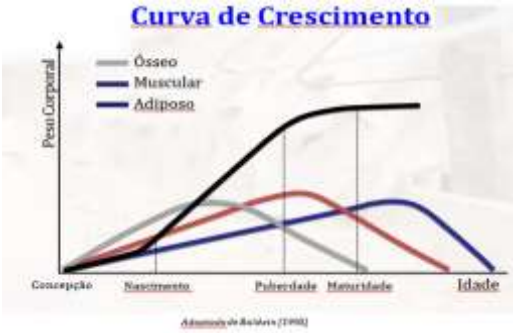


Figura 1: Curva de crescimento de bovinos com diferentes tamanhos corporais





Principais conclusões e leitura adicional

Como você pode ver, existem muitas técnicas diferentes de análise de dados à sua disposição. Para transformar seus dados brutos em insights acionáveis, é importante considerar que tipo de dados você possui (são qualitativos ou quantitativos?) Neste

[16] https://www.minitab.com
[17] https://jupyter.org/try
[18] https://www.python.org/
[19] https://www.r-project.org/about.html
[20] https://www.sas.com/en_us/home.html
[21] https://www.ibm.com/products/spss-statistics
[22] https://www.tableau.com/products

post, apresentamos sete das técnicas de análise de dados mais úteis, mas há muitas outras para serem descobertas!

População, Amostra, Parâmetros e estatísticas

População: o agregado de todas as unidades de amostra definidas arbitrariamente

Parâmetros: constantes que descrevem a população como um todo

Amostras:

- uma agregação de unidades de amostra
- Qualquer subconjunto (ou coleção) de unidades de uma população de unidades
- Podem ser as próprias unidades (por exemplo, um punhado de peças Reeses de um grande frasco de peças)
- ou mais frequentemente uma medida das unidades (por exemplo, uma lista de alturas de 10 árvores ocorrendo em um povoamento de 1200 árvores).
- As amostras que são selecionadas aleatoriamente, ou semelhantes ao aleatório, podem ser utilizadas para análise estatística
- Medir todas as unidades (árvores, recreação, pássaros, etc.) é impraticável, senão impossível.
- Amostrar apenas algumas unidades economiza dinheiro.
- Amostrar apenas algumas unidades economiza tempo.
- Algumas medições são destrutivas:
    - corte de árvores para inspecionar padrões de anéis ou análise de caule, captura de vida selvagem para examinar sua morfologia, etc.
- A amostragem torna os métodos estatísticos atraentes e poderosos.

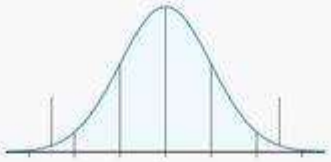
Variáveis: uma característica que pode variar de uma amostra até a próxima

Estatística: parâmetro de uma amostra (distribuição de amostra)

TOP 10

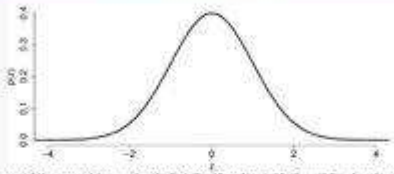
Types of Distribution in Statistics With Formulas

Normal Distribution



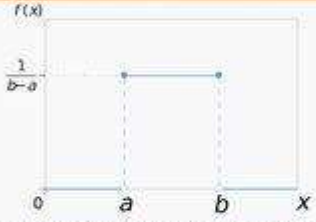
Sometimes, the normal distribution is also called the bell curve. It occurs naturally in several cases; for example, the normal distribution can be seen in tests such as GRE and SAT. Furthermore, there are several groups that follow the normal distribution pattern.

T- Distribution



It is one of the most important distribution in statistics. It is also known as Student's t- distribution, which is the probability distribution. That is used to estimate the parameters of the population when the given sample size is small.

Uniform Distribution



The basic form of a continuous distribution is known as uniform distribution. It has the constant probability that forms a rectangular distribution. And it implies that each value has the same length of distribution.

Characteristics of uniform distribution

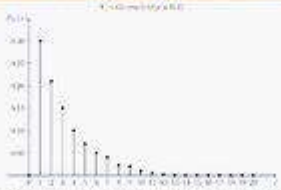
The density function combines to unity.
Each of its input function has equal weightage.
The uniform function's mean is given by:

$$\mu = \frac{(a+b)}{2}$$

The variance of the uniform distribution is given by:

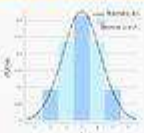
$$V(x) = \frac{(b-a)^2}{12}$$

Bernoulli distribution



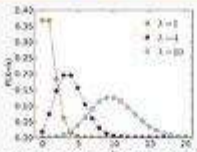
The basic form of a continuous distribution is known as uniform distribution. It has the constant probability that forms a rectangular distribution. And it implies that each value has the same length of distribution.

Binomial distribution



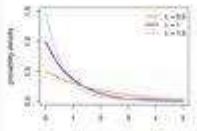
It is a probability distribution that concludes the value that takes one of two independent values under a set of assumptions or parameters. Besides, the binomial distribution's assumptions must have a single result with the same probability of success.

Poisson distribution



It is a tool that is used to predict a certain probability of the event when you know the value of happening of a certain event.

Exponential distribution



It is also known as a negative exponential distribution that represents the time between the trails in a Poisson process.

Some of the formulas of it

An exponential random variable's expected value is given by:

$$E(X) = \frac{1}{\lambda}$$

An exponential random variable's variance value is given by:

$$Var(X) = \frac{1}{\lambda^2}$$

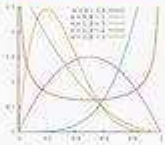
The exponential random variable's moment generating function is given by:

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

An exponential random variable's characteristic function is given by:

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

Beta distribution



It is the family of continuous probability distributions that set under the interval [0,1], which is expressed by alpha and beta.

Properties of beta distribution

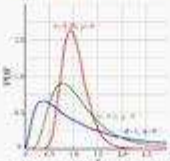
The terms to measure the central tendency are:

Mean
Harmonic Mean
Mode
Median
Geometric Mean

Beta-binomial distribution

It is the simplest Bayesian model that is widely used in intelligence testing, epidemiology, and marketing. A distribution is said to be beta-binomial.

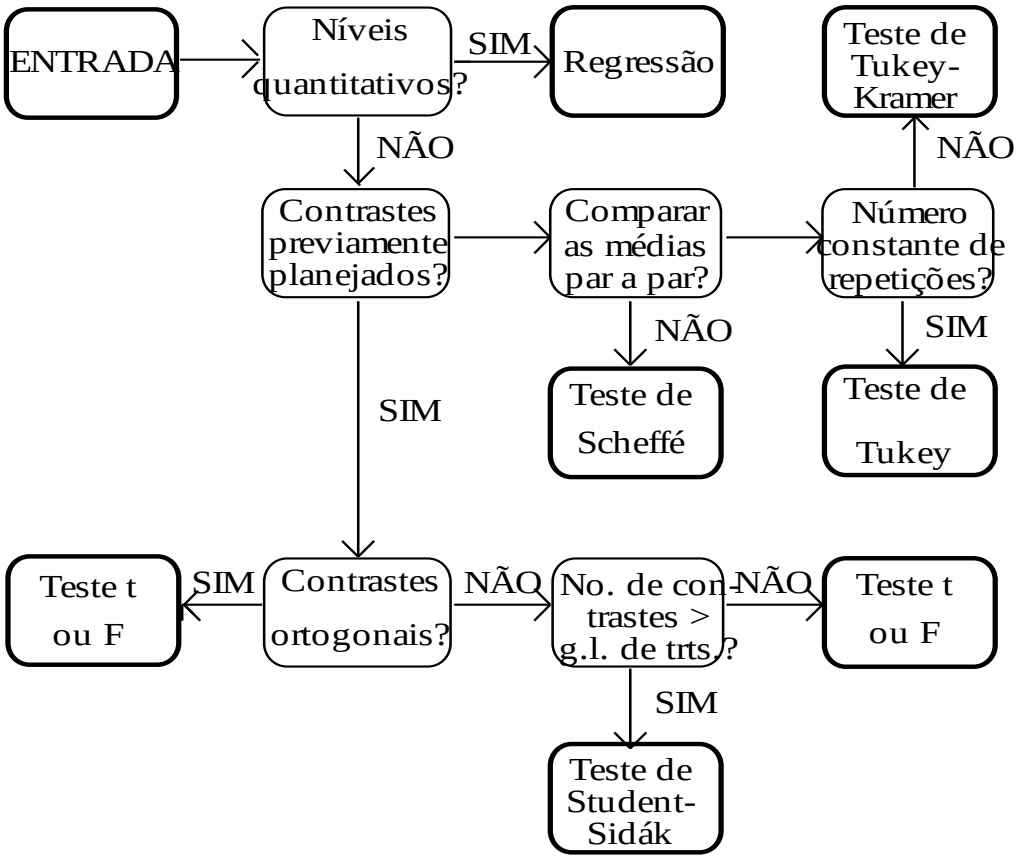
Log-normal distribution



f the log to the power is normally distributed, then the variable is taken as lognormally distributed. Or we can say that ln(x) is normally distributed and that the variable x is assumed to have a log-normal distribution.

Um roteiro para escolher o teste adequado

As recomendações discutidas neste capítulo e no precedente serão agora sumarizadas por via de um roteiro, que facilitará a escolha do teste apropriado:



Vários outros testes existem, alguns com finalidades bem específicas.