



ARGUMENTOS

# UNIDADE

# 3

José Nogueira

Livro Eletrônico

## Aula 5 – Aplicações: Teoria dos Conjuntos, Estatística e Probabilidade

### Objetivo da Aula

Relacionar os conceitos de Teoria dos Conjuntos, estatística e probabilidade, apresentando o relacionamento entre as disciplinas e a área de computação.

### Apresentação

Nesta aula, iremos relacionar conceitos de Teoria dos Conjuntos, estatística e probabilidade, apresentando o relacionamento entre as disciplinas e a área de computação. Devemos destacar que a Teoria dos Conjuntos é o alicerce da probabilidade e da estatística. Podemos destacar a ampla utilização da probabilidade em protocolos de rede de computadores; já na área de estatística, podemos destacar a fundamentação e as ferramentas para analisar estrutura de dados e fornecer a compreensão da melhor solução. Já a Teoria dos Conjuntos é a base, por exemplo, para a programação, bem como para a área de banco de dados e linguagens formais.

### 1. Interdisciplinas

A Teoria dos Conjuntos, com seus diagramas de Venn, regras e operações, permite-nos estudar as relações entre os elementos de um conjunto e entre os conjuntos em si. Esse campo de conhecimento nos fornece uma linguagem e ferramentas que nos permitem descrever e manipular coleções de objetos com precisão e clareza.

Isso é ótimo para ser utilizado em estatística, que, por sua vez, é uma área metodológica do raciocínio lógico que lida com coleta, organização, análise, interpretação e apresentação de dados. Os conjuntos e toda a teoria formada em cima deles auxiliam esse processo, ao entregar em nossas mãos as ferramentas corretas para compreender os conjuntos que são de interesse para o estudo estatístico.

A estatística é ótima para nos ajudar a tomar decisões com base em evidências e, até mesmo, prever certas tendências e modas, porém ela não é capaz de quantificar a certeza que temos sobre os resultados incertos. Nesse ponto, entra a probabilidade, um ramo de estatística que é responsável pelas áreas mais incertas desse campo de conhecimento.

Probabilidade é a média de quão provável é um evento acontecer. Ela é usada para quantificar a incerteza, em situações em que há uma variedade de possíveis resultados, e é uma ferramenta importante para a análise de dados e a tomada de decisões em muitas áreas. A probabilidade continua dependendo da coleta e da análise de dados fornecidas pela estatística em geral, assim como das ferramentas e operações dispostas pela Teoria dos Conjuntos, para que ela seja capaz de ser executada.

Nessa rápida introdução, podemos observar, facilmente, como as três áreas estão intimamente conectadas e relacionadas, já que a estatística e a probabilidade são usadas juntas com frequência para analisar dados e inferir conclusões sobre populações. A Teoria dos Conjuntos também é frequentemente utilizada em problemas de probabilidade, especialmente quando se trata de calcular a probabilidade de um evento que pode ser descrito em termos de conjuntos de resultados possíveis.

## 2. Teoria dos Conjuntos

Como dito acima, a Teoria dos Conjuntos é muito utilizada para resolver problemas de estatística e de probabilidade. Seus conceitos básicos permitem uma melhor compreensão dos conjuntos abstratos formados nessas outras disciplinas.

Destaco aqui:

- Conjunto.
  - Agrupamento de elementos que possui uma característica em comum.
- Elementos.
  - Os indivíduos no grupo.
- Operações.
  - União.
    - Um novo conjunto que une todos os elementos de seu progenitor.
  - Interseção.
    - Um novo conjunto que possui todos os elementos em comum de dois ou mais conjuntos progenitores.
  - Complementar.
    - O complementar de um conjunto em relação a outro conjunto é o conjunto que contém todos os elementos que não estão presentes no primeiro conjunto, mas que estão presentes no segundo conjunto.

Os diagramas de Venn são mais uma ferramenta extremamente utilizada na Teoria dos Conjuntos, para que fique ainda mais fácil a compreensão desses conjuntos e de seus conceitos básicos e operações. O fato é que a representação gráfica facilita muito o entendimento de sistemas de qualquer nível, inclusive complexos.

### 3. Estatística

Os conceitos da Teoria dos Conjuntos são usados na estatística descritiva para coletar, organizar e analisar dados. A estatística descritiva é a parte da estatística que lida com a descrição e o resumo de dados. Ela envolve a organização dos dados em tabelas, gráficos e outras formas e a identificação de padrões e tendências nos dados.

Os conceitos da Teoria dos Conjuntos são especialmente úteis na estatística descritiva para descrever conjuntos de dados. Por exemplo, podemos usar a união de dois conjuntos de dados para combinar dados de diferentes fontes, podemos usar a interseção de dois conjuntos para identificar elementos em comum entre dois conjuntos de dados, assim como podemos usar o complementar de um conjunto para identificar os elementos que não estão presentes em um conjunto de dados.

Além disso, a Teoria dos Conjuntos também é usada na estatística descritiva para calcular medidas de tendência central, como a média, mediana e moda. A média é a soma de todos os valores em um conjunto de dados, dividida pelo número de valores no conjunto. Já a mediana é o valor que divide o conjunto de dados em duas partes iguais, com metade dos valores acima e metade dos valores abaixo desse valor. Por fim, a moda é o valor mais comum em um conjunto de dados.

Por exemplo, suponha que temos os seguintes dados de notas em uma prova:

$$S = \{5, 6, 7, 7, 8, 8, 9\}$$

Daí, podemos calcular a média desse conjunto, que seria a soma de todos os elementos dividida pela sua amplitude.

$$\text{Média} = (5 + 6 + 7 + 7 + 8 + 8 + 9)/7 = 7,14$$

Podemos definir a mediana, que é o elemento mais central do conjunto ordenado em ordem crescente ou a média dos dois elementos mais centrais. Nesse caso, o elemento que divide o conjunto ao meio, em dois conjuntos de amplitude iguais, é o quarto elemento, com valor 7. Por fim, podemos também afirmar que a moda é 7 e 8, pois são os dois elementos que mais se repetem nesse conjunto.

## 4. Probabilidade

A teoria da probabilidade é um ramo da matemática que lida com a análise de eventos aleatórios. Esses eventos são aqueles que não podem ser previstos com certeza. Eles podem ser quantificados por meio da probabilidade, que é uma medida da chance de que o evento ocorra.

Um exemplo de experimento aleatório é o lançamento de um dado. O resultado do lançamento é imprevisível e pode ser qualquer um dos seis valores possíveis (1, 2, 3, 4, 5 ou 6). O conjunto de todos os resultados possíveis desse experimento é chamado de “espaço amostral”. No caso do lançamento de um dado, o espaço amostral é  $\{1, 2, 3, 4, 5, 6\}$ .

Os eventos são subconjuntos do espaço amostral que correspondem a um conjunto de resultados possíveis. Por exemplo, o evento “obter um número par no lançamento de um dado” é o conjunto  $\{2, 4, 6\}$ , que é um subconjunto do espaço amostral.

A teoria da probabilidade fornece ferramentas matemáticas para quantificar a chance de um evento ocorrer. A probabilidade é uma medida numérica entre 0 e 1, que indica a chance de um evento ocorrer. Um evento impossível tem probabilidade 0, e um evento certo tem probabilidade 1. A probabilidade de um evento pode ser calculada dividindo o número de resultados favoráveis pelo número total de resultados possíveis.

Por exemplo, a probabilidade de obter um número par em um lançamento de dado é de  $3/6$  ou  $1/2$ , já que existem três resultados favoráveis (2, 4 ou 6) e seis resultados possíveis.

Em resumo, a teoria da probabilidade é um ramo da matemática que estuda eventos aleatórios. Ela envolve a análise de experimentos aleatórios, espaço amostral e eventos. A probabilidade é uma medida numérica que indica a chance de um evento ocorrer e pode ser calculada dividindo o número de resultados favoráveis pelo número total de resultados possíveis.

## 5. Aplicações

Como podemos perceber, as três áreas são tão entrelaçadas que é necessário o conhecimento de uma para ser capaz de aplicar a outra. Vamos, agora, observar alguns casos em que essas áreas de conhecimento podem ser aplicadas.

Vamos começar com o caso de colisões em uma rede de computação. Suponha uma rede de **ethernet** (rede local), em que há diversas máquinas conectadas a um mesmo sistema de transmissão. Se duas ou mais máquinas tentarem transmitir usando o mesmo protocolo, como **ethernet**, ocorrerá uma colisão. Essa colisão pode fazer com que os pacotes de dados

se sobreponham e se corrompam, o que pode levar a erros de transmissão e atrasos na entrega de dados.

Para evitar que essas colisões ocorram, é comum utilizar um protocolo de detecção de colisão, como CSMA/CD. O CSMA utiliza conceitos da Teoria dos Conjuntos, probabilidade e estatística para determinar quando os dispositivos podem transmitir dados, em uma rede compartilhada, tentando evitar colisões de pacotes de dados.

A Teoria dos Conjuntos é usada para modelar/classificar o comportamento dos dispositivos em uma rede compartilhada. O conjunto dos dispositivos que tentam acessar a rede, em determinado momento, é avaliado para determinar se a transmissão é possível ou não. Para isso, é utilizado um conjunto de regras que garante que apenas um dispositivo transmita dados em dado momento.

A probabilidade entra nessa equação para determinar a probabilidade de colisões em uma rede compartilhada. Quando dois ou mais dispositivos tentam transmitir dados simultaneamente, há uma chance de que seus pacotes de dados colidam e sejam corrompidos. A probabilidade de colisão é avaliada para determinar se a rede está sobrecarregada e se é necessário aguardar antes de transmitir dados.

Já a estatística é usada, nesse caso, para analisar o tráfego de rede e determinar se há congestionamento ou problemas de desempenho na rede. Por exemplo, é possível usar técnicas estatísticas para monitorar o número de colisões na rede, o tamanho dos pacotes de dados e a taxa de transmissão. Essas informações podem ser usadas para ajustar as configurações da rede e melhorar o desempenho geral.

A gerência de redes de computadores é o processo de monitorar, gerenciar e controlar redes de computadores, para garantir que elas funcionem corretamente e atendam às necessidades dos usuários e da organização. Isso envolve o monitoramento de dispositivos de rede, como roteadores, **switches**, servidores, dispositivos de armazenamento e outros componentes, para garantir que eles estejam funcionando adequadamente e que as conexões de rede estejam estabelecidas de forma confiável. A Teoria dos Conjuntos, a estatística e a probabilidade são usadas em várias áreas da gerência de redes de computadores.

A Teoria dos Conjuntos é utilizada na gerência de redes de computadores para modelar e analisar conjuntos de dispositivos de rede e recursos. Por exemplo, um administrador de rede pode usar a Teoria dos Conjuntos para criar grupos de dispositivos de rede, com base em características comuns, como a localização física, a função do dispositivo, a configuração ou outras características relevantes. Isso pode ajudar a simplificar o gerenciamento da rede e a identificar problemas com mais facilidade.

A estatística é usada na gerência de redes de computadores para coletar e analisar dados de desempenho da rede. Por exemplo, um administrador de rede pode coletar informações sobre a largura de banda da rede, a taxa de transmissão, o tempo de resposta, entre outras métricas de desempenho. Esses dados podem ser analisados estatisticamente para identificar problemas de desempenho, identificar gargalos na rede e ajustar a configuração da rede para melhorar o desempenho geral.

A probabilidade é usada na gerência de redes de computadores para prever a probabilidade de falhas e outros eventos imprevistos na rede. Por exemplo, um administrador de rede pode usá-la para prever a probabilidade de um roteador falhar ou de um cabo de rede ser danificado. Isso pode ajudar a planejar o gerenciamento da rede com antecedência, realizar manutenção preventiva e preparar planos de contingência para lidar com problemas de rede imprevistos.

Criptografia é uma área da computação que se preocupa com a proteção da informação, por meio de técnicas de codificação e decodificação. É usada para garantir a confidencialidade, integridade e autenticidade dos dados em trânsito ou armazenados.

A Teoria dos Conjuntos é utilizada para modelar os problemas de criptografia; a probabilidade, para avaliar a segurança dos algoritmos criptográficos. Já a estatística é usada para analisar e testar os algoritmos criptográficos, a fim de determinar sua força e segurança perante possíveis ataques.

Exemplificando, a análise de frequência de caracteres é uma técnica estatística usada para quebrar sistemas de criptografia baseados em substituição simples. A Teoria dos Conjuntos é utilizada para modelar problemas criptográficos, como o problema da fatoração de números inteiros grandes, que serve de base da criptografia de chave pública. Já a probabilidade é constantemente usada para avaliar a segurança dos algoritmos criptográficos e determinar a probabilidade de um ataque bem-sucedido.

Para implementar um banco de dados de sistema de informação bem-sucedido, é necessário, primeiro, modelar ele logicamente. Peter Chen desenvolveu uma técnica chamada de “Modelo de Entidade e Relacionamento” (MER). Esse modelo descreve as entidades, os atributos e os relacionamentos entre essas entidades de maneira clara e estruturada. Novamente, vemos essas três áreas do conhecimento atuando – Teoria dos Conjuntos, estatística e probabilidade – ao mesmo tempo.

O MER utiliza conceitos da Teoria dos Conjuntos para representar as entidades e os seus relacionamentos. Por exemplo, as entidades são representadas como conjuntos de objetos que compartilham as mesmas características. Os relacionamentos são representados como conjuntos de pares ordenados que estabelecem a ligação entre as entidades.

Já a estatística é utilizada na análise e testagem do MER, a fim de garantir que ele atenda às necessidades do sistema de informação. A exemplo, é feita a análise estatística para avaliar a distribuição dos dados, identificar padrões ou anomalias nos dados e, então, determinar a relevância dos atributos para a solução do problema em questão.

A probabilidade também pode ser aplicada no MER, de acordo com a necessidade de avaliar a incerteza dos dados e a probabilidade de ocorrência de eventos específicos. O MER pode ser usado para prever a probabilidade de ocorrência de determinado evento, como o número de vendas em um período ou a probabilidade de um cliente fazer uma compra adicional, por exemplo. Ambos os exemplos incluem técnicas de probabilidade aplicadas em cima do banco de dados modelados pelo sistema de entidade e relacionamento explicitado por Peter Chen.

## Considerações Finais da Aula

Nesta aula, abordamos a interdisciplinaridade entre teoria de conjunto, estatística e probabilidade. Vimos que as três áreas estão fortemente conectadas e, juntas, apoiam o processo de tomada de decisão. Vale destacar um caso muito utilizado, na computação, na área de **Business Intelligence** (BI), de que pais que iam comprar fralda à noite sempre compravam cerveja. Ao identificar esse padrão, os produtos (fralda e cerveja) foram colocados próximos, o que ocasionou o aumento das vendas de cerveja. Com base na análise estatística dos dados e na análise da probabilidade, podemos criar modelos que podem indicar tendências e prever a probabilidade de o evento ocorrer.

Por fim, abordamos, também, a aplicação das três áreas em alguns exemplos da área de computação, como rede de computadores, banco de dados ou mesmo Diagrama de Entidade e Relacionamento (DER), proposto por Peter Chen.

## Materiais Complementares



### O que é Business Intelligence?

2023, Microsoft. Power BI.

Ferramenta que se utiliza da estatística e probabilidade para apresentar *insights* para a tomada de decisões estratégicas. A ferramenta analisa dados históricos e atuais e apresenta as descobertas em formatos gráficos.

Link para acesso: <https://powerbi.microsoft.com/pt-br/what-is-business-intelligence/>  
(acesso em 2 jun. 2023.)

 Artigo**Métodos Estatísticos no Contexto de Big Data**

2021, Carolina Hilda Schleger, Benjamim Zucolotto e Rafael Zancan Frantz.

Material apresenta a importância da estatística para a área de Big Data.

Link para acesso: <https://www.publicacoeseventos.unijui.edu.br/index.php/salaocoonhecimento/article/view/20747/19459> (acesso em 2 jun. 2023.)

 Artigo**Metodologia de Base para Ciência de Dados**

2015, John B. Rollins. IBM.

Material apresenta a utilização da estatística na área de Ciência de Dados.

Link para acesso: <https://www.ibm.com/downloads/cas/B1WQ0GM2> (acesso em 2 jun. 2023.)

**Referências**

GUPTA, Bhisham; GUTTMAN, Irwin. Estatística e probabilidade com aplicações para engenheiros e cientistas. Rio de Janeiro: LTC, 2016. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788521632931>. Acesso em: 9 abr. 2023.