

## Trabalho 2

# Análise de Dados

---

### Contexto

O Conjunto de Dados de Admissões de Pós-Graduação de Berkeley em 1973 representa um marco histórico na análise de admissões em instituições de ensino superior. Ele surgiu no contexto de debates sobre desigualdade de gênero nos processos de seleção para programas de pós-graduação nos Estados Unidos durante os anos 1970, uma época em que movimentos pelos direitos civis e igualdade de gênero estavam em ascensão.

Na década de 1970, havia um crescente questionamento sobre práticas discriminatórias em ambientes acadêmicos e profissionais. A Universidade da Califórnia, Berkeley, uma das principais instituições de ensino superior dos EUA, publicou dados relacionados ao processo de admissão em seus programas de pós-graduação para investigar alegações de desigualdade de gênero.

### Problematização

O Conjunto de Dados de Admissões de Pós-Graduação de Berkeley em 1973 levanta questões centrais sobre a equidade nos processos seletivos em instituições de ensino superior. Os dados, à primeira vista, indicam uma desigualdade de gênero nas admissões, sugerindo que mulheres tinham menor probabilidade de serem aceitas em comparação aos homens. Essa evidência inicial gerou um debate sobre a possibilidade de discriminação institucional contra mulheres no ambiente acadêmico.


### Atividade

Desenvolver uma análise exploratória de dados que identifique padrões e propondo interpretações sobre viés e desigualdade na admissão de pós-graduação da Universidade de Berkeley em 1973.

Segue abaixo o escopo o trabalho:

#### **Preprocessamento dos dados (3 pontos)**

1. Carregar o dataset;


- 
2. Verificar suas informações gerais;
  3. Verificar dados ausentes;
  4. Verificar e remover outliers;
  5. Verificar e corrigir inconsistências;
  6. Criar novas variáveis derivadas das existentes;
  7. Converter tipos de dados;
  8. Salvar o dataset pré-processado.

### **Análise de dados (7 pontos)**

1. Carregar o dataset e verificar suas informações gerais;
2. Examinar as estatísticas descritivas das variáveis numéricas, verificando médias, medianas, desvios padrão, mínimos e máximos.
3. Analisar a distribuição de frequências das variáveis categóricas.
4. Verificar a representatividade de homens e mulheres nas aplicações e admissões por departamento.
5. Comparar o número total de candidatos (admitidos e rejeitados) por gênero.
6. Comparar as taxas de admissão por departamento e gênero.
7. Comparar as taxas de admissão agregadas e desagregadas por departamento.
8. Verificar a relação entre o número de candidatos e as taxas de admissão/rejeição por gênero e departamento.
9. Comparar as taxas de admissão agregadas para homens e mulheres.
10. Analisar como as taxas de admissão de homens e mulheres variam entre os departamentos.
11. Identificar quais departamentos têm as maiores e menores taxas de rejeição.
12. Avaliar a proporção de inscrições feitas por homens e mulheres nos departamentos mais e menos competitivos.
13. Comparar as taxas de admissão gerais com as taxas de admissão por departamento.
14. Verificar se existem preferências de departamentos por gênero que podem influenciar os resultados observados.

### **Questionamentos a serem respondidos durante a análise de dados**

1. Há departamentos ou gêneros que apresentam taxas de admissão ou rejeição desproporcionalmente altas ou baixas?
2. Há equilíbrio na distribuição de gêneros nos departamentos?
3. Como os padrões de admissão diferem entre os departamentos e gêneros?
4. Existe um viés aparente contra algum gênero?
5. Há departamentos mais favoráveis a um gênero específico?
6. Quais departamentos são mais competitivos e como isso afeta os gêneros?
7. Como as conclusões mudam ao desagregar os dados por departamento?

- 
8. Justifique o viés de gênero nas admissões de pós-graduação de Berkeley.

## Observações

1. Exigido o uso do dataset neste link: <https://discovery.cs.illinois.edu/dataset/berkeley/>;
2. Dúvidas serão respondidas pelo fórum da atividade;
3. Envio dos links do github e vídeo com máximo de 10 min com a explicação dos códigos;
4. Atividade em grupo de 5 pessoas, sendo uma avaliação do trabalho em grupo e uma avaliação individual do vídeo explicativo.
5. Prazo definido no AVA sem possibilidade de adiamento.

## Referências bibliográficas

1. **BICKEL, Peter J.; HAMMEL, Eugene A.; O'CONNELL, J. William.** Sex bias in graduate admissions: Data from Berkeley. Statistics and public policy, p. 113-130, 1977.
2. UNIVERSITY OF ILLINOIS. Berkeley's 1973 Graduate Admissions Dataset. Data Science Discovery, 2022. Disponível em: <https://discovery.cs.illinois.edu/dataset/berkeley/>. Acesso em: 6 dez. 2024.
3. WIKIPÉDIA. \*Paradoxo de Simpson\*. Disponível em: [https://pt.wikipedia.org/wiki/Paradoxo\\_de\\_Simpson](https://pt.wikipedia.org/wiki/Paradoxo_de_Simpson)>. Acesso em: 6 dez. 2024.