

Universidade de São Paulo - USP  
Instituto de Ciências Matemáticas e de Computação - ICMC

Análise de Dados de um processo seletivo fictício

Paulo Eduardo De Vincenzi  
Nº USP: 12743130

São Carlos, SP  
25 de agosto de 2025

## 1. Objetivo

Comparar o perfil de alunos selecionados ou não em um processo seletivo fictício.

## 2. Descrição dos Dados

O banco de dados do processo seletivo contém 3.000 cadastros de participantes. Cada registro corresponde a um candidato e inclui informações descritas a seguir:

- **Id\_aluno**: Identificador do aluno;
- **Selecionado** (Sim/Não): Indica se o aluno foi ou não selecionado;
- **Sexo**: F/M;
- **OrientacaoSexual**: 'Bissexual', 'Heterossexual', 'Homossexual', 'Pansexual', 'Outro', 'Prefiro não responder';
- **Idade**, em anos completos;
- **Raça-cor**: 'Amarela', 'Branca', 'Indígena', 'Parda', 'Preta', 'Não informada';
- **Deficiência** (Se possui deficiência e qual).

A distribuição geral dessas variáveis está apresentada nas Figuras 1 a 5, que mostram a composição da amostra.

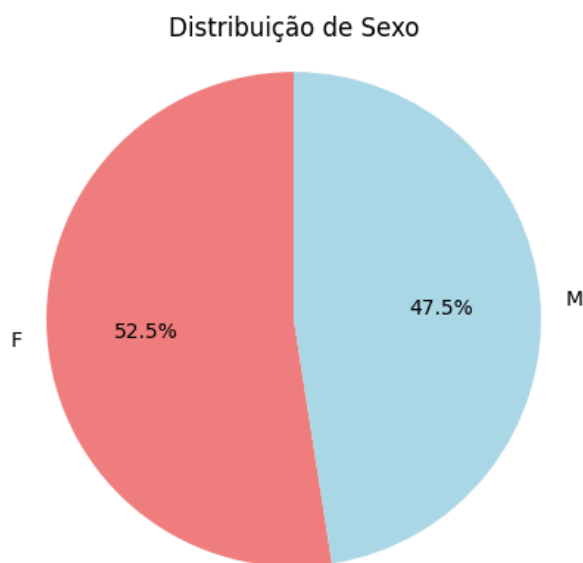


Figura 1 - Distribuição dos participantes por sexo.

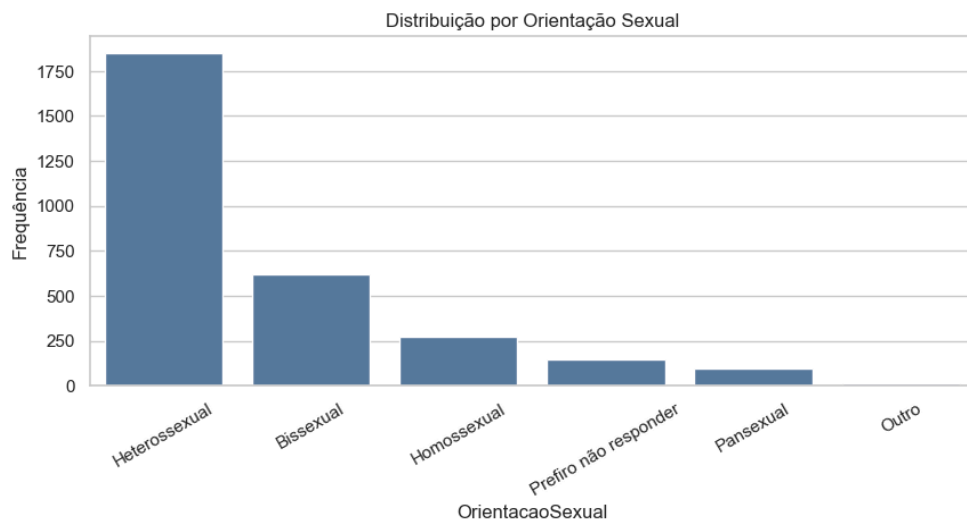


Figura 2 - Distribuição dos participantes por orientação sexual.

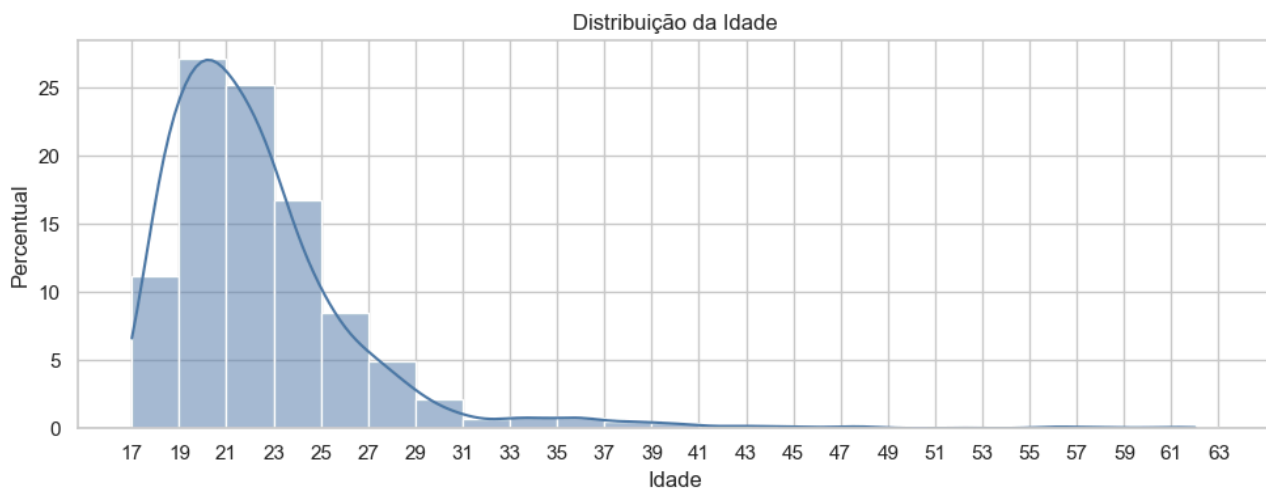


Figura 3 - Distribuição dos participantes por idade.

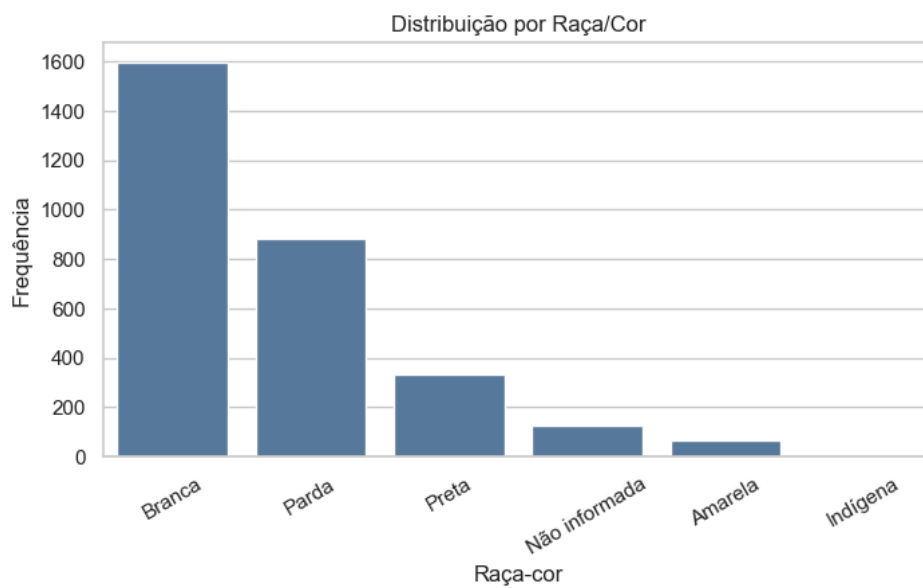


Figura 4 - Distribuição dos participantes por raça/cor.

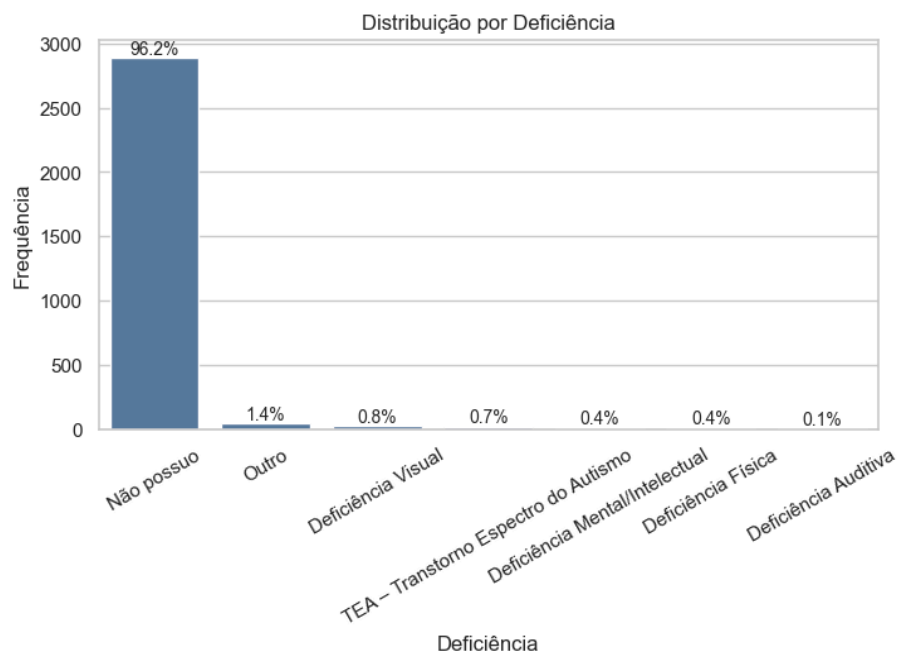


Figura 5 - Distribuição dos participantes por deficiência.

Observa-se que a amostra é majoritariamente composta por: jovens brancos de ambos os sexos, entre 17 e 29 anos, com concentração em 18 anos, sem deficiência e heterossexuais, conforme evidenciado nas figuras.

### 3. Análise Descritiva

A taxa de candidatos selecionados em cada categoria de atributo (sexo, orientação sexual, idade, raça/cor, e deficiência) está apresentada nas Figuras 6 a 10 a seguir.



Figura 6 - Taxa de candidatos selecionados por sexo.

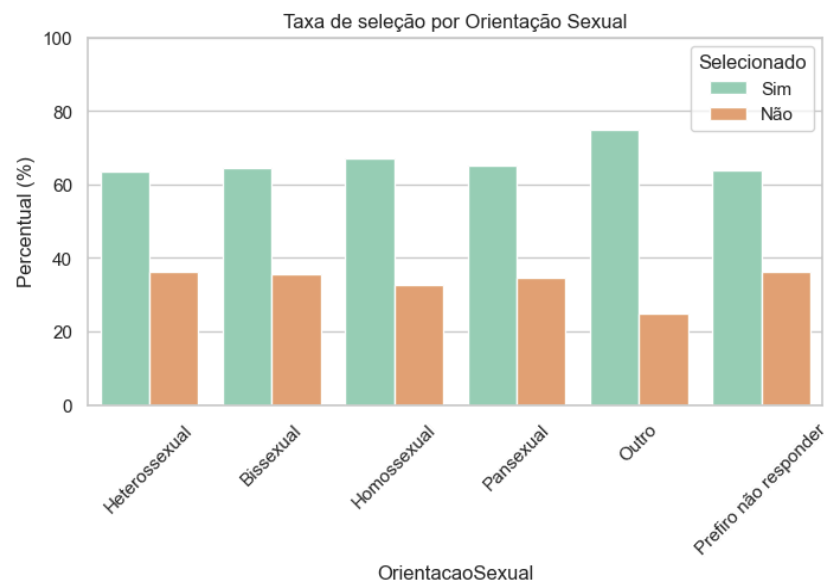


Figura 7 - Taxa de candidatos selecionados por orientação sexual.

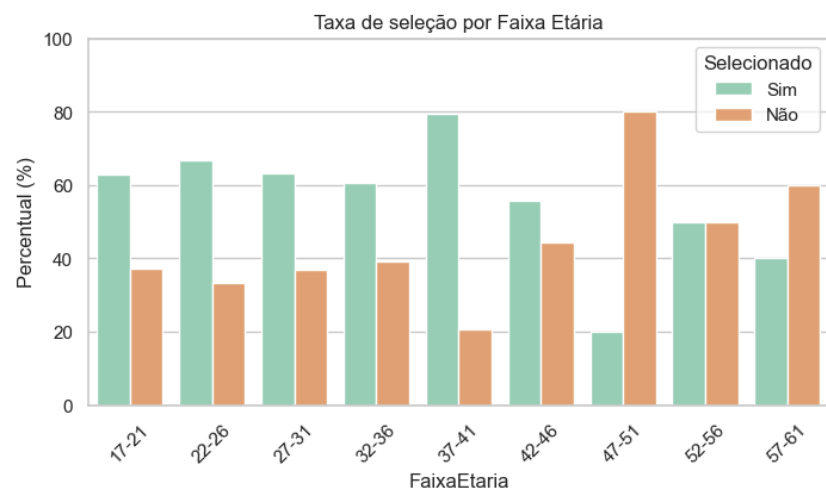


Figura 8 - Taxa de candidatos selecionados por faixa etária.

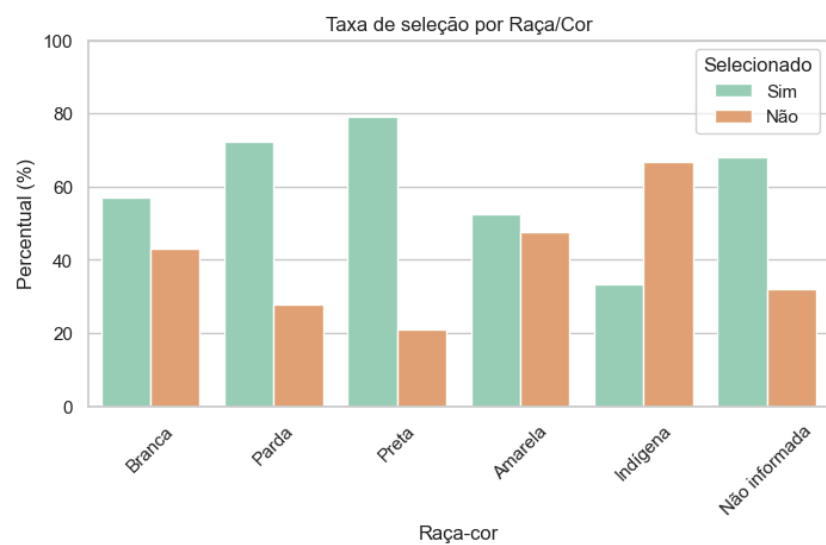


Figura 9 - Taxa de candidatos selecionados por raça/cor.

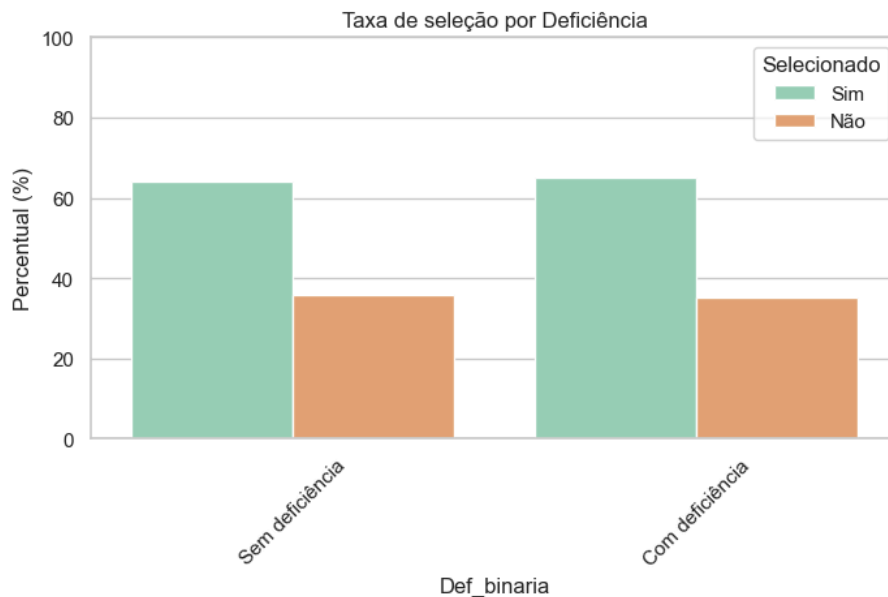


Figura 10 - Taxa de candidatos selecionados por deficiência.

As figuras acima permitem observar nítidos padrões na taxa de seleção dos candidatos do processo seletivo:

- Em relação ao sexo, a proporção de aprovados é semelhante entre homens e mulheres;
- Em relação à orientação sexual, a proporção de aprovados é praticamente semelhante entre os diferentes grupos;
- Para as faixas etárias, candidatos jovens e adultos até os 40 anos apresentam uma taxa de seleção ligeiramente maior que candidatos com idade mais avançada;
- Quanto à raça/cor, observa-se uma taxa de aprovação maior entre indivíduos de raça preta e parda em comparação com os de raça branca, que constituem a maioria da amostra;
- Por fim, não há diferença significativa entre a seleção entre indivíduos deficientes e não deficientes.

Conclui-se que, no processo seletivo, **a taxa de aprovação é relativamente maior para candidatos negros ou pardos na faixa etária de 18 a 40 anos**, em comparação com os demais participantes.

#### 4. Análise inferencial

Para avaliar se existe relação estatisticamente significativa entre as variáveis de interesse (sexo, orientação sexual, faixa etária, raça/cor e deficiência) e o resultado do processo seletivo (selecionado ou não selecionado), **foram realizados testes de associação**.

Considerando que todas as tabelas apresentaram frequências adequadas para o uso do teste, optou-se por aplicar exclusivamente o **teste qui-quadrado de Pearson**, que é apropriado para avaliar associações em tabelas de contingência com mais de duas categorias ou frequências suficientes.

A interpretação dos resultados baseia-se no p-valor considerando um nível de confiança de 95%:

- $p < 0,05$ : evidência de associação significativa entre a variável e o resultado do processo seletivo.
- $p \geq 0,05$ : não há evidência estatística de associação.

A seguir apresenta-se os testes de associação entre o número de selecionados e o(a):

- Sexo

Tabela 1 - distribuição do número de selecionados e não selecionados por sexo.

	Selecionado	Não Selecionado
Feminino	559	1015
Masculino	514	912

Resultados:

- $\text{Qui}^2 = 0,0699$
- Graus de liberdade = 1
- p-valor =  $0,7914 > 0,05$

Interpretação:

- p-valor não significativo
- Não há evidências de que homens e mulheres tenham taxas de aprovação diferentes.

- **Orientação Sexual**

Tabela 2 - distribuição da orientação sexual de selecionados e não selecionados

	Selecionado	Não Selecionado
Bissexual	220	400
Heterossexual	674	1174
Homossexual	90	184
Outro	4	12
Pansexual	33	62
Prefiro não responder	52	92

**Resultados:**

- $\chi^2 = 2,2334$
- Graus de liberdade = 5
- p-valor =  $0,8160 > 0,05$

**Interpretação:**

- p-valor não significativo
- Não há evidências de que diferentes orientações sexuais tenham diferentes taxas de seleção.

- **Idade**

Tabela 3 - distribuição de faixas etárias de selecionados e não selecionados

	Selecionado	Não Selecionado
17-21	572	969
22-26	375	747
27-31	84	144
32-36	22	34
37-41	6	23
42-46	4	5
47-51	4	1
52-56	2	2
57-61	3	2



**Resultados:**

- $\chi^2 = 13,3830$
- Graus de liberdade = 8
- p-valor = 0,0993 > 0,05

**Interpretação:**

- p-valor não significativo
- Não há evidências de que diferentes idades apresentam diferentes taxas de seleção.

O teste de associação para a variável idade apresentou um p-valor não significativo a 95% de confiança ( $p = 0,0993$ ), porém próximo do limite de 0,05. Considerando um nível de confiança de 90%, o p-valor é menor que 0,10, indicando que a idade passa a ser considerada estatisticamente significativa, sugerindo diferenças nas taxas de seleção entre faixas etárias.

- **Raça/Cor**

Tabela 4 - distribuição da raça/cor de selecionados e não selecionados

	Selecionado	Não Selecionado
Amarela	30	33
Branca	687	909
Indígena	2	1
Não informada	40	85
Parda	244	636
Preta	70	263

**Resultados:**

- $\chi^2 = 98,9457$
- Graus de liberdade = 5
- p-valor = 0,0000 < 0,05

**Interpretação:**

- p-valor significativo
- Há evidências de que diferentes raças/cores têm diferentes taxas de seleção.

- **Deficiência**

Tabela 5 - distribuição de deficiência de selecionados e não selecionados

	Selecionado	Não Selecionado
Deficiência Auditiva	1	3
Deficiência Física	3	8
Deficiência Mental/Intelectual	4	7
Deficiência Visual	11	14
Não Possui	1033	1853
Outro	11	31
TEA	10	11

**Resultados:**

- $\chi^2 = 4,2481$
- Graus de liberdade = 6
- p-valor =  $0,6431 > 0,05$

**Interpretação:**

- p-valor não significativo
- Não há evidências de que indivíduos deficientes, ou não, têm diferentes taxas de seleção.

## 5. Conclusão

No geral, não houve evidências de diferença significativa na taxa de seleção quanto a sexo, orientação sexual ou deficiência. Contudo, a idade mostrou algum indício de diferença (p-valor = 0,0993), em nível de 90% de confiança, e a raça/cor apresentou associação significativa (p-valor  $< 0,01$ , indicando confiança de mais de 99%). Assim, houve maior seleção de indivíduos jovens negros e pardos.

## 6. Repositório

O repositório que contém o código usado para analisar o banco de dados está disponível em:

[https://github.com/PauloVincenzi/analise\\_processo\\_seletivo\\_ficticio\\_usp](https://github.com/PauloVincenzi/analise_processo_seletivo_ficticio_usp)