

Distribuição de frequências para variáveis qualitativas

Prof. Walmes M. Zeviani

Departamento de Estatística
Universidade Federal do Paraná



- ▶ Tabelas de frequência.
- ▶ Tipos de frequência.
- ▶ Cálculo de frequência.
- ▶ Tabelas de frequência cruzada.
- ▶ Gráficos para distribuição de frequência.



Figura 1. Photo by nappy from Pexels. Fonte <https://rb.gy/nigarg>.

Tabelas de frequência para uma variável qualitativa

Classe

Uma classe é uma das categorias nas quais uma variável qualitativa pode ser classificada.

Frequência da classe

Frequência da classe é o número de observações no conjunto de dados que pertence a uma particular classe.

Frequência relativa de classe

Frequência relativa de classe é a frequência de classe dividida pelo número total de observações no conjunto de dados.

Percentual da classe

Percentual da classe é a frequência relativa de classe multiplicada por 100.

Uma tabela de dados

Tabela 1. Um recorte aleatório de 10 observações da tabela de dados resultante da pesquisa feita no Moodle com alunos matriculados em Estatística que já responderam à enquete. Total de registros: 977 respondentes.

Área	Email	Adequação	Idade	CH
Sociais	outros	Não sei opi.	22	120
Biolog	hotmail	Não sei opi.	48	8
Humanas	gmail	Adeq.	24	210
Eng	hotmail	Tot. adeq.	20	300
Eng	gmail	Tot. adeq.	20	300
Sociais	gmail	Tot. adeq.	33	120
Exatas	hotmail	Não sei opi.	23	0
Biolog	outros	Adeq.	25	150
Eng	hotmail	Adeq.	25	420
Exatas	outros	Tot. adeq.	23	200

A variável ordenada

Tabela 2. Tabela anterior com registros ordenados alfabeticamente pelos rótulos da variável **Área**.

Área	Email	Adequação	Idade	CH
Biolog	hotmail	Não sei opi.	48	8
Biolog	outros	Adeq.	25	150
Eng	hotmail	Tot. adeq.	20	300
Eng	gmail	Tot. adeq.	20	300
Eng	hotmail	Adeq.	25	420
Exatas	hotmail	Não sei opi.	23	0
Exatas	outros	Tot. adeq.	23	200
Humanas	gmail	Adeq.	24	210
Sociais	outros	Não sei opi.	22	120
Sociais	gmail	Tot. adeq.	33	120

A tabela de frequência do recorte

Tabela 3. Tabela de frequência considerando apenas os registros do recorte de 10 observações para exemplificar os cálculos.

Área	Frequência	Freq. relativa
Eng	3	0.30
Biolog	2	0.20
Exatas	2	0.20
Sociais	2	0.20
Humanas	1	0.10
Total	10	1.00

A tabela de frequência de todos os respondentes

Tabela 4. Tabela de frequência para todos respondentes da disciplina.

Área	Frequência	Freq. relativa
Eng	241	0.25
Saúde	212	0.22
Biolog	177	0.18
Exatas	175	0.18
Humanas	122	0.12
Sociais	50	0.05
Total	977	1.00

- ▶ Frequência (absoluta, f_a) de classe:

f_a = número de observações da classe.

- ▶ Frequência relativa (f_r) de classe:

$$f_r = \frac{\text{número de observações da classe}}{\text{total de observações}}.$$

Gráficos para frequência de uma variável qualitativa

Gráfico de barras verticais

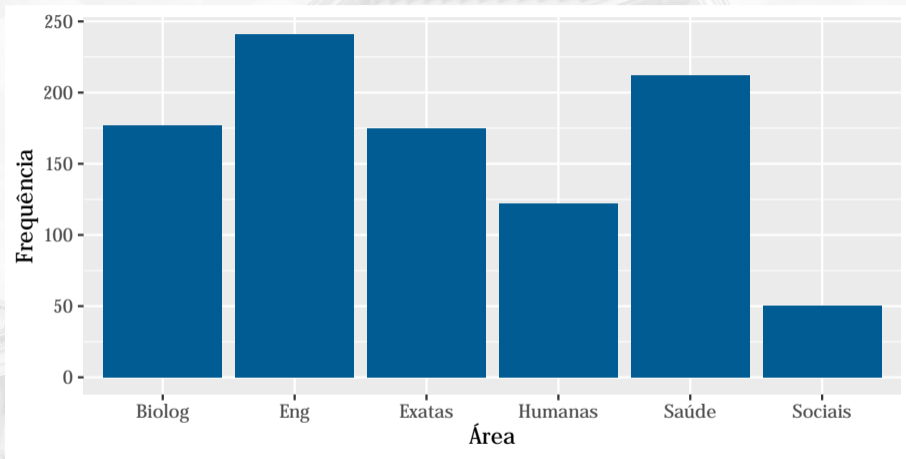


Figura 2. Gráfico de barras com a frequência absoluta para as classes da variável Área.

Gráfico de barras horizontais

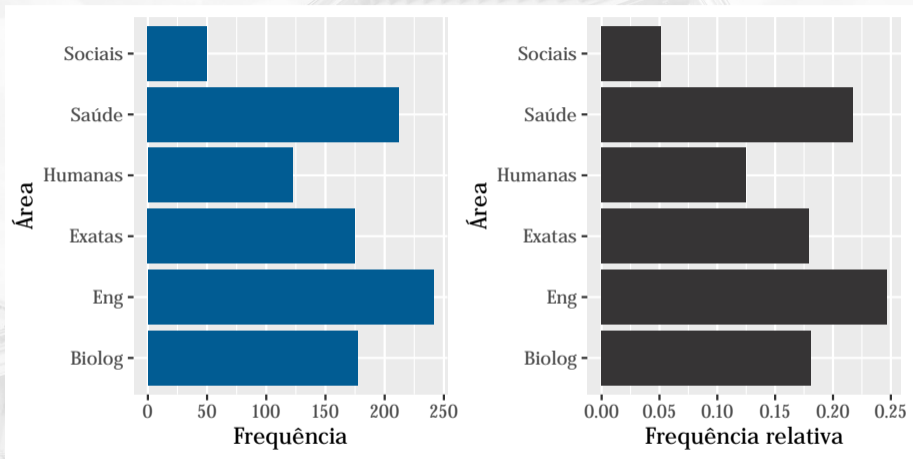


Figura 3. Gráfico de barras com a frequência absoluta (esq.) e relativa (dir.) para as classes da variável Área.

Gráfico de barras empilhadas

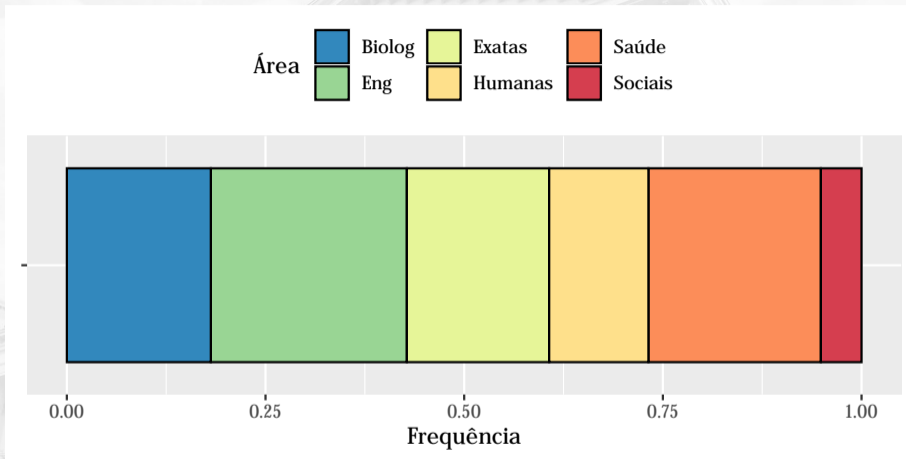


Figura 4. Gráfico de barras empilhadas com a frequência relativa para as classes da variável Área.

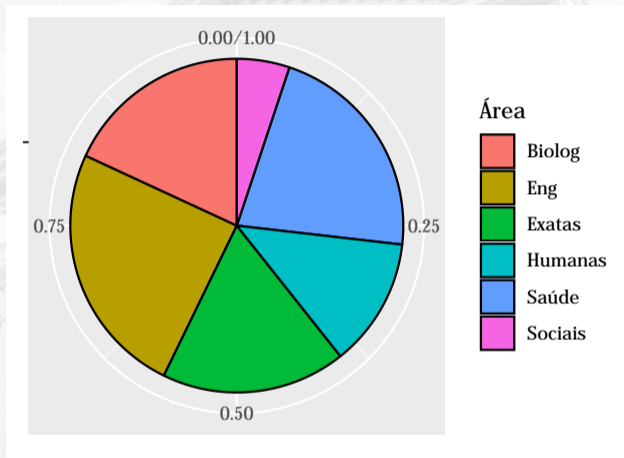


Figura 5. Gráfico de setores com a frequência relativa para as classes da variável Área.

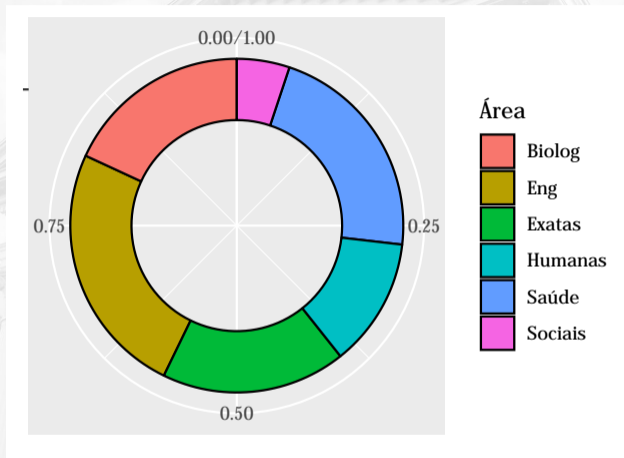


Figura 6. Gráfico de rosca com a frequência relativa para as classes da variável Área.

- ▶ Variáveis nominais.
 - ▶ Não há ordenação natural das classes.
 - ▶ Não há ordem para exibição das frequências das classes.
 - ▶ Ordem alfabética auxilia procurar pela classe quando existem muitas.
 - ▶ A ordenação pela frequência auxilia identificar as classes predominantes e minoritárias.
- ▶ Variáveis ordinais.
 - ▶ Há ordenação natural das classes.
 - ▶ Procurar manter a ordem das classes para uma exibição coerente.
 - ▶ Quando for o caso, pode-se ordenar pela frequência.

Tabelas de frequência para duas variáveis qualitativas

Tabela de frequência cruzada

- ▶ Considera as combinações de duas variáveis simultaneamente.
- ▶ Cada combinação de classes é chamada cédula.
- ▶ Agora tem-se:
 - ▶ As frequências das cédulas.
 - ▶ As frequências marginais.

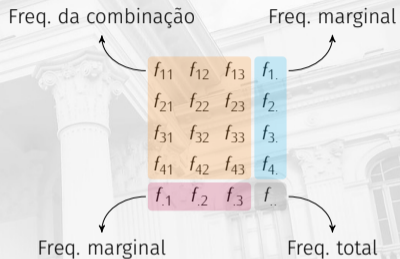


Figura 7. Esquema de disposição das frequências em uma tabela para duas variáveis. Fonte: os autores.

A tabela de dados

Tabela 5. Tabela com registros ordenados alfabeticamente pelas classes da variável **Email** seguida da variável Área.

Área	Email	Adequação	Idade	CH
Eng	gmail	Tot. adeq.	20	300
Humanas	gmail	Adeq.	24	210
Sociais	gmail	Tot. adeq.	33	120
Biolog	hotmail	Não sei opi.	48	8
Eng	hotmail	Tot. adeq.	20	300
Eng	hotmail	Adeq.	25	420
Exatas	hotmail	Não sei opi.	23	0
Biolog	outros	Adeq.	25	150
Exatas	outros	Tot. adeq.	23	200
Sociais	outros	Não sei opi.	22	120

Tabela de frequência cruzada para o recorte

Tabela 6. Tabela de frequência (absoluta) cruzada para as variáveis Email e Área considerando as observações do recorte aleatório.

Área	gmail	hotmail	outros	Sum
Biolog	0	1	1	2
Eng	1	2	0	3
Exatas	0	1	1	2
Humanas	1	0	0	1
Sociais	1	0	1	2
Sum	3	4	3	10

Tabela de frequência relativa cruzada para o recorte

Tabela 7. Tabela de frequência relativa cruzada para as variáveis Email e Área considerando as observações do recorte aleatório.

Área	gmail	hotmail	outros
Biolog	0.00	0.10	0.10
Eng	0.10	0.20	0.00
Exatas	0.00	0.10	0.10
Humanas	0.10	0.00	0.00
Sociais	0.10	0.00	0.10

Tabela de frequência cruzada de todos os respondentes

Tabela 8. Tabela de frequência (absoluta) cruzada para as variáveis Email e Área para todos os respondentes.

Área	gmail	hotmail	outros	ufpr	Sum
Biolog	109	39	11	18	177
Eng	124	74	24	19	241
Exatas	98	49	22	6	175
Humanas	76	31	12	3	122
Saúde	121	70	15	6	212
Sociais	25	14	8	3	50
Sum	553	277	92	55	977

Tabela 9. Tabela de frequência relativa cruzada, em percentual, para as variáveis Email e Área para todos os respondentes.

Área	gmail	hotmail	outros	ufpr	Sum
Biolog	11.16	3.99	1.13	1.84	18.12
Eng	12.69	7.57	2.46	1.94	24.67
Exatas	10.03	5.02	2.25	0.61	17.91
Humanas	7.78	3.17	1.23	0.31	12.49
Saúde	12.38	7.16	1.54	0.61	21.70
Sociais	2.56	1.43	0.82	0.31	5.12
Sum	56.60	28.35	9.42	5.63	100.00

- ▶ Frequência relativa de cédula ou combinação:

f_a = número observado da combinação de classes.

- ▶ Frequência relativa de cédula ou combinação:

$f_r = \frac{\text{número observado da combinação de classes}}{\text{número total de observações}}$

- ▶ Considera duas ou mais variáveis **simultaneamente**.
- ▶ Permite investigar relação de **associação** entre as variáveis.
- ▶ As **margens** contêm as distribuições de frequência de cada variável.

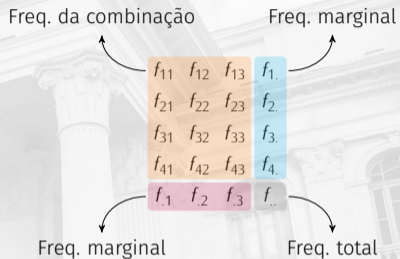


Figura 8. Esquema de disposição das frequências em uma tabela para duas variáveis. Fonte: os autores.

Gráficos para frequência de duas variáveis qualitativas

Gráficos de barras empilhadas

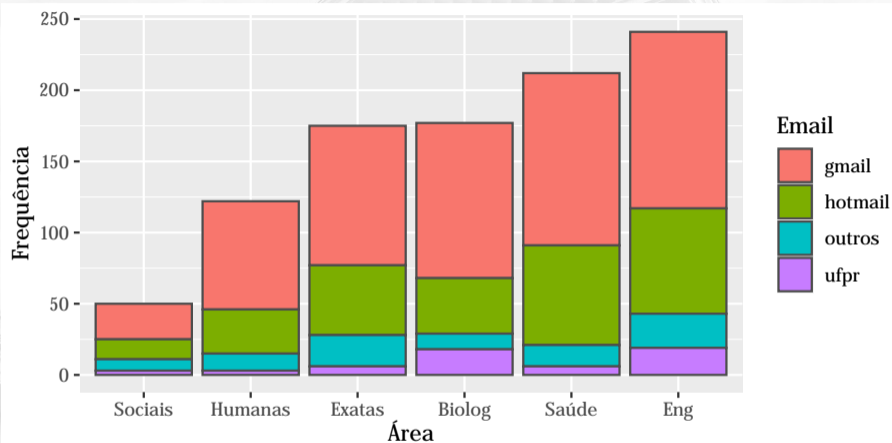


Figura 9. Gráfico de barras empilhadas com a frequência absoluta para as combinações das variáveis Email e Área.

Gráficos de barras lado a lado

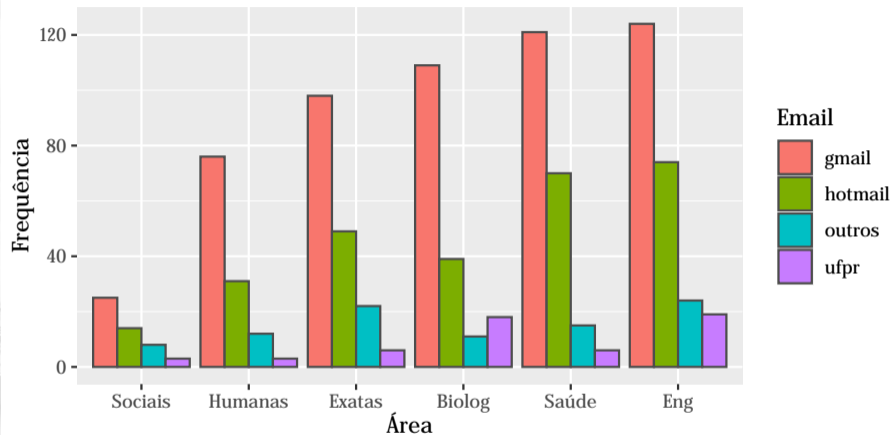


Figura 10. Gráfico de barras lado a lado com a frequência absoluta para as combinações das variáveis Email e Área.

Gráficos de barras empilhadas relativo

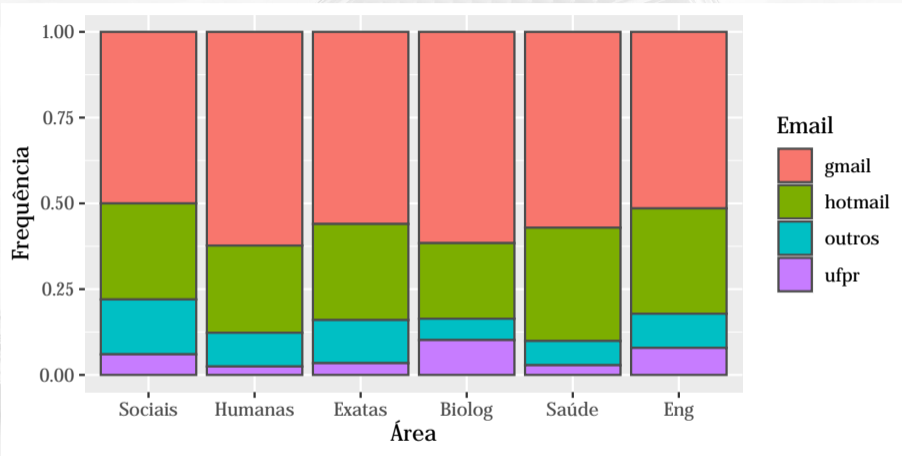


Figura 11. Gráfico de barras empilhadas relativo com a frequência relativa para as combinações das variáveis Email e Área.

Gráfico de mosaico

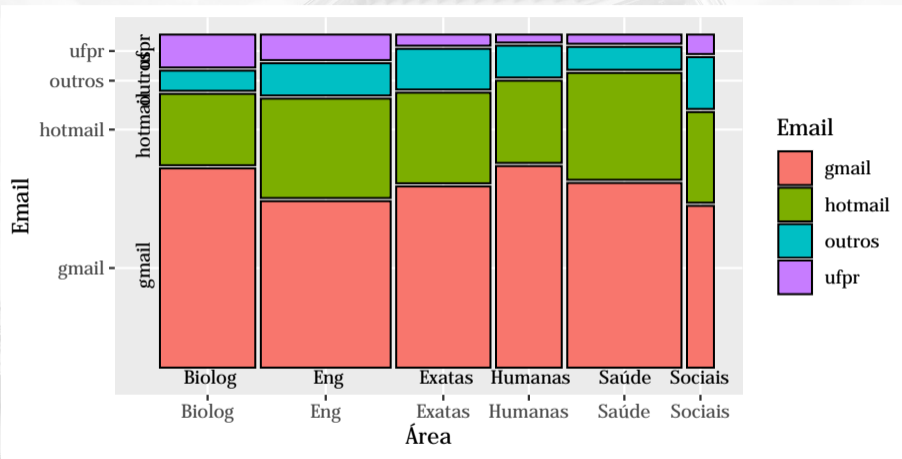


Figura 12. Gráfico de mosaico com a frequência relativa para as combinações das variáveis Email e Curso.

Quando usar cada tipo

- ▶ Barras empilhadas.
 - ▶ Comparar as frequências marginais da variável primária.
 - ▶ Comparar a composição da variável secundária dentro das classes da primária.
- ▶ Barras lado a lado.
 - ▶ Comparar as frequências da variável secundária dentro de uma classe da primária.
 - ▶ Comparar entre combinações das variáveis.
- ▶ Barras empilhadas relativas.
 - ▶ Para comparar a composição da variável secundária dentro das classes da variável primária.
- ▶ Mosaico.
 - ▶ Comparar as frequências relativas marginais da variável primária.
 - ▶ Comparar a composição da variável secundária dentro das classes da variável primária.
 - ▶ Comparar as frequências relativas entre combinações.
 - ▶ Verificar independência entre variáveis qualitativas.



Considerações finais

Revisão

- ▶ Análise descritiva para variáveis qualitativas.
 - ▶ Tabelas de frequência: simples e cruzada.
 - ▶ Gráficos de frequência: barras, setores, mosaico.
- ▶ Função da análise.
 - ▶ Uma variável: descrever o comportamento → distribuição de frequências.
 - ▶ Duas ou mais: investigar relações.



Figura 13. Photo by nappy from Pexels. Fonte <https://rb.gy/nigarg>.