

Análise exploratória I

Gráficos e tabelas para variáveis qualitativas

Prof. Me. Lineu Alberto Cavazani de Freitas

CE003 – Estatística II

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Análise exploratória

- ▶ Parte primordial de qualquer análise estatística é chamada **análise descritiva** ou **exploratória**.
- ▶ Consiste basicamente de **tabelas**, **resumos numéricos** e **análises gráficas** das variáveis disponíveis em um conjunto de dados.
- ▶ Trata-se de uma etapa de extrema importância e deve preceder qualquer análise mais sofisticada.
- ▶ As técnicas de análise exploratória visam **resumir** e **apresentar** as informações de um conjunto de dados brutos.

Análise exploratória

- ▶ Tentar compreender um conjunto de dados sem algum método que permita resumir as informações é inviável.
- ▶ A análise exploratória é a primeira forma de tentarmos entender o que acontece nos nossos dados.
- ▶ Uma das tarefas é a etapa de consistência dos dados, isto é, verificar se os dados coletados são condizentes com a realidade.



Figura 1. Extraído de pixabay.com.

Análise exploratória

- ▶ O conjunto de técnicas aplicáveis está diretamente associado ao **tipo das variáveis de interesse** (quantitativas x qualitativas) e suas ramificações.
- ▶ Podemos conduzir análises focadas nas variáveis uma a uma (**análises univariadas**).
- ▶ Bem como conduzir análises focadas em avaliar a relação entre as variáveis (**análises multivariadas**).



Figura 2. Extraído de pixabay.com.

Análise exploratória

Podemos fazer uso diversas técnicas, tais como

- ▶ Tabelas de frequência absolutas.
- ▶ Tabelas de frequência relativas.
- ▶ Tabelas de frequência acumuladas.
- ▶ Tabelas para múltiplas variáveis.
- ▶ Gráficos (para análise uni e multivariada).
- ▶ Medidas de posição central.
- ▶ Medidas de posição relativa.
- ▶ Medidas de forma.
- ▶ Medidas de dispersão.
- ▶ Medidas de associação.



Análise descritiva univariada para variáveis qualitativas

Análise descritiva univariada para variáveis qualitativas

- ▶ Uma variável qualitativa representa um atributo que pode ser expresso por meio de **rótulos** ou **categorias**.
- ▶ Podem ser classificadas em **nominais** (sem ordenação natural entre os níveis) ou **ordinais** (com ordenação natural entre os níveis).
- ▶ As categorias também são chamadas de **classes** ou **níveis**.
- ▶ Na análise descritiva de uma variável qualitativa estamos interessados em avaliar as **frequências** das classes.

Tipos de frequência

- ▶ **Frequência absoluta** (f_a): número de observações no conjunto de dados que pertence a uma determinada classe.
- ▶ **Frequência relativa** (f_r): frequência de classe dividida pelo número total de observações no conjunto de dados.
 - ▶ Pode ser apresentada em forma de percentual, quando multiplicada por 100.
- ▶ **Frequência acumulada** (F_a ou F_r): frequência absoluta ou relativa acumulada conforme disposição das classes.
 - ▶ Não faz muito sentido para variáveis qualitativas nominais.

Tipos de frequência

Exemplos

- ▶ Frequência absoluta e relativa dos alunos por gênero.
 - ▶ 16 do sexo masculino.
 - ▶ 10 do sexo feminino.
 - ▶ $16/26 = 0,615$ do sexo masculino (61,5%).
 - ▶ $10/26 = 0,385$ do sexo feminino (38,5%).

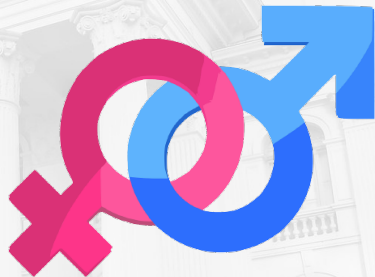


Figura 3. Extraído de pixabay.com.

Tabelas de frequência para uma variável qualitativa

- ▶ Utilizando apenas os dados brutos é difícil responder questões de interesse.
- ▶ Para reduzir os dados originais de forma que fique mais claro o entendimento dos mesmos são utilizadas as **tabelas de frequência**.
- ▶ No caso de variáveis qualitativas consiste em listar os possíveis níveis da variável e fazer a contagem de quantas vezes cada nível aparece nos dados brutos.



Figura 4. Extraído de pixabay.com.

Tabelas de frequência para uma variável qualitativa

- ▶ Cada **linha** da tabela diz respeito a um **nível** da variável categórica.
- ▶ As **colunas** podem apresentar diferentes tipos de **frequência** (absoluta, relativa, acumulada).
- ▶ Alguns cuidados para a apresentação dos resultados dizem respeito ao tipo de variável categórica em questão: nominal ou ordinal.
- ▶ Os níveis de variáveis **nominais não apresentam uma ordenação natural**, portanto, na apresentação dos resultados pode ser interessante ordenar os níveis por frequência ou por ordem alfabética.
- ▶ Esta estratégia não é recomendada para variáveis **ordinais**, pois estas **apresentam uma ordenação natural** e esta ordenação deve ser preferencialmente mantida na exposição dos resultados.

Tabelas de frequência para uma variável qualitativa nominal

Tabela 1. Tabela de frequências para a origem dos alunos.

| Origem | Frequência | Freq. Relativa |
|----------------------|------------|----------------|
| Curitiba | 10 | 0.38 |
| Interior do Paraná | 4 | 0.15 |
| Outro Estado | 9 | 0.35 |
| Região metropolitana | 3 | 0.12 |
| Total | 26 | 1.00 |

Tabelas de frequência para uma variável qualitativa nominal

Tabela 2. Tabela de frequências para a origem dos alunos.

| Origem | Frequência | Freq. Relativa |
|----------------------|------------|----------------|
| Curitiba | 10 | 0.38 |
| Outro Estado | 9 | 0.35 |
| Interior do Paraná | 4 | 0.15 |
| Região metropolitana | 3 | 0.12 |
| Total | 26 | 1.00 |

Tabelas de frequência para uma variável qualitativa nominal

Tabela 3. Tabela de frequências para a origem dos alunos.

| Origem | Frequência | Percentual |
|----------------------|------------|------------|
| Curitiba | 10 | 38 % |
| Outro Estado | 9 | 35 % |
| Interior do Paraná | 4 | 15 % |
| Região metropolitana | 3 | 12 % |
| Total | 26 | 100 % |

Tabelas de frequência para uma variável qualitativa ordinal

Tabela 4. Tabela de frequências para a escolaridade dos alunos.

| Escolaridade | Frequência | Freq. Relativa | Freq. Acumulada | Freq. Rel. Acumulada |
|-----------------|------------|----------------|-----------------|----------------------|
| Ensino médio | 21 | 0.81 | 21 | 0.81 |
| Ensino superior | 5 | 0.19 | 26 | 1.00 |
| Mestrado | 0 | 0.00 | 26 | 1.00 |
| Doutorado | 0 | 0.00 | 26 | 1.00 |
| Total | 26 | 1.00 | 26 | 1.00 |

Tabelas de frequência para uma variável qualitativa ordinal

Tabela 5. Tabela de frequências para a escolaridade dos alunos.

| Escolaridade | Frequência | Percentual | Freq. Acumulada | Percentual Acumulado |
|-----------------|------------|------------|-----------------|----------------------|
| Ensino médio | 21 | 81 % | 21 | 81 % |
| Ensino superior | 5 | 19 % | 26 | 100 % |
| Mestrado | 0 | 0 % | 26 | 100 % |
| Doutorado | 0 | 0 % | 26 | 100 % |
| Total | 26 | 100 % | 26 | 100 % |

Gráficos para representação de frequências de uma variável qualitativa

- ▶ A representação por meio de tabelas é útil mas nem sempre eficiente.
- ▶ Em diversos casos pode ser mais conveniente utilizar um **gráfico**.
- ▶ “Uma imagem vale mais que mil palavras”.
- ▶ Os cuidados com a ordenação dos níveis de acordo com o tipo da variável se mantêm.

Algumas possibilidades são:

- ▶ Gráfico de barras verticais.
- ▶ Gráfico de barras horizontais.
- ▶ Gráfico de barras empilhadas.
- ▶ Gráfico de setores.

Gráfico de barras

► Gráfico de barras verticais ou horizontais.

- Utiliza os possíveis **níveis** das variáveis **em um eixo**.
- As **frequências ou porcentagens** ficam **no outro eixo**.
- O tamanho da barra correspondente à frequência ou percentual.

► Gráfico de barras empilhadas.

- Usa-se **uma única barra**.
- A barra é dividida de acordo com a **contribuição relativa** de cada nível da variável.
- Representa-se a frequência relativa ou o percentual.

Gráfico de barras verticais

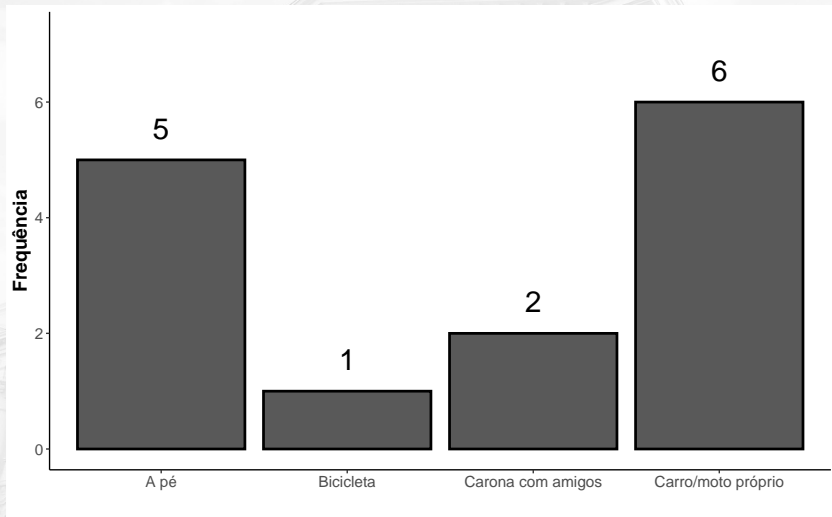


Figura 5. Gráfico de barras verticais para o meio de transporte dos alunos.

Gráfico de barras verticais

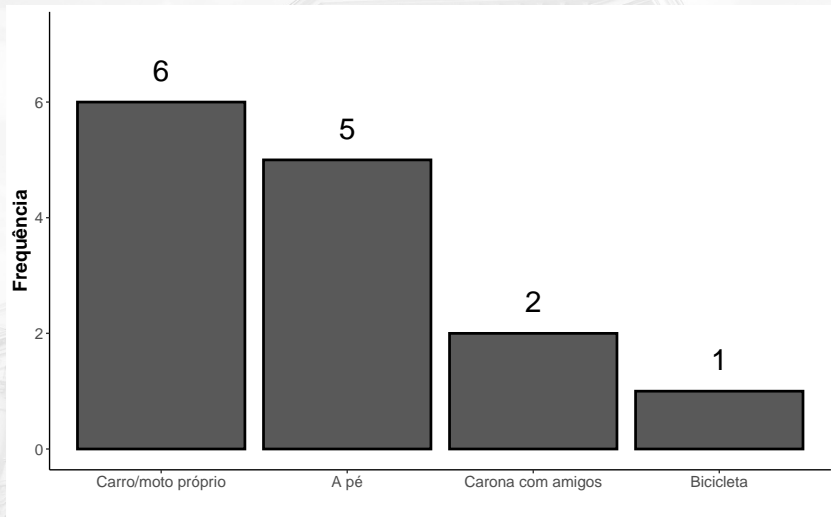


Figura 6. Gráfico de barras verticais para o meio de transporte dos alunos.

Gráfico de barras horizontais

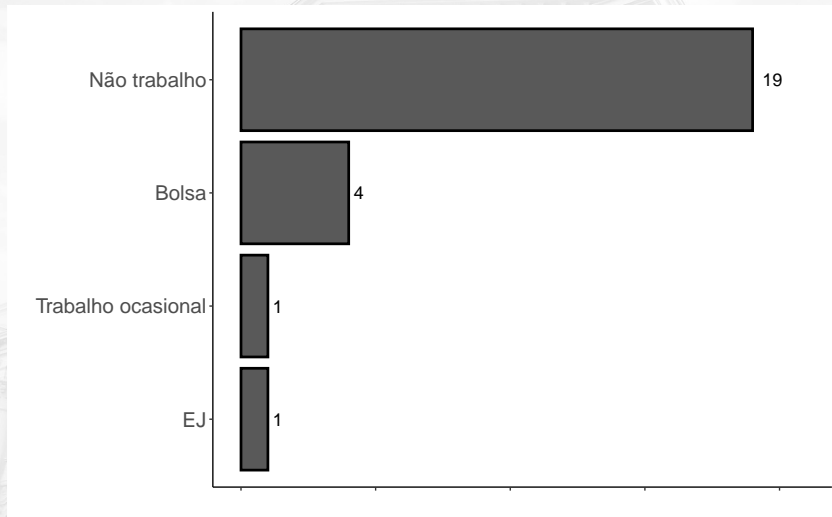


Figura 7. Gráfico de barras horizontais para a situação de emprego dos alunos.

Gráfico de barras empilhadas

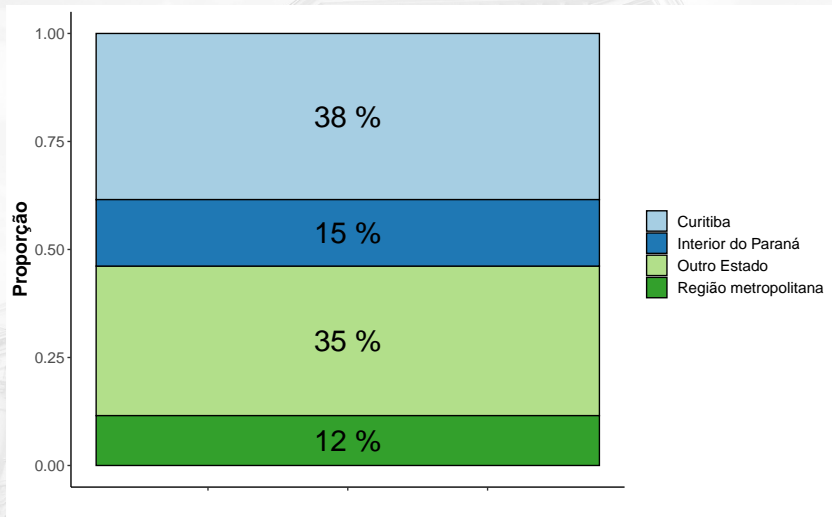


Figura 8. Gráfico de barras empilhadas para a origem dos alunos.

Gráfico de setores

- ▶ Consiste em **repartir um círculo** em setores de tamanhos proporcionais às **frequências relativas** ou às **porcentagens** de cada valor.
- ▶ Pode ser usados para representar variáveis com **poucos níveis**.
- ▶ Apesar de muito usado e preferido em diversas áreas, deve ser evitado.
- ▶ O cérebro humano tem dificuldade em relacionar frequências com áreas relativas.
- ▶ Para variáveis com muitos níveis, o gráfico tende a ficar visualmente poluído e pouco informativo.
- ▶ Outro problema é que níveis com frequências iguais a o deixam de aparecer no gráfico, diferente de um gráfico de barras.

Gráfico de setores

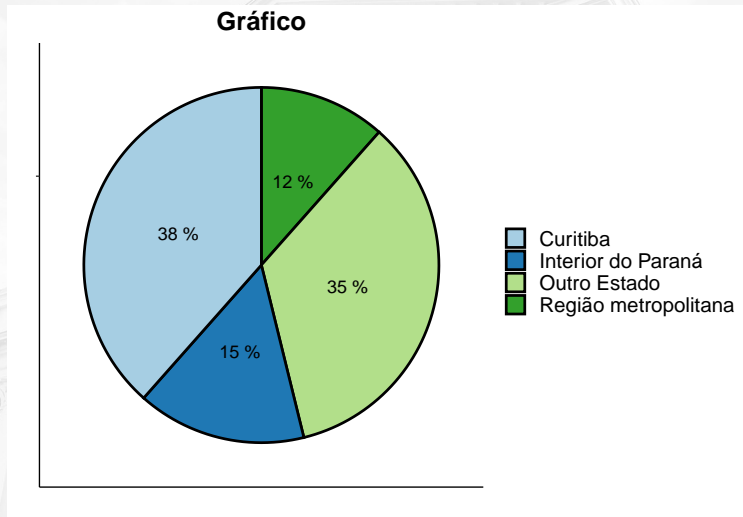


Figura 9. Gráfico de setores para a variável...



O que foi visto:

- ▶ Introdução à análise exploratória.
- ▶ Análise exploratória univariada para variáveis qualitativas.

Próximos assuntos:

- ▶ Análise exploratória univariada para variáveis quantitativas.