

Conceitos básicos

Gilberto Pereira Sassi

Universidade Federal da Bahia
Instituto de Matemática e Estatística
Departamento de Estatística

Programa do curso

- 1 Estatística descritiva: tabela de distribuição de frequência e gráficos (gráfico de barras e histograma), associação entre variáveis;
- 2 Medidas de resumo: medidas de posição e medidas de dispersão;
- 3 Probabilidade: axiomas de probabilidade, espaço amostral, ponto amostral;
- 4 Probabilidade condicional: regra de produto das probabilidades, teorema de probabilidade total e teorema de Bayes;
- 5 Variável aleatória discreta: valor esperado e variância;
- 6 Modelos probabilísticos para variável aleatória discreta: Uniforme, Bernoulli, Binomial, Poisson, Geométrico;
- 7 Variável aleatória contínua: valor esperado e variância;
- 8 Modelos probabilísticos: Uniforme, Normal, Exponencial, t-Student, Qui-quadrado;
- 9 Distribuições amostrais e teorema do limite central;
- 10 Estimativa intervalar: média e proporção;
- 11 Teste de Hipóteses: Erros tipo I e II, nível de significância, poder do teste;
- 12 Teste para média e proporção;
- 13 p-valor;
- 14 Teste de aderência: qui-quadrado;
- 15 Associação entre duas variáveis quantitativas: gráfico de dispersão e coeficiente de correlação linear de Pearson;
- 16 Regressão linear simples;
- 17 Qualidade do ajuste: análise de resíduo.

Critério de avaliação

Critério de avaliação

- Duas provas: P_1 e P_2 ;
- Nota final: $NF = \frac{P_1 + P_2}{2}$.

Datas de provas

- Primeira prova: 28/04/2020;
- Segunda prova: 30/06/2020;
- Segunda chamada da primeira prova: 07/07/2020;
- Segunda chamada da segunda prova: 09/07/2020.

Página: listas e slides do curso

<https://gilberto-sassi.github.io>

Referências

Principal

- 1 Barbetta, Pedro Alberto – Estatística Aplicada às Ciências Sociais;
- 2 Bussab, Wilton de Oliveira; Morettin, Pedro Alberto – Estatística Básica;

Complementar

- 1 Soares, José Francisco; Siqueira, Armindo Lúcia – Introdução a estatística médica;
- 2 Triola, Mario F. – Introdução a estatística.

Contato

E-mail

gilberto.sassi@ufba.br

Onde me encontrar

Sala 220 do Instituto de Matemática e Estatística.

Tipos de inferência

Dedutiva

Usa-se argumentos lógicos para chegar a conclusões a partir de premissas.

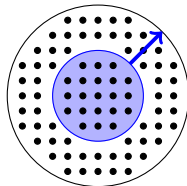
Exemplo:

- Premissa: Todo ser humano nascido em solo brasileiro, tem direito a cidadania brasileira.
- João nasceu em Salvador, então João tem direito a cidadania brasileira.

Indutiva

Processo de generalização da parte ao todo. (Objeto de estudo desse curso).

Exemplo: Em uma pesquisa eleitoral, entrevista-se parte do eleitorado para estudar a intenção de votos para cada candidato.



Nomenclatura

População Todos elementos alvo do estudo;

Amostra Parte da população;

Parâmetro Característica da população;

Variável Característica de um elemento da população. Usamos uma letra maiúscula para representar essa característica;

Estimativa Característica de amostra.

Exemplo

Intenção de voto para Lula em outubro de 2018.

População Todos eleitores aptos a votar em outubro de 2018;

Amostra Três mil eleitores que serão questionados se pretendem, ou não, em Lula;

Parâmetro Proporção de eleitores que votariam em Lula;

Estimativa Proporção de eleitores entre os três mil eleitores da amostra;

Variável Intenção de voto de cada eleitor (vota ou não no Lula).

Classificação de variáveis

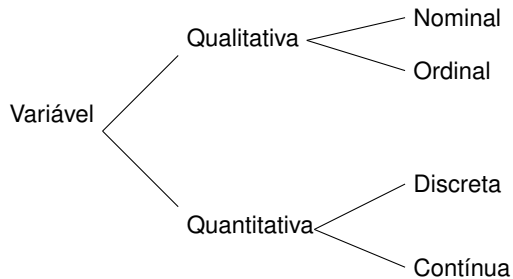
Variáveis podem ser variáveis em quatro categorias:

Variável Qualitativa Nominal Característica não-numérica com valores possíveis sem hierarquia. **Exemplo:** Gênero (valores possíveis: Masculino, Feminino, Outros).

Variável Qualitativa Ordinal Característica não-numérica com valores possíveis com hierarquia. **Exemplo:** Escolaridade (valores possíveis: Ensino Fundamental, Ensino Médio, Ensino Superior).

Variável Quantitativa Discreta Característica numérica geralmente proveniente de um processo de contagem. **Exemplo:** Número de filhos (valores possíveis: 0, 1, 2, 3, 4, ...).

Variável Quantitativa Contínua Característica numérica possivelmente não inteiro. **Exemplo:** Salário (um operário pode ganhar R\$1200,34).



Estatística Descritiva

Objetivo

Estudar padrões e comportamentos presentes na amostra (ou na população) usando gráficos e medidas resumo.

Vamos introduzir as técnicas gráficas usando exemplos.

Exemplo – Tabela 2.1 do livro do Morettin

Um pesquisador está interessado em fazer um levantamento sobre alguns aspectos socioeconômicos dos empregados da seção de orçamentos da companhia MB. Usando informações obtidos do departamento pessoal, ele obteve uma amostra de 36 funcionários e armazenou as seguintes variáveis:

Variável	Representação	Tipo
Estado civil	X	Qualitativa Nominal
Escolaridade	Y	Qualitativa Ordinal
Número de filhos	Z	Quantitativa Discreta
Salário	S	Quantitativa Contínua
Idade	U	Quantitativa Discreta
Região da procedência	V	Qualitativa Nominal

Exemplo

Arquivo Editar Exibir Inserir Formatar Planilha Dados Ferramentas Janela Ajuda							
Arial 10 a a a a							
K10 $f(x)$ Σ =							
	A	B	C	D	E	H	I
2	N	Estado Civil	Grau de Instrução	N de Filhos	Salário (x Sal Min)	Região de Procedência	Idade
3	1	solteiro	ensino fundamental	0	4,00	interior	26,25
4	2	casado	ensino fundamental	1	4,56	capital	32,83
5	3	casado	ensino fundamental	2	5,25	capital	36,42
6	4	solteiro	ensino médio	0	5,73	outra	20,83
7	5	solteiro	ensino fundamental	0	6,26	outra	40,58
8	6	casado	ensino fundamental	0	6,66	interior	28,00
9	7	solteiro	ensino fundamental	0	6,86	interior	41,00
10	8	solteiro	ensino fundamental	0	7,39	capital	43,33
11	9	casado	ensino médio	1	7,59	capital	34,83
12	10	solteiro	ensino médio	0	7,44	outra	23,50
13	11	casado	ensino médio	2	8,12	interior	33,50
14	12	solteiro	ensino fundamental	0	8,46	capital	27,92
15	13	solteiro	ensino médio	0	8,74	outra	37,42
16	14	casado	ensino fundamental	3	8,95	outra	44,17
17	15	casado	ensino médio	0	9,13	interior	30,42
18	16	solteiro	ensino médio	0	9,35	outra	38,67
19	17	casado	ensino médio	1	9,77	capital	31,58
20	18	casado	ensino fundamental	2	9,80	outra	39,58
21	19	solteiro	superior	0	10,53	interior	25,67
22	20	solteiro	ensino médio	0	10,76	interior	37,33
23	21	casado	ensino médio	1	11,06	outra	30,75
24	22	solteiro	ensino médio	0	11,59	capital	34,17
25	23	solteiro	ensino fundamental	0	12,00	outra	41,00
26	24	casado	superior	0	12,79	outra	26,08
27	25	casado	ensino médio	2	13,23	interior	32,42
28	26	casado	ensino médio	2	13,60	outra	35,00
29	27	solteiro	ensino fundamental	0	13,85	outra	46,58

Tabela de distribuição de frequência

Contagem do número de vezes que um valor possível aparece na amostra.

Variável	Frequência	Frequência Relativa	Porcentagem
categoria 1	n_1	$f_1 = \frac{n_1}{n}$	$100 \cdot f_1 \%$
categoria 2	n_2	$f_2 = \frac{n_2}{n}$	$100 \cdot f_2 \%$
\vdots	\vdots	\vdots	\vdots
categoria k	n_k	$f_k = \frac{n_k}{n}$	$100 \cdot f_k \%$
Total	$n = n_1 + n_2 + \cdots + n_k$	1	100%

n_i Número de elementos da população com variável igual a categoria $i, i = 1, 2, \dots, k$;

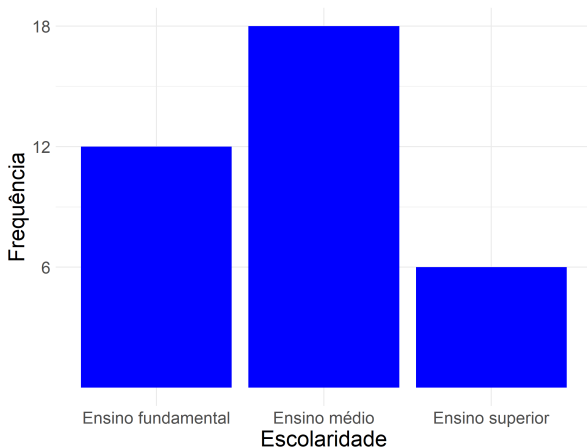
f_i Proporção ou frequência relativa calculada por $f_i = \frac{n_i}{n}, i = 1, 2, \dots, k$, em que $n = n_1 + n_2 + \cdots + n_k$.

Exemplo

Variável Escolaridade (Y).

Y	Frequência	Frequência Relativa	Porcentagem
Ensino Fundamental	12	0,3333	33,33%
Ensino Médio	18	0,5	50%
Ensino Superior	6	0,1667	16,67%
Total	36	1	100%

Exemplo – continuação



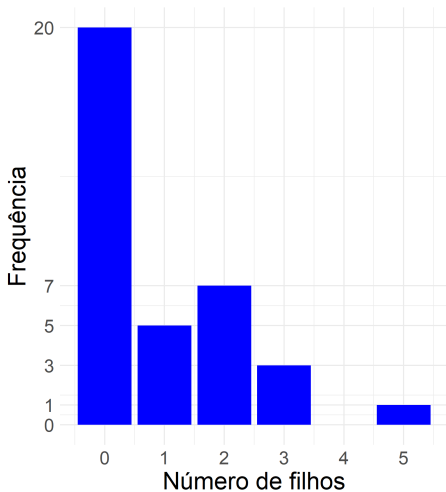
Interpretação: O grau de escolaridade mais presente entre os funcionários é “Ensino Médio”, podemos inclusive dizer que esse atributo é maioria entre os funcionários da companhia MB.

Exemplo

Número de filhos (Z).

Z	Frequência	Frequência relativa	Porcentagem
0	20	0,5556	55,56%
1	5	0,1386	13,86%
2	7	0,1944	19,44%
3	3	0,0833	8,33%
4	0	0,00	0,00%
5	1	0,0278	2,78%
Total	36	1	100%

Exemplo – continuação



Interpretação: A maioria dos funcionários tem até 2 filhos.

Exemplo

Salário (S).

S	Frequência	Frequência relativa	Porcentagem
[4, 8)	10	0,2778	27,78%
[8, 12)	12	0,3333	33,33%
[12, 16)	8	0,2222	22,22%
[16, 20)	5	0,1389	13,89%
[20, 24]	1	0,0278	2,78%
Total	36	1	100%

Exemplo – continuação

Simplificação

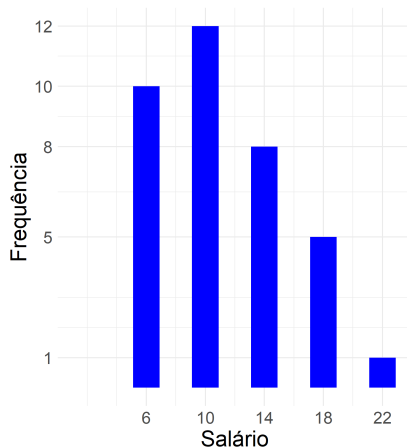
Representamos cada faixa da variável contínua pelo seu “ponto médio”.

Salário	Ponto Médio	Frequência
[4, 8)	6	10
[8, 12)	10	12
[12, 16)	14	8
[16, 20)	18	5
[20, 24]	22	1

Observação

Em nosso exemplo, estamos afirmando que todos os funcionários da faixa salarial [4, 8) ganham aproximadamente 6 salários mínimos.

Exemplo – continuação



Interpretação: Notamos que o salário mais frequente é dez salários mínimos.

Definição

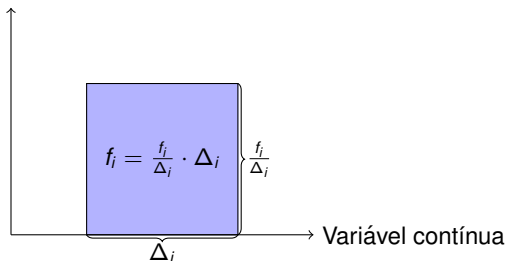
Gráficos de barras contíguas, com as bases proporcionais a largura das classes/faixas e a área de cada retângulo proporcional à respectiva frequência relativa.

Notação

Δ_i : amplitude da i -ésima barra;

$\frac{f_i}{\Delta_i}$: altura da barra chamada de densidade de frequência.

Densidade de frequência

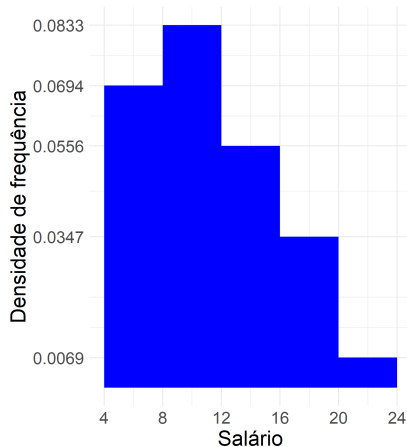


Exemplo

Salário (S).

Salário	Amplitude (Δ_i)	Frequência (f_i)	Densidade de frequência ($\frac{f_i}{\Delta_i}$)
[4, 8)	4	0,2778	$\frac{0,2778}{4} = 0,0694$
[8, 12)	4	0,3333	$\frac{0,3333}{4} = 0,0833$
[12, 16)	4	0,2222	$\frac{0,2222}{4} = 0,0556$
[16, 20)	4	0,1389	$\frac{0,1389}{4} = 0,0347$
[20, 24]	4	0,0278	$\frac{0,0278}{4} = 0,0069$

Exemplo – continuação



Interpretação: A maioria dos funcionários ganham entre 4 e 12 salários mínimos.