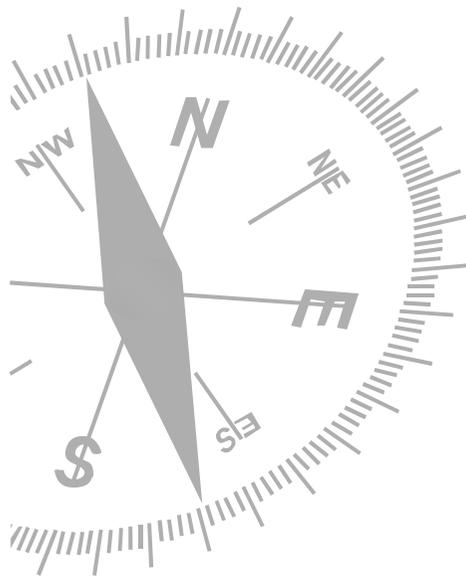


UNIVERSIDADE FEDERAL
DA PARAÍBA

DISTRIBUIÇÃO DE FREQUÊNCIA



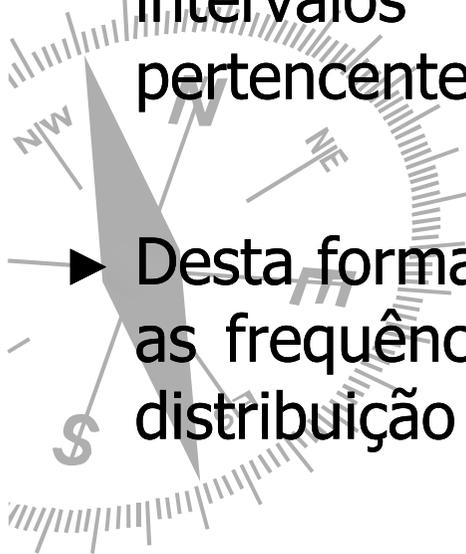
Departamento de Estatística

Luiz Medeiros

<http://www.de.ufpb.br/~luiz/>

DISTRIBUIÇÃO DE FREQUÊNCIA

- ▶ Quando se estuda uma massa de dados é de frequente interesse resumir as informações de variáveis.
- ▶ Costuma-se, frequentemente, para uma melhor compreensão dos mesmos, distribuí-los em classes ou intervalos determinando-se o número de indivíduos pertencentes a cada classe ou intervalo.
- ▶ Desta forma, um arranjo tabular dos dados, juntamente com as frequências correspondentes aos mesmos é denominado distribuição de frequência ou tabela de frequência.



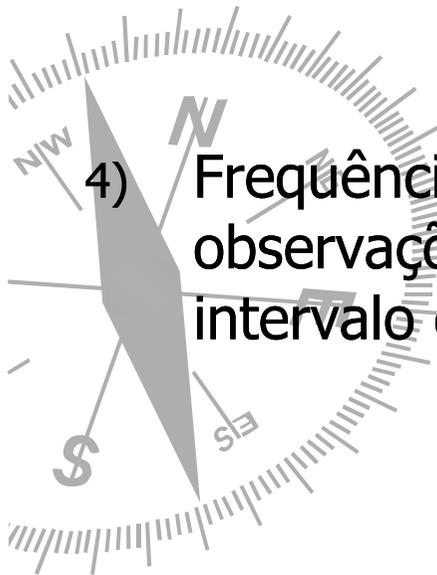
Definições

- 1) Dados Brutos: Conjunto de dados que ainda não foram numericamente organizados, obtidos após a crítica dos valores.
- 2) Rol: É um arranjo dos dados brutos em ordem crescente.
- 3) Amplitude Total (AT): É a diferença entre o maior e o menor valor observado.

$$AT = X_{MÁX} - X_{MÍN}$$

- 4) Frequência absoluta ou simples (F_i): É o número de observações que se encontra presente em uma classe ou intervalo específico.

$$\sum F_i = n$$



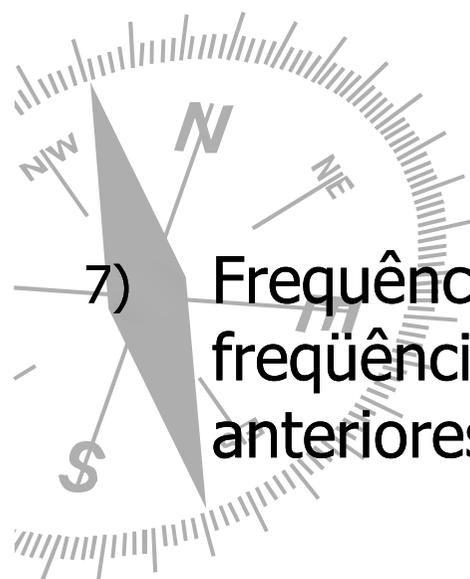
- 5) Frequência percentual ($f_i\%$): Representa o percentual de um certo valor na amostra.

$$f_i\% = (F_i / n) \times 100$$

$$\sum f_i\% = 100$$

- 6) Frequência simples acumulada (Fac_i): É a soma das frequências simples das classes ou dos valores anteriores.

$$Fac_i = F_1 + F_2 + \dots + F_i$$



- 7) Frequência percentual acumulada ($fac_i\%$): É a soma das frequências relativas percentual das classes ou dos valores anteriores.

$$fac_i\% = f_1\% + f_2\% + \dots + f_i\%$$

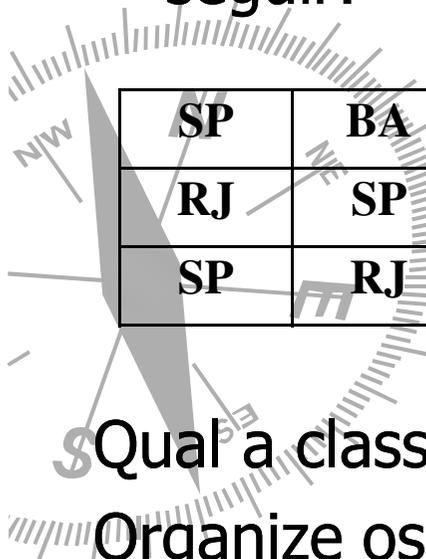
DISTRIBUIÇÃO DE FREQUÊNCIA POR VALOR

- ▶ Utilizada para dados de uma variável qualitativa ou discreta.
- ▶ Considera-se os diferentes valores das observações ou categorias e o número de vezes (frequência) que cada valor aparece nos dados.



Exemplo

- ▶ De acordo com a Secretaria de Turismo da Paraíba, os quatro estados com maior participação no número de turistas que desembarcaram no aeroporto de João Pessoa em 2009 foram: São Paulo (SP), Rio de Janeiro (RJ), Bahia (BA) e Paraná (PR). Os dados de uma amostra de 30 turistas abordados no aeroporto são apresentados a seguir:



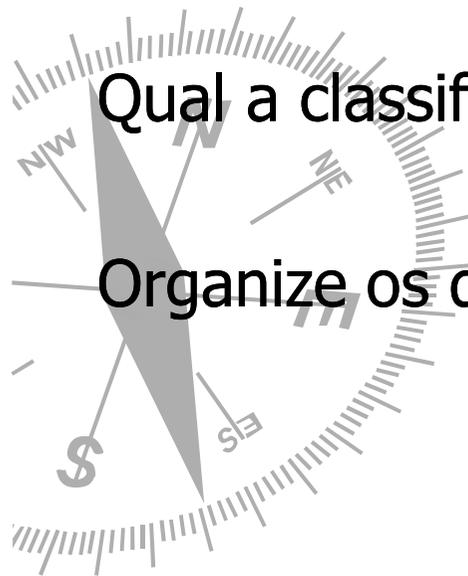
SP	BA	BA	SP	RJ	BA	SP	SP	PR	PR
RJ	SP	SP	RJ	PR	RJ	SP	RJ	SP	BA
SP	RJ	PR	SP	BA	SP	PR	BA	SP	SP

Qual a classificação dessa variável?

Organize os dados em uma tabela com F_i , $f_i\%$.

Exemplo

- ▶ O C.A. de engenharia resolveu realizar um levantamento do número de alunos reprovados, em pelo menos uma disciplina, no 1º período do curso nos últimos semestres. Os valores encontrados foram: 22 25 23 22 23 26 25 29 23 30 24 24 22 24 22 24 25 22 28 23 24 24 24 24 24 24 25 25 25 30 25 28 29 24 26 30

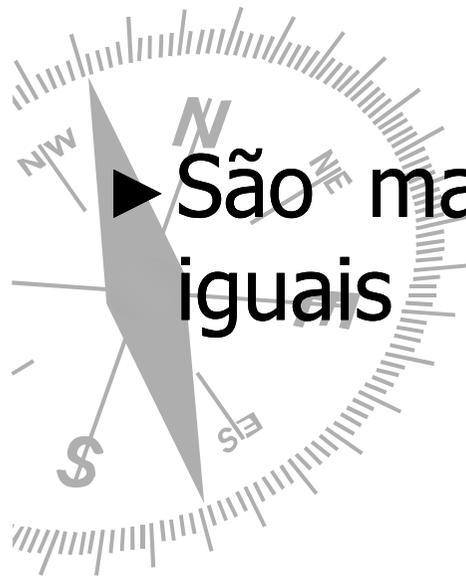


Qual a classificação dessa variável?

Organize os dados em uma tabela com F_i , $f_i\%$, F_{ac_i} e $f_{ac_i}\%$.

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO OU CLASSES

- ▶ Os dados são agrupados em classes
- ▶ Considera-se a frequência em cada classe (nº de observações pertencentes)



- ▶ São mais comuns tabelas com intervalos iguais

Classes	% da população	Renda Média Mensal Familiar (R\$ / pessoa)	
		de	até
Classe A 1	1%	9.733,47	acima
Classe A 2	4%	6.563,73	9.733,47
Classe B 1	9%	3.479,36	6.563,73
Classe B 2	15%	2.012,67	3.479,36
Classe C 1	21%	1.194,53	2.012,67
Classe C 2	22%	726,26	1.194,53
Classe D	25%	484,97	726,26
Classe E	3%	276,7	484,97
Classe Média (renda aprox.)		1.500,00	2.600,00

Para a construção de uma distribuição de frequência por intervalo é necessário algumas definições adicionais.

- 8) Número de classes (k): Representa o total de classes da variável. Não existe uma fórmula exata para o cálculo do número de classes. Seja n o tamanho da amostra, temos duas alternativas

1- $K \approx \text{raiz}(n)$

2- Fórmula de Sturges:

$$k = 1 + 3,3 \times \log_{10}(n)$$

Obs1: Alguns autores sugerem que a distribuição de frequência deve ter entre 5 e 20 classes.

Obs2: Arredondar o número de classes para o inteiro mais próximo.

- 9) Amplitude das classes (h): É a divisão entre a amplitude total e o número de classes. O valor só poderá ser arredondado para mais.

$$h = AT / k$$

Obs: Não se pode arredondar a amplitude das classes (h) para menos.

10) Limites das Classes: São os números extremos das classes. Utilizaremos a seguinte representação:

$LI_i \text{ |— } LS_i$ (Compreende todos os valores de LI_i a LS_i , excluindo LS_i)

1ª Classe:

Limite Inferior: $LI_1 = X_{\min}$

Limite Superior: $LS_1 = LI_1 + h$

2ª Classe:

Limite Inferior: $LI_2 = LS_1$

Limite Superior: $LS_2 = LI_2 + h$

⋮

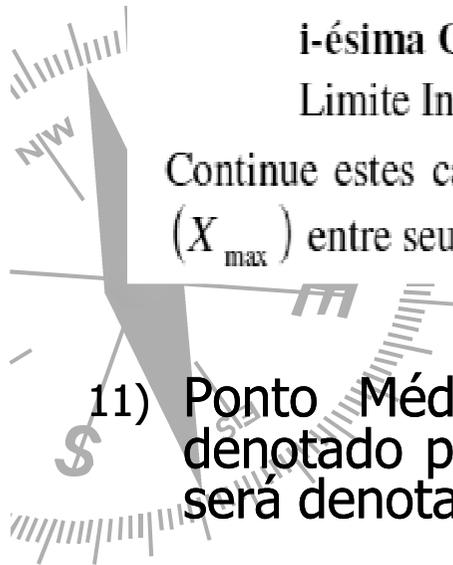
i-ésima Classe:

Limite Inferior: $LI_i = LS_{i-1}$

Limite Superior: $LS_i = LI_i + h$

Continue estes cálculos até que seja obtido um intervalo que contenha o maior valor da amostra (X_{\max}) entre seus limites.

11) Ponto Médio (X_i): É a média aritmética entre o limite superior, denotado por LS_i , e o limite inferior, denotado por LI_i . O ponto médio será denotado por X_i , onde $X_i = (LI_i + LS_i)/2$



Exemplo : Os dados abaixo representam o Preço (R\$) do produto A vendido em 25 diferentes estabelecimentos. A partir da distribuição de frequência que iremos construir podemos observar diversas características deste conjunto de dados como, por exemplo, identificar qual a faixa de preços com maior frequência.

Dados Brutos: 20,5 - 19,5 - 15,6 - 24,1 - 9,9 - 15,4 - 12,7 - 5,4 - 17,0 - 28,6 - 16,9 - 7,8 - 23,3 - 11,8 - 18,4 - 13,4 - 14,3 - 19,2 - 9,2 - 16,8 - 8,8 - 22,1 - 20,8 - 12,6 - 15,9

Rol: 5,4 - 7,8 - 8,8 - 9,2 - 9,4 - 9,9 - 11,8 - 12,6 - 12,7 - 13,4 - 14,3 - 15,4 - 15,6 - 15,9 - 16,8 - 16,9 - 17,0 - 18,4 - 19,2 - 19,5 - 20,5 - 20,8 - 22,1 - 24,1 - 28,6

Passos:

1. Ordenar as informações
2. Variável aleatória Preço é de natureza contínua implicando numa distribuição intervalar (Por classe).
3. Calcular a amplitude total.
4. Calcular o número de intervalos.
5. Calcular a amplitude do intervalo.

Exemplo

- ▶ Os dados abaixo representam o valor da hora de trabalho de 30 profissionais de Administração na cidade de João Pessoa.

20,5	9,5	15,3	17,2	24,1	19,9
15,4	12,7	7,4	16,5	15,3	26,2
14,9	7,8	23,3	15,9	11,8	18,4
13,4	14,3	16,2	16,7	9,2	16,8
9,8	20,1	17,8	17,1	12,6	15,9



Qual a classificação dessa variável?

Organize os dados em uma tabela completa.

- ▶ A distribuição de frequência abaixo refere-se às notas da turma 03 de Cálculo das probabilidades e estatística I.

NOTAS	F_i
0 ---- 2	4
2 ---- 4	8
4 ---- 6	12
6 ---- 8	8
8 ---- 10	4
TOTAL	36

- ▶ Considerando nota mínima para aprovação igual a 6. Qual foi o percentual de aprovados?
- ▶ A professora considera o desempenho de uma turma satisfatório se pelo menos 60% das notas estiverem distribuídas entre 4 (incluso) e 8 (excluso). A turma teve um desempenho satisfatório? Justifique sua resposta.

