

Bioestatística

Isolde Previdelli
itsprevidelli@uem.br
isoldeprevidelli@gmail.com

07 de Março 2017



APRESENTAÇÃO DOS DADOS

Análise descritiva

Tipos de variáveis

Tabelas de frequências

Tabelas de frequências

Gráficos

Tabelas de frequências

Gráficos

Medidas de resumo numérico



Análise descritiva

- ▶ Análise descritiva é primeiro contato do pesquisador com os dados
- ▶ Principal objetivo é explorar, resumir e sumarizar o comportamento dos dados.
- ▶ Tabelas de frequências, gráficos e medidas de resumo numérico.



Variável qualitativa

 **Nominal** - São categorias que não podem ser ordenadas.

Exemplo: Sexo, condição de saúde de um paciente, bairro, etc

 **Ordinal** - São categorias que seguem uma ordem natural.

Exemplo: Grau de instrução, intensidade de maratonista, gravidade de uma doença, etc.



Variável quantitativa ●

-  **Discreta** - Seu campo de variação é um conjunto finito ou infinito enumerável, em geral são contagens.
Exemplo: Quantidade de pacientes com determinada doença, número de ovos do mosquito *aedes aegypti* em uma determinada região, etc
-  **Contínua** - Seu campo de variação é um conjunto infinito não-enumerável
Exemplo: peso, altura e pressão sistólica de pacientes, etc.



 **Rank/Posto** - Rank pode ser definido como o número que corresponde a posição de cada observação de uma amostra de valores.

Exemplo: Uma amostra com os valores 0, 1, 2, 2, 2, 4, 6.
Os ranks para esta amostra: 1,2,4,4,4,5,6

 **Porcentagem** - É a proporção de indivíduos pertencentes a uma categoria e o número total de indivíduos considerados na amostra simples.

Exemplo: número de óbitos por acidentes de trânsito por faixa etária em um região dividido pelo total de óbitos por acidentes de trânsito.



-  **Taxas** - número de casos de um resultado de interesse que ocorre em um determinado período de tempo dividido pelo tamanho da população nesse período.
Exemplo: Razão entre o número total de óbitos de uma determinada região em um determinado período de tempo dividido pela população total desta região no mesmo período de tempo.
-  **Escores** - é uma pontuação dada aos indivíduos de uma amostra para mensurar determinado risco ou chance que ocorrer um determinado evento de interesse.
Exemplo: atendimento em um pronto socorro que após algumas informações serem coletadas, cada indivíduo recebe um escore e desta forma seu atendimento é classificado como urgente, pouco urgente ou sem urgência.



 **Variáveis latentes** - São variáveis não observáveis ou que não podem ser medidas diretamente. Desta forma são criadas escalas para poder mensurá-las.

Exemplo: Variáveis que representam sentimentos como felicidade ou depressão que podem ser medidas através de instrumentos analíticos psicológicos.



Tabela de frequência para variável raça

São ideais para resumir variáveis qualitativas e quantitativas discreta desde que esta última tenha um conjunto pequeno de valores diferentes.

Raça	Freq. Abs. (n_i)	Freq. Rel. (f_i)
Branca	96	0,508
Afro-americana	26	0,138
Outras	67	0,354
Total(n)	189	1

Tabela: SHAHBABA B., pág.23

n_i : frequência absoluta da classe

n : total da amostra

$f_i = \frac{n_i}{n}$, frequência relativa da classe

TABELAS

Exemplo

9



29

Tabela de frequência para uma variável quantitativa

Considere um experimento em que foi investigado em uma amostra de vinte e um pacientes de uma clínica médica, se o aumento da ingestão de cálcio aumenta a pressão arterial:

Nível	Freq. Abs. (n_i)	Freq. Rel. (f_i)
95-100	3	0,14
100-105	4	0,19
105-110	1	0,05
110-115	6	0,29
115-120	3	0,14
120-125	3	0,14
125-130	0	0,00
130-135	1	0,05

Tabela: http://www.ics.uci.edu/~babaks/BWR/Home_files/calcium.txt

HISTOGRAMA

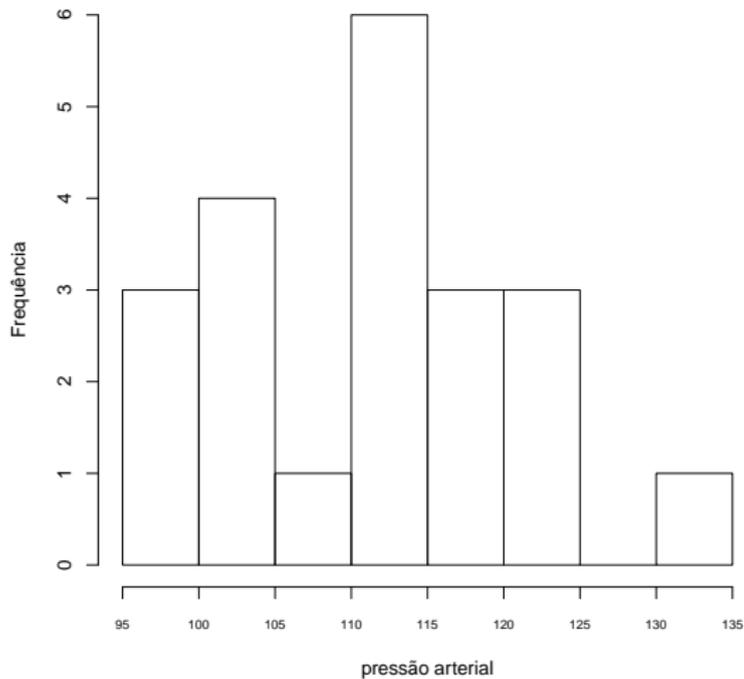
Exemplo

10



29

Histograma do exemplo anterior



TABELAS

Exemplo

11



29

Tabela de frequência para uma variável quantitativa

E se juntarmos as três últimas classes? Como ficaria a tabela de frequência?

Nível	Freq. Abs. (n_i)	Freq. Rel. (f_i)
95-100	3	0,14
100-105	4	0,19
105-110	1	0,05
110-115	6	0,29
115-120	3	0,14
120-135	4	0,19

Tabela: http://www.ics.uci.edu/~babaks/BWR/Home_files/calcium.txt

HISTOGRAMA

Exemplo

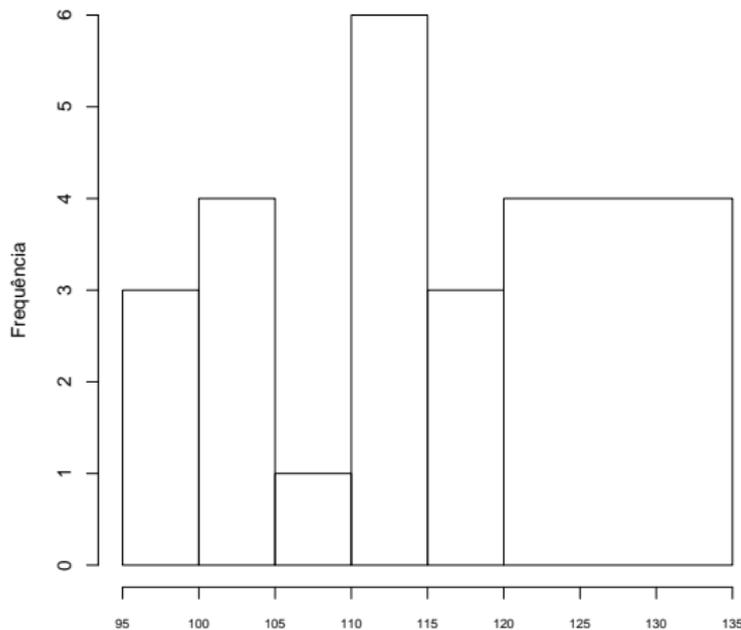
12



29

E o novo Histograma?

Histograma do exemplo anterior





Medidas de resumo numérico

-  Tendência central
-  Dispersão
-  Separatrizes
-  Assimetria e curtose



Média aritmética

 Indica o valor em torno do qual há um equilíbrio na distribuição dos dados:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

obs: $\bar{x} + K = \bar{x} + K$

obs: $\bar{x} * K = \bar{x} * K$



Mediana e Moda

-  **Mediana:** É valor central em um conjunto de dados ordenado:
-  **Moda:** É o valor mais frequente em um conjunto de dados



Amplitude total ●

 É a diferença entre o valor máximo e mínimo

$$\Delta = X_{(max.)} - X_{(min.)}$$



Desvio médio absoluto

É o cálculo da média dos desvios absolutos.

$$dma = \sum_{i=1}^n \frac{|x_i - \bar{x}_{obs}|}{n}$$



Variância •

 Quantifica a variabilidade dos dados em torno da média.

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$$

obs: $s^2 + K = s^2$

obs: $s^2 * K = s^2 * K^2$



Desvio Padrão

 Quantifica a variabilidade dos dados na mesma escala em que eles foram medidos.

$$s = \sqrt{\text{Var}(x)} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}}$$



Coefficiente de variação •

 É uma medida relativa que avalia o percentual de variabilidade em relação a média observada.

$$cv = 100 \frac{s}{\bar{x}}$$



SHAHBABA B., pág.33

- ⚡ Considere a seguinte amostra: $x_1 = (74, 80, 79, 85, 81)$, a média é: $\bar{x}_1 = 79,8$
- ⚡ Se trocarmos o valor 74 por 47, $x_2 = (47, 80, 79, 85, 81)$? O que acontece?
- ⚡ A média muda para $\bar{x}_2 = 74,4$.
- ⚡ Esse novo valor resume bem os dados?
- ⚡ A mediana de x_1 e x_2 é igual a 80, ou seja, é uma medida mais robusta a *outliers*.



SHAHBABA B., pág.35

- Considerare agora duas amostras sobre medidas da pressão do sangue (mmHg) para dois tipos de pacientes:
- Paciente A, $x = (95, 98, 96, 95, 96)$
- Paciente B, $y = (85, 106, 88, 105, 96)$
- As médias são iguais: $\bar{x} = \bar{y} = 96$
- As medianas também são iguais: $\tilde{x} = \tilde{y} = 96$


SHAHBABA B., pág.35

Paciente A			Paciente B		
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
95	-1	1	85	-11	121
98	2	4	106	10	100
96	0	0	88	-8	64
95	-1	1	105	9	81
96	0	0	96	0	0
Σ	0	6	Σ	0	367

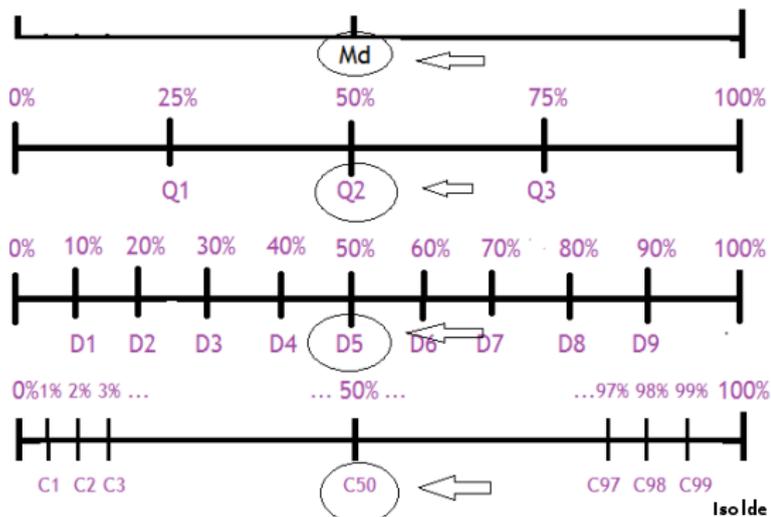
$$s_x^2 = \frac{6}{4} = 1,5 \quad s_y^2 = \frac{367}{4} = 91,5$$

$$s_x = \sqrt{1,5} = 1,22 \quad s_y = \sqrt{91,5} = 9,56$$



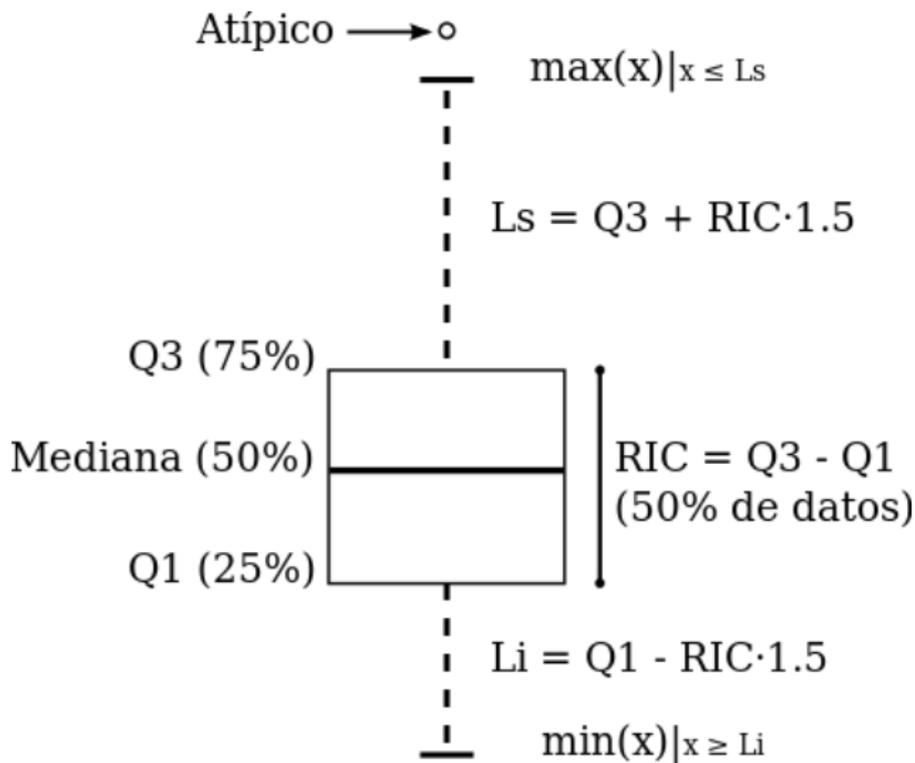
Percentis, Decis e Quartis

 São medidas que dividem os dados em um número de partes iguais.



Separatrizes

Boxplot





Assimetria

Analisa a proximidade ou o afastamento de um conjunto de dados em relação à distribuição Gaussiana.

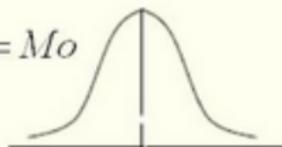
$$As = \frac{\mu_3}{\sigma^3}$$

- Se $As = 0$, distribuição é simétrica
- Se $As > 0$, distribuição assimétrica a direita (positiva)
- Se $As < 0$, distribuição assimétrica a esquerda (negativa)



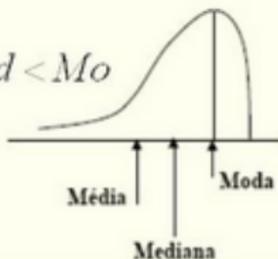
Distr. Simétrica

$$\bar{X} = Med = Mo$$



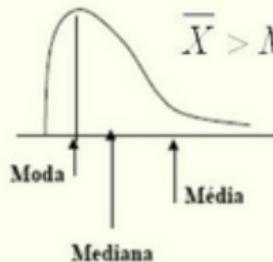
Assimetria negativa

$$\bar{X} < Med < Mo$$



Assimetria positiva

$$\bar{X} > Med > Mo$$





Assimetria

caracterizar a forma da distribuição quanto ao seu achatamento.

$$k = \frac{\mu_4}{\sigma^4}$$

 $k < 0$, distribuição Platicúrtica

 $k = 0$, distribuição Mesocúrtica

 $k > 0$, distribuição Leptocúrtica



Exemplos curtose

