

# Probabilidade e estatística - Aula 13

## Estatística descritiva - Apresentação de dados

Dr. Giannini Italino Alves Vieira

Universidade Federal do Ceará - Campus de Crateús

2024

1 Diagrama de Ramo e folhas

2 Distribuição de frequências

3 Histograma

4 Diagrama de caixas

# Estatística descritiva - Apresentação de dados

- Na aula passada fizemos um estudos sobre algumas medidas úteis para resumir dados, tais como: média da amostra, variância da amostra, mediana, amplitude, moda e os quartis.
- Na aula de hoje veremos alguns recursos para podermos fazer apresentações de dados, tais como: diagrama de ramo e folhas, distribuição de frequências, histograma e o diagrama de caixa.
- Resumos e apresentações de dados, quando bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

# Estatística descritiva - Apresentação de dados

- Na aula passada fizemos um estudos sobre algumas medidas úteis para resumir dados, tais como: média da amostra, variância da amostra, mediana, amplitude, moda e os quartis.
- Na aula de hoje veremos alguns recursos para podermos fazer apresentações de dados, tais como: diagrama de ramo e folhas, distribuição de frequências, histograma e o diagrama de caixa.
- Resumos e apresentações de dados, quando bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

# Estatística descritiva - Apresentação de dados

- Na aula passada fizemos um estudos sobre algumas medidas úteis para resumir dados, tais como: média da amostra, variância da amostra, mediana, amplitude, moda e os quartis.
- Na aula de hoje veremos alguns recursos para podermos fazer apresentações de dados, tais como: diagrama de ramo e folhas, distribuição de frequências, histograma e o diagrama de caixa.
- Resumos e apresentações de dados, quando bem constituídos, são fundamentais para fazermos uma boa análise estatística, uma vez que, a partir destes, podemos extrair características importantes dos dados e também ter noção sobre o tipo de modelo que é mais apropriado na solução do problema.

# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.

# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.

# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.



# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.

# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.

# Diagrama de Ramo e folhas

- Um diagrama de ramo e folhas é um bom recurso para se obter uma apresentação visual de um conjunto de dados.
- Contudo, esse recurso só pode ser utilizado em dados, digamos  $x_1, x_2, \dots, x_n$ , em que cada número  $x_i$  consiste em, no mínimo, dois dígitos.

## Construção de um diagrama de Ramo e folhas

Para construirmos um diagrama de ramo e folhas procedemos da seguintes forma:

- Divida cada número  $x_i$  em duas partes: um ramo, contendo um ou mais dígitos iniciais e uma folha, contendo os dígitos restantes;
- Liste os valores do ramo em uma coluna vertical e, ao lado de cada ramo, coloque a folha para cada observação;
- Escreva as unidades para os ramos e folhas no gráfico.

# Exemplo

Exemplo: As medições humanas fornecem uma ampla área de aplicação para métodos estatísticos. O artigo "A Longitudinal Study of the Development of Elementary School Children's Private Speech" (Merrill-Palmer Q., 1990: 443-463) relatou um estudo com crianças falando sozinhas (fala privada). Pensou-se que a fala privada seria relacionado ao QI, porque o QI deve medir a maturidade mental, e era sabido que a fala privada diminui à medida que os alunos progridem nas séries primárias. O estudo incluiu 33 alunos cujas pontuações de QI na primeira série são dados aqui:

82	96	99	102	103	103	106	107	108	108	108	108
109	110	110	111	113	113	113	113	115	115	118	118
119	121	122	122	127	132	136	140	146			

# Exemplo

Ramo	Folha	Frequência
8	2	1
9	6 9	2
10	2 3 3 6 7 8 8 8 8 9	10
11	0 0 1 3 3 3 3 5 5 8 8 9	12
12	1 2 2 7	4
13	2 6	2
14	0 6	2

Ramo: dígitos das dezenas e das centenas; Folha: dígitos das unidades

- Note que em um diagrama ordenado de ramo e folhas é bem fácil encontrar características dos dados. Por exemplo, no diagrama acima temos que

$q_2 = Md = 113$ ,  $q_1 = 108$ ,  $q_3 = 119$  108 e 113 são modas.

# Exemplo

Ramo	Folha	Frequência
8	2	1
9	6 9	2
10	2 3 3 6 7 8 8 8 8 9	10
11	0 0 1 3 3 3 3 5 5 8 8 9	12
12	1 2 2 7	4
13	2 6	2
14	0 6	2

Ramo: dígitos das dezenas e das centenas; Folha: dígitos das unidades

- Note que em um diagrama ordenada de ramo e folhas é bem fácil encontrar características dos dados. Por exemplo, no diagrama acima temos que

$q_2 = Md = 113$ ,  $q_1 = 108$ ,  $q_3 = 119$  108 e 113 são modas.

# Exemplo

Ramo	Folha	Frequência
8	2	1
9	6 9	2
10	2 3 3 6 7 8 8 8 8 9	10
11	0 0 1 3 3 3 3 5 5 8 8 9	12
12	1 2 2 7	4
13	2 6	2
14	0 6	2

Ramo: dígitos das dezenas e das centenas; Folha: dígitos das unidades

- Note que em um diagrama ordenada de ramo e folhas é bem fácil encontrar características dos dados. Por exemplo, no diagrama acima temos que

$q_2 = Md = 113$ ,  $q_1 = 108$ ,  $q_3 = 119$  108 e 113 são modas.

# Exemplo

Ramo	Folha	Frequência
8	2	1
9	6 9	2
10	2 3 3 6 7 8 8 8 8 9	10
11	0 0 1 3 3 3 3 5 5 8 8 9	12
12	1 2 2 7	4
13	2 6	2
14	0 6	2

Ramo: dígitos das dezenas e das centenas; Folha: dígitos das unidades

- Note que em um diagrama ordenada de ramo e folhas é bem fácil encontrar características dos dados. Por exemplo, no diagrama acima temos que

$q_2 = Md = 113$ ,  $q_1 = 108$ ,  $q_3 = 119$  108 e 113 são modas.



# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- ▶ Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- ▶ Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- ▶ Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- ▶ Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- ▶ Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- ▶ Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- ▶ Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- ▶ Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- ▶ Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- ▶ Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- ▶ Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- ▶ Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

# Distribuição de frequências

- Um outro recurso útil para se fazer apresentação de dados é por meio da distribuição de frequências.
- A distribuição de frequências apresenta um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas.

## Distribuição de frequências

Para construir uma distribuição de frequências procedemos da seguinte maneira:

- ▶ Dividir a faixa de dados em intervalos de classes (intervalos em que os dados assumem valores);
- ▶ Se possível, os intervalos de classe devem ter mesmo comprimento, de modo a aumentar a informação visual da distribuição de frequências;
- ▶ Especificar a frequência com que os dados assumem valores em cada intervalo de classe considerado.

- A escolha do número de intervalos depende do número de observações da amostra e da dispersão dos dados. Contudo, em muitos casos, trabalha-se bem se o número de intervalos de classe considerado for aproximado de  $\sqrt{n}$ , em que  $n$  é o número de observações da amostra;

## Exemplo

- Considere novamente os dados do exemplo anterior.
- Temos que uma distribuição de frequências para esses dados é dada por

Int. de classe	Frequência	Frequência relativa
$80 < x \leq 90$	1	$1/33$
$90 < x \leq 100$	2	$2/33$
$100 < x \leq 110$	12	$12/33$
$110 < x \leq 120$	10	$10/33$
$120 < x \leq 130$	4	$4/33$
$130 < x \leq 140$	3	$3/33$
$140 < x \leq 150$	1	$1/33$

Distribuição de frequências para os dados do exemplo 1.

- A escolha do número de intervalos depende do número de observações da amostra e da dispersão dos dados. Contudo, em muitos casos, trabalha-se bem se o número de intervalos de classe considerado for aproximado de  $\sqrt{n}$ , em que  $n$  é o número de observações da amostra;

## Exemplo

- Considere novamente os dados do exemplo anterior.
- Temos que uma distribuição de frequências para esses dados é dada por

Int. de classe	Frequência	Frequência relativa
$80 < x \leq 90$	1	$1/33$
$90 < x \leq 100$	2	$2/33$
$100 < x \leq 110$	12	$12/33$
$110 < x \leq 120$	10	$10/33$
$120 < x \leq 130$	4	$4/33$
$130 < x \leq 140$	3	$3/33$
$140 < x \leq 150$	1	$1/33$

Distribuição de frequências para os dados do exemplo 1.



- A escolha do número de intervalos depende do número de observações da amostra e da dispersão dos dados. Contudo, em muitos casos, trabalha-se bem se o número de intervalos de classe considerado for aproximado de  $\sqrt{n}$ , em que  $n$  é o número de observações da amostra;

## Exemplo

- Considere novamente os dados do exemplo anterior.
- Temos que uma distribuição de frequências para esses dados é dada por

Int. de classe	Frequência	Frequência relativa
$80 < x \leq 90$	1	$1/33$
$90 < x \leq 100$	2	$2/33$
$100 < x \leq 110$	12	$12/33$
$110 < x \leq 120$	10	$10/33$
$120 < x \leq 130$	4	$4/33$
$130 < x \leq 140$	3	$3/33$
$140 < x \leq 150$	1	$1/33$

Distribuição de frequências para os dados do exemplo 1.

- A escolha do número de intervalos depende do número de observações da amostra e da dispersão dos dados. Contudo, em muitos casos, trabalha-se bem se o número de intervalos de classe considerado for aproximado de  $\sqrt{n}$ , em que  $n$  é o número de observações da amostra;

## Exemplo

- Considere novamente os dados do exemplo anterior.
- Temos que uma distribuição de frequências para esses dados é dada por

Int. de classe	Frequência	Frequência relativa
$80 < x \leq 90$	1	$1/33$
$90 < x \leq 100$	2	$2/33$
$100 < x \leq 110$	12	$12/33$
$110 < x \leq 120$	10	$10/33$
$120 < x \leq 130$	4	$4/33$
$130 < x \leq 140$	3	$3/33$
$140 < x \leq 150$	1	$1/33$

Distribuição de frequências para os dados do exemplo 1.

# Exemplo

- Na distribuição de frequências, podemos ainda apresentar a frequência relativa, que é calculada como sendo a razão entre a frequência observada em cada intervalo de classe sobre o número total de observações na amostra.
- Note que quando usamos a distribuição de frequências para apresentar dados então temos algumas "perdas" de informações. Por exemplo, se quiséssemos saber qual o valor foi mais frequente na amostra então não conseguimos obter essa informação a partir da distribuição de frequências.

# Exemplo

- Na distribuição de frequências, podemos ainda apresentar a frequência relativa, que é calculada como sendo a razão entre a frequência observada em cada intervalo de classe sobre o número total de observações na amostra.
- Note que quando usamos a distribuição de frequências para apresentar dados então temos algumas "perdas" de informações. Por exemplo, se quiséssemos saber qual o valor foi mais frequente na amostra então não conseguimos obter essa informação a partir da distribuição de frequências.

# Histograma

- Um histograma é um recurso para se dispor, visualmente, a distribuição de frequências dos dados.

Para construir um histograma procedemos da seguinte forma:

- ▶ Marque os limites dos intervalos de classe em um eixo horizontal;
- ▶ Nomeie o eixo vertical com as frequências ou frequências relativas;
- ▶ Acima de cada intervalo de classe, desenhe um retângulo em que sua altura seja igual a frequência (ou frequência relativa) correspondente àquele intervalo de classe.

- Por exemplo, se considerarmos a distribuição de frequências do exemplo anterior, temos que o histograma é

# Histograma

- Um histograma é um recurso para se dispor, visualmente, a distribuição de frequências dos dados.

Para construir um histograma procedemos da seguinte forma:

- ▶ Marque os limites dos intervalos de classe em um eixo horizontal;
  - ▶ Nomeie o eixo vertical com as frequências ou frequências relativas;
  - ▶ Acima de cada intervalo de classe, desenhe um retângulo em que sua altura seja igual a frequência (ou frequência relativa) correspondente àquele intervalo de classe.
- 
- Por exemplo, se considerarmos a distribuição de frequências do exemplo anterior, temos que o histograma é

# Histograma

- Um histograma é um recurso para se dispor, visualmente, a distribuição de frequências dos dados.

Para construir um histograma procedemos da seguinte forma:

- ▶ Marque os limites dos intervalos de classe em um eixo horizontal;
  - ▶ Nomeie o eixo vertical com as frequências ou frequências relativas;
  - ▶ Acima de cada intervalo de classe, desenhe um retângulo em que sua altura seja igual a frequência (ou frequência relativa) correspondente àquele intervalo de classe.
- 
- Por exemplo, se considerarmos a distribuição de frequências do exemplo anterior, temos que o histograma é

# Histograma

- Um histograma é um recurso para se dispor, visualmente, a distribuição de frequências dos dados.

Para construir um histograma procedemos da seguinte forma:

- ▶ Marque os limites dos intervalos de classe em um eixo horizontal;
  - ▶ Nomeie o eixo vertical com as frequências ou frequências relativas;
  - ▶ Acima de cada intervalo de classe, desenhe um retângulo em que sua altura seja igual a frequência (ou frequência relativa) correspondente àquele intervalo de classe.
- Por exemplo, se considerarmos a distribuição de frequências do exemplo anterior, temos que o histograma é



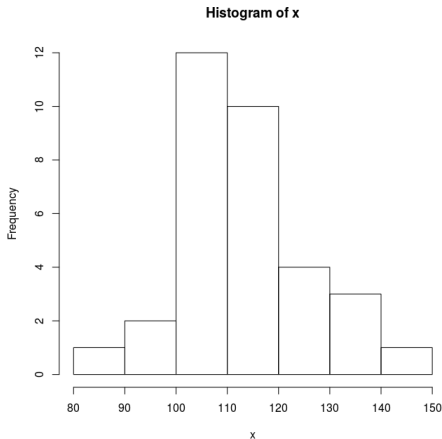
# Histograma

- Um histograma é um recurso para se dispor, visualmente, a distribuição de frequências dos dados.

Para construir um histograma procedemos da seguinte forma:

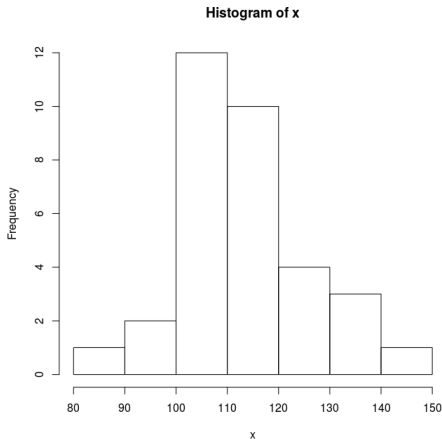
- ▶ Marque os limites dos intervalos de classe em um eixo horizontal;
  - ▶ Nomeie o eixo vertical com as frequências ou frequências relativas;
  - ▶ Acima de cada intervalo de classe, desenhe um retângulo em que sua altura seja igual a frequência (ou frequência relativa) correspondente àquele intervalo de classe.
- 
- Por exemplo, se considerarmos a distribuição de frequências do exemplo anterior, temos que o histograma é

# Histograma dos dados do problema anterior



Note que o histograma, assim como o diagrama de ramo e folhas, fornece uma impressão visual da forma da distribuição dos dados, assim como informações sobre tendência central e o espalhamento ou dispersão dos dados.

# Histograma dos dados do problema anterior



Note que o histograma, assim como o diagrama de ramo e folhas, fornece uma impressão visual da forma da distribuição dos dados, assim como informações sobre tendência central e o espalhamento ou dispersão dos dados.

# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é uma medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é uma medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é um medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é um medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é uma medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.



# Diagrama de caixas - Box plots

- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é uma medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Diagrama de caixas - Box plots

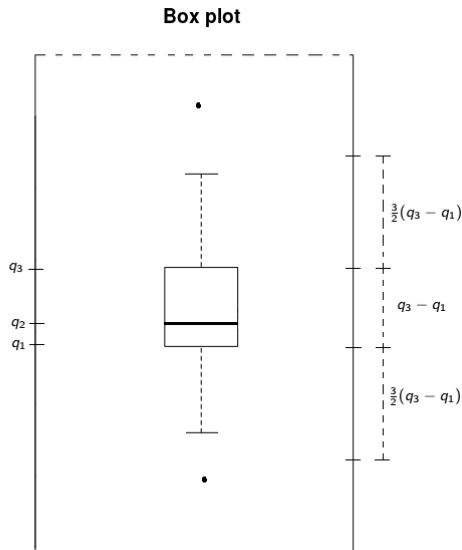
- O diagrama de ramo e folhas e o histograma são bons recursos de impressões visuais de dados, na medida que fornecem informações sobre características dos dados.
- Veremos agora um outro importante recurso, de apresentação gráfica dos dados, que descreve simultaneamente várias características importantes dos dados, tais como: centro, dispersão, simetria e identificações de observações atípicas (outliers).

## Diagrama de caixas

Para construir um diagrama de caixas procedemos da seguinte forma:

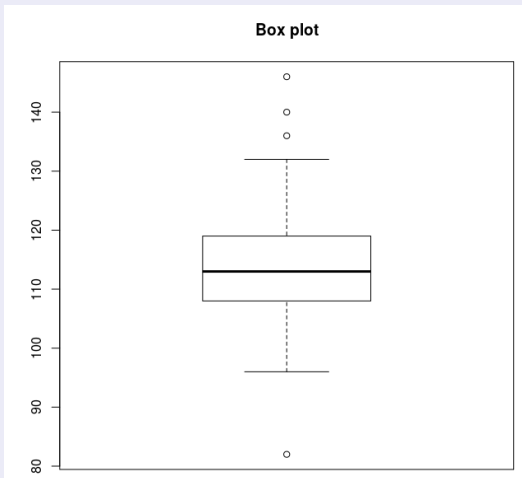
- Crie um retângulo em que estão representadas a mediana e os quartis;
- A partir do retângulo, para cima, segue uma linha até o maior ponto amostral que não exceda  $L_s = q_3 + \frac{3}{2}(q_3 - q_1)$  ( $q_3 - q_1$  é chamado de distância interquartil e é um medida de dispersão alternativa ao desvio padrão);
- De modo análogo, da parte inferior do retângulo, para baixo, segue uma linha até o menor valor amostral que não seja menor do que  $L_i = q_1 - \frac{3}{2}(q_3 - q_1)$
- Um ponto que estiver acima de  $L_s$  ou abaixo de  $L_i$  é chamado de outlier.

# Ilustrando a ideia do diagrama de caixas



# Exemplo

- Se considerarmos os dados do primeiro exemplo dessa aula, temos que o diagrama de caixas é o seguinte:



## Cont. do exemplo

- Na figura acima temos que  $q_1 = 108$ ,  $q_2 = \text{Md} = 113$  e  $q_3 = 119$ .
- Temos que a linha superior se estende até observação 132, uma vez que ela é a maior observação abaixo de  $L_s = q_3 + \frac{3}{2}(q_3 - q_1) = 119 + \frac{3}{2}(119 - 108) = 135.5$ .
- Similarmente, a linha inferior se estende até a observação 96, uma vez que ela é a menor observação acima de  $L_i = q_1 - \frac{3}{2}(q_3 - q_1) = 108 - \frac{3}{2}(119 - 108) = 91.5$ .
- Note que há também quatro outliers nessa amostra.

## Cont. do exemplo

- Na figura acima temos que  $q_1 = 108$ ,  $q_2 = \text{Md} = 113$  e  $q_3 = 119$ .
- Temos que a linha superior se estende até observação 132, uma vez que ela é a maior observação abaixo de  $L_s = q_3 + \frac{3}{2}(q_3 - q_1) = 119 + \frac{3}{2}(119 - 108) = 135.5$ .
- Similarmente, a linha inferior se estende até a observação 96, uma vez que ela é a menor observação acima de  $L_i = q_1 - \frac{3}{2}(q_3 - q_1) = 108 - \frac{3}{2}(119 - 108) = 91.5$ .
- Note que há também quatro outliers nessa amostra.

## Cont. do exemplo

- Na figura acima temos que  $q_1 = 108$ ,  $q_2 = \text{Md} = 113$  e  $q_3 = 119$ .
- Temos que a linha superior se estende até observação 132, uma vez que ela é a maior observação abaixo de  $L_s = q_3 + \frac{3}{2}(q_3 - q_1) = 119 + \frac{3}{2}(119 - 108) = 135.5$ .
- Similarmente, a linha inferior se estende até a observação 96, uma vez que ela é a menor observação acima de  $L_i = q_1 - \frac{3}{2}(q_3 - q_1) = 108 - \frac{3}{2}(119 - 108) = 91.5$ .
- Note que há também quatro outliers nessa amostra.

## Cont. do exemplo

- Na figura acima temos que  $q_1 = 108$ ,  $q_2 = \text{Md} = 113$  e  $q_3 = 119$ .
- Temos que a linha superior se estende até observação 132, uma vez que ela é a maior observação abaixo de  $L_s = q_3 + \frac{3}{2}(q_3 - q_1) = 119 + \frac{3}{2}(119 - 108) = 135.5$ .
- Similarmente, a linha inferior se estende até a observação 96, uma vez que ela é a menor observação acima de  $L_i = q_1 - \frac{3}{2}(q_3 - q_1) = 108 - \frac{3}{2}(119 - 108) = 91.5$ .
- Note que há também quatro outliers nessa amostra.