DISCIPLINA: Tratamento e Análise de Dados / Informações - TADI

PROF^{a}.: Karla Lima

EACH-USP

Aula 5: 06/04/2016

Elementos típicos da distribuição

Para ressaltar as características de cada distribuição precisamos introduzir conceitos que se expressem através de números, que nos permitam traduzir essas características. Esses conceitos são denominados elementos típicos da distribuição. São eles:

- Medidas de Posição;
- Medidas de variabilidade ou dispersão;
- medidas de assimetria e curtose;

As medidas de posição mais importantes são as medidas de tendência central, que recebem este nome pelo fato dos dados observados tenderem, em geral, a se agrupar em torno dos valores centrais.

Medidas de tendência central

- Média.
- Mediana.
- Moda.

Separatrizes - outras medidas de posição

- Quartis.
- Decis.
- Percentis.

Média para dados não agrupado ou dados simples

É o quociente entre a soma dos valores da variável e o número desses valores:

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

em que,

- $\bar{x}$ é a média;
- x_i é a i-ésima observação;
- n é o número total de observações.

Média para dados não agrupado ou dados simples

Exemplo: Sabendo-se que a produção leiteira diária de uma certa vaca, durante uma semana, foi de

$$10, 14, 13, 15, 16, 18, 12 \quad \text{litros.}$$

Qual a produção média da semana?

Média para dados agrupados sem intervalos de classe - Variável Discreta

Quando os dados estiverem agrupados sem intervalos de classe, ou seja, os dados estão apresentados na forma de uma variável discreta. Neste caso, temos

$$\bar{x}_s = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i},$$

em que,

- $\bar{x}_s$ é a média quando os dados são agrupados sem intervalo de classes;
- x_i é o valor que a i-ésima variável assume;
- f_i é a frequência da i-ésima variável;
- k é o número de categorias da variável.

Média para dados agrupados sem intervalos de classe - Variável Discreta

Exemplo: Variável número de filhos dos empregados casados.

Número de filhos	frequência
0	2
1	6
2	10
3	12
4	4

Calcule a média baseando-se na tabela acima.

Média para dados agrupados com intervalos de classe - Variável Contínua

Neste caso, temos dados agrupados em intervalos, ou seja, os dados são apresentados na forma de uma variável contínua. Temos,

$$\bar{x}_c = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i},$$

em que,

- $\bar{x}_c$ é a média quando os dados são agrupados com intervalo de classes;
- m_i é o ponto médio da i-ésima classe;
- f_i é a frequência da i-ésima variável;
- k é o número de intervalos de classe.

Média para dados agrupados com intervalos de classe - Variável Contínua

Exemplo: Variável salário.

Salários	frequência
$4 \vdash 8$	10
$8 \vdash 12$	12
$12 \vdash 16$	8
$16 \vdash 20$	5
$20 \vdash 24$	1

Calcule a média baseando-se na tabela acima.

Mediana para dados não agrupados ou dados simples

- É a realização que ocupa a posição central da série de observações, quando estão ordenadas.
- Quando os valores de uma série estão ordenados e n é o número de elementos da série, o valor mediano será:
  - Se n for ímpar: o termo de ordem
$$\frac{n+1}{2};$$
  - Se n for par: a média aritmética dos termos de ordem
$$\frac{n}{2} \text{ e } \frac{n}{2}+1.$$

Mediana para dados não agrupados ou dados simples

Exemplo (Monitoramento do chumbo no ar): A Agência de Proteção Ambiental Americana estabeleceu um padrão de qualidade do ar para o chumbo: um máximo de $1,5\mu g/m^3$. As medidas mostradas abaixo foram registradas no local do Edifício 5 do "World Trade Center", em dias diferentes, logo após a destruição causada pelos ataques terroristas de 11 de setembro de 2001. Encontre a média e mediana para essa amostra de medidas de níveis de chumbo no ar:

$$5.40, 1.10, 0.42, 0.73, 0.48, 1.10$$

- Neste caso, a média é

$$\bar{x} = \frac{5.40 + 1.10 + 0.42 + 0.73 + 0.48 + 1.10}{6} = \frac{9.23}{6} = 1.538$$

Mediana para dados não agrupados ou dados simples

- Independentemente do valor da média, é também de se notar que o conjunto de dados contém um valor (5.40) que está bem afastado dos demais.
- Nesse caso, o nível de chumbo, $5.40 \mu g/m^3$ foi medido um dia após o desmoronamento das torres e havia níveis elevados de poeira e fumaça.
- Uma desvantagem da média é que ela é sensível a qualquer valor excepcional.

Mediana para dados não agrupados ou dados simples

Para o cálculo da mediana, primeiro ordene os valores. Então

$$5.40, 1.10, 0.42, 0.73, 0.48, 1.10$$

A mediana é a média aritmética do terceiro e quarto termo, ou seja,

$$\tilde{x} = \frac{0.73 + 1.10}{2} = \frac{1.83}{2} = 0.915\ \mu g/m^3$$

- A mediana é muito diferente da média.
- A razão dessa grande diferença é o efeito que 5.40 $\mu g/m^3$ tem sobre a média.
- Se esse valor fosse reduzido para 1.20 $\mu g/m^3$ teríamos $\bar{x} = 0.838\ \mu g/m^3$ mas a mediana não se alteraria.

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

- Se os dados estão apresentados na forma de uma variável discreta, eles já estão naturalmente ordenados.
- Assim, basta verificar se o número de elementos da série é ímpar ou par e aplicar o mesmo raciocínio do caso anterior.
- Para facilitar a localização dos termos centrais, construimos a frequência acumulada da série.

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

Exemplo:

x_i	f_i
2	1
5	4
8	10
10	6
12	2
Total	23

- Temos, $\sum_{i=1}^{5} f_i = 23$ (ímpar).
- Portanto, a série admite apenas um termo central que ocupa a posição

$$\left(\frac{23+1}{2}\right)^o = 12^o$$

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

Construindo a frequência acumulada podemos localizar com facilidade o décimo segundo elemento da série.

x_i	f_i	fac_i
2	1	1
5	4	5
8	10	15
10	6	21
12	2	23

- O elemento que ocupa a primeira posição na série é 2.
- Depois aparecem 4 elementos iguais a 5. Estes elementos ocupam na série as posições de segundo a quinto.
- Depois aparecem mais dez elementos iguais a 8 que ocupam na série as posições de sexto a décimo quinto.
- Consequentemente, o elemento que ocupa a décima segunda posição vale 8. Daí, $\tilde{x}_s = 8$.

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

Interpretação: 50% dos valores da série são menores ou iguais a 8 e 50% dos valores da série são maiores ou iguais a 8.

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

Exemplo: Calcular a mediana da série.

Número de filhos	frequência
0	2
1	6
2	10
3	12
4	4
Total	34

- O número de elementos da série é 34 (par) e a série admite dois termos centrais que ocupam as posições: $(34/2)^a = 17^a$ e $(34/2+1)^a = 18^a$.
- Para localizar estes elementos vamos construir a frequência acumulada da série

Mediana para dados agrupados sem intervalos de classe - Variável Discreta

Número de filhos	frequência	frequência acumulada
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34

- As duas primeiras posições da série são ocupadas por elementos iguais a zero.
- Da terceira à oitava posição os elementos são iguais a 1.
- Da nona à décima oitava posição os elementos são iguais a 2. O elemento que ocupa a 17^a e 18^a posição é o 2. Logo a mediana é

$$\tilde{x}_s = 2.$$

Mediana para dados agrupados com intervalos de classe - Variável Contínua

- Os dados são apresentados na forma de uma variável contínua.
- Utilizaremos um exemplo, para generalizar a fórmula de cálculo da mediana.
- A fórmula da mediana para dados agrupados com intervalo de classe é dada por:

$$\tilde{x}_c = l_{m_d} + \frac{\frac{n}{2} - fac_{ant}}{f_{m_d}} h,$$

em que

- l_{m_d}: limite inferior da classe mediana;
- n: número de elementos da série;
- fac_{ant}: frequência acumulada da classe anterior à classe mediana;
- f_{m_d}: frequência simples da classe mediana;
- h: amplitude do intervalo de classe

Mediana para dados agrupados com intervalos de classe - Variável Contínua

Exemplo: Alturas

Estaturas (cm)	frequência
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
Total	40

Moda para dados não agrupados ou dados simples

- É o valor de maior frequência em um conjunto de dados. Basta identificar o elemento de maior frequência.
- Notação: m_o.
- **Exemplo 1**:
$$X = 2, 8, 3, 5, 4, 5, 3, 5, 5, 1.$$
- O elemento de maior frequência é 5. Portanto, $m_o = 5$. Neste caso, temos uma sequência unimodal.
- **Exemplo 2**:
$$X = 6, 10, 5, 6, 10, 2.$$
- Os elementos de maior frequência são 6 e 10. Portanto, $m_o = 6$ e $m_o = 10$. Neste caso, temos uma sequência bimodal.

Moda para dados não agrupados ou dados simples

- Podemos encontrar sequências trimodais, tetramodais e assim sucessivamente. Estas frequências serão chamadas de forma genérica por sequências polimodais.
- A moda pode sempre ser utilizada para dados qualitativos.

Moda para dados não agrupados ou dados simples

Exemplo: Suponha que os dados abaixo mostrem a preferência de refrigerantes em uma amostra de 50 pessoas.

Guaraná	Sprite	Coca-cola	Coca-cola	Coca-cola light
Coca-cola light	Pepsi	Coca-cola	Guaraná	Coca-Cola
Coca-cola	Pepsi	Sprite	Coca-cola light	Coca-Cola
Sprite	Guaraná	Coca-cola	Coca-cola	Coca-Cola
Coca-cola light	Pepsi	Sprite	Guaraná	Coca-Cola
Coca-cola	Guaraná	Guaraná	Guaraná	Coca-cola light
Guaraná	Pepsi	Pepsi	Coca-cola	Coca-Cola
Coca-cola	Coca-cola light	Coca-cola light	Guaraná	Guaraná
Guaraná	Guaraná	Sprite	Coca-cola	Coca-Cola
Coca-cola	Coca-Cola	Coca-cola	Coca-cola	Coca-cola light

Quem é a moda?

Moda para dados agrupados sem intervalos de classe - Variável Discreta

- É o valor com maior frequência.
- Na apresentação da variável discreta, as frequências já estão computadas na segunda coluna. Basta identificar o elemento de maior frequência.

Moda para dados agrupados sem intervalos de classe - Variável Discreta

Exemplo: Variável número de filhos dos empregados casados.

Número de filhos	frequência
0	2
1	6
2	10
3	12
4	4

Neste caso, $m_o = 3$.

Moda para dados agrupados sem intervalos de classe - Variável Discreta

Exemplo:

x_i	frequência
1	2
2	5
3	4
4	5
5	1

Neste caso, $m_o = 2$ e $m_o = 4$.

Moda para dados agrupados com intervalos de classe - Variável Contínua

- Para determinar a moda de uma variável contínua, podemos optar por vários processos. Daremos destaque à moda de Czuber.
- Chamaremos de classe modal, na apresentação da variável contínua, a classe que apresentar a maior frequência.

Moda para dados agrupados com intervalos de classe - Variável Contínua

Fórmula de Czuber:

$$m_o = l_{m_o} + \frac{f_{m_o} - f_{ant}}{2f_{m_o} - (f_{ant} + f_{post})} h,$$

onde

- m_o: moda;
- l_{m_o}: limite inferior da classe modal;
- f_{m_o}: frequência simples da classe modal;
- f_{ant}: frequência simples da classe anterior à modal;
- f_{post}: frequência simples da classe posterior à modal;
- h: amplitude do intervalo de classe.

Moda para dados agrupados com intervalos de classe - Variável Contínua

Vamos entender a fórmula de Czuber:

- Identifica-se classe modal e caracteriza-se o seu limite inferior l_{m_o} e o seu limite superior L_{m_o}.
- Unindo-se os pontos A e D e os pontos B e C, os segmentos de reta determinados se interceptam no ponto P.
- Em seguida projeta-se verticalmente este ponto no eixo horizontal obtendo o ponto M, que identifica-se como sendo a moda da distribuição.
- A amplitude do intervalo é h e está dividida em duas partes. Se chamarmos a primeira parte de x então o restante é $h - x$.
- Estes valores correspondem as alturas dos triângulos ABP e CDP respectivamente.

Moda para dados agrupados com intervalos de classe - Variável Contínua

Vamos entender a fórmula de Czuber:

- Como estes triângulos são semelhantes, os lados e as alturas são proporcionais. Então:

$$\frac{AB}{CD} = \frac{x}{h-x}.$$

- Usando a propriedade das propoções, podemos afirmar:

$$\frac{AB+CD}{AB} = \frac{x+h-x}{x} \Leftrightarrow \frac{AB+CD}{AB} = \frac{h}{x}$$

- Daí,

$$x = \frac{AB}{AB+CD}h$$

- Note que, $AB = AE - BE = f_{m_o} - f_{ant}$.
- Temos ainda, $CD = CF - DF = f_{m_o} - f_{post}$.

Moda para dados agrupados com intervalos de classe - Variável Contínua

Logo,

$$x = \frac{f_{m_o} - f_{ant}}{f_{m_o} - f_{ant} + f_{m_o} - f_{post}} h = \frac{f_{m_o} - f_{ant}}{2f_{m_o} - (f_{ant} + f_{post})} h$$

Como a moda é identificada como sendo o ponto M da figura, podemos afirmar que

$$m_o = l_{m_o} + x.$$

Então,

$$m_o = l_{m_o} + \frac{f_{m_o} - f_{ant}}{2f_{m_o} - (f_{ant} + f_{post})} h.$$

Moda para dados agrupados com intervalos de classe - Variável Contínua

Exemplo: Variável salário.

Salários	frequência
$4 \vdash 8$	10
$8 \vdash 12$	12
$12 \vdash 16$	8
$16 \vdash 20$	5
$20 \vdash 24$	1

$$m_o = 8 + \frac{12 - 10}{24 - (10 + 8)} \times 4 = 8 + \frac{8}{6} = 8 + 1,3 = 9,3.$$

Utilização das medidas de tendência central

- Na maioria das situações, não necessitamos calcular as três medidas de tendência central.
- Normalmente precisamos de apenas uma das medidas para caracterizar o centro da série.
- Qual a medida que deve ser usada?
- A medida ideal em cada caso é aquela que melhor representa a maioria dos dados da série.
- Quando todos os dados de uma série estatística são iguais, a média, a mediana e a moda coincidirão com este valor e, portanto qualquer uma delas representa bem a série. No entanto, este caso dificilmente ocorrerá na prática.

Utilização das medidas de tendência central

- Na maioria das vezes, teremos valores diferenciados para a série e consequentemente a medida irá representar bem, apenas os dados da série e consequentemente a medida irá representar bem, apenas os dados da série que se situam próximos a este valor.
- Os dados muito afastados em relação ao valor da medida não serão bem representados por ela.
- Desta forma, se uma série apresenta forte concentração de dados em sua área central, a média, a mediana e a moda ficam bem situadas em sua área central representando bem a série.
- Como a mais conhecida é a média, optamos por esta medida de tendência central.
- Concluindo, devemos optar pela média quando houver forte concentração de dados na área central da série.

Utilização das medidas de tendência central

- Se um a série apresenta forte concentração de dados em seu início, a mediana e a moda estarão posicionadas mais no ínicio da série, representando bem esta concentração.
- A média que é fortemente afetada por alguns valores posicionados no final da série se deslocará para a direita desta concentração não a representando bem.
- Como a mais conhecida entre mediana e moda é a mediana, esta será a medida mais indicada.
- Devemos optar pela mediana quando houver forte concentração dos dados no início ou no final da série.
- A moda deve ser a opção como medida de tendência central em séries que apresentam um elemento típico, ou seja, um valor cuja frequência é muito superior à frequência dos outros elementos da série. A moda também é usada em dados qualitativos.

Separatrizes

- A mediana separa o conjunto de dados em dois grupos que apresentam o mesmo número de valores.
- Existem medidas de posição ligadas à mediana relativamente à sua característica (dividir os dados em partes iguais).
- Estas medidas são conhecidas por **separatrizes** (valores que dividem um conjunto de dados em partes iguais), são elas:
  - mediana;
  - quartis;
  - quintis;
  - decis;
  - percentis.

Quartis

Os quartis dividem um conjunto de dados em quatro partes iguais. Assim,

- $Q_1 = 1^o$ quartil, deixa 25% dos elementos à esquerda;
- $Q_2 = 2^o$ quartil, coincide com a mediana, deixa 50% dos elementos à esquerda;
- $Q_3 = 3^o$ quartil, deixa 75% dos elementos à esquerda;

Quartis - Exemplo

O nível de albumina no sangue, um indicador do estado nutricional, foi medido em um grupo de 60 pacientes, obtendo-se os resultados (g/dl) apresentados em forma ordenada na tabela:

4,44	4,47	4,48	4,51	4,54	4,54	4,61	4,64	4,66	4,68
4,68	4,69	4,71	4,73	4,73	4,76	4,76	4,81	4,86	4,86
4,87	4,88	4,90	4,90	4,95	4,95	4,96	4,97	4,98	4,98
4,99	5,00	5,01	5,01	5,01	5,02	5,04	5,05	5,08	5,09
5,09	5,10	5,11	5,11	5,16	5,17	5,18	5,18	5,19	5,24
5,24	5,26	5,27	5,27	5,29	5,32	5,35	5,46	5,50	5,85

Quartis - Exemplo

O nível de albumina no sangue, um indicador do estado nutricional, foi medido em um grupo de 60 pacientes, obtendo-se os resultados (g/dl) apresentados em forma ordenada na tabela:

4,44	4,47	4,48	4,51	4,54	4,54	4,61	4,64	4,66	4,68
4,68	4,69	4,71	4,73	4,73	4,76	4,76	4,81	4,86	4,86
4,87	4,88	4,90	4,90	4,95	4,95	4,96	4,97	4,98	4,98
4,99	5,00	5,01	5,01	5,01	5,02	5,04	5,05	5,08	5,09
5,09	5,10	5,11	5,11	5,16	5,17	5,18	5,18	5,19	5,24
5,24	5,26	5,27	5,27	5,29	5,32	5,35	5,46	5,50	5,85

- O primeiro quartil (Q1), deixa 25% dos dados abaixo e 75% dos dados acima dele.
- **Conclusão:** O primeiro quartil corresponde ao nível de albumina de $4,73$ g/dl (decilitros), ou seja, 25% dos valores dos níveis de albumina estão abaixo de $4,73$ g/dl e 75% estão acima.

Quintis

- Se dividirmos a sequência ordenada em cinco partes, cada uma ficará com 20% de seus elementos.
- Os elementos que separam estes grupos são chamados quintis.
- Assim, o primeiro quintil, que indicaremos por K_1, separa a sequência ordenada, deixando à sua esquerda 20% de seus valores e à sua direita 80% dos seus valores.
- De modo análogo definimos os outros quintis: K_2, K_3, K_4, K_5.

Decis

- Se dividirmos a sequência ordenada em dez partes, cada uma ficará com 10% de seus elementos.
- Os elementos que separam estes grupos são chamados decis.
- Assim, o primeiro decil, que indicaremos por D_1, separa a sequência ordenada, deixando à sua esquerda 10% de seus valores e à sua direita 90% dos seus valores.
- De modo análogo definimos os outros decis: $D_2, D_3, \ldots, D_{10}$.

Percentis

- Se dividirmos a sequência ordenada em 100 partes, cada uma ficará com 1% de seus elementos.
- Os elementos que separam estes grupos são chamados percentis.
- Assim, o primeiro percentil, que indicaremos por P_1, separa a sequência ordenada, deixando à sua esquerda 1% de seus valores e à sua direita 99% dos seus valores.
- De modo análogo definimos os outros percentis: $P_2, P_3, \ldots, P_{100}$.

Observação

Se observarmos que os quartis, quintis e decis são múltiplos dos percentis, então basta estabelecer a fórmula de cálculo dos percentis. Todas as outras medidas podem ser identificadas como percentis. Desta forma:

$Q_1 = P_{25}$	$K_1 = P_{20}$	$D_1 = P_{10}$
$Q_2 = P_{50}$	$K_2 = P_{40}$	$D_2 = P_{20}$
$Q_3 = P_{75}$	$K_3 = P_{60}$	$D_3 = P_{30}$
	$K_4 = P_{80}$	$D_4 = P_{40}$
		$D_5 = P_{50}$
		$D_6 = P_{60}$
		$D_7 = P_{70}$
		$D_8 = P_{80}$
		$D_9 = P_{90}$

Dados simples (ou brutos)

- Devemos ordenar os elementos.
- Identificamos à medida que queremos obter com o percentil correspondente, P_i.
- Calculamos $i\%$ de n, ou seja,

$$\frac{i \times n}{100}$$

  para localizar a posição do percentil i na série ordenada (rol).
- Em seguida, identificamos o elemento que ocupa esta posição.

Dados simples (ou brutos)

- Se
$$\frac{i \times n}{100}$$
for um número inteiro, então P_i que estamos procurando identificar é um dos elementos do rol.
- Se
$$\frac{i \times n}{100}$$
não for um número inteiro, então P_i é um elemento intermediário entre os elementos que ocupam as posições aproximadas por falta e por excesso de $\frac{i \times n}{100}$. Neste caso, o P_i é definido como sendo a média dos valores que ocupam estas posições.

Dados simples (ou brutos) - Exemplos

1. Calcule Q_1 da sequência

$$X : 2; 5; 8; 5; 5; 10; 1; 12; 12; 11; 13; 15.$$

2. Calcule K_3 da sequência

$$X : 2; 8; 7,5; 6; 10; 12; 2; 9$$

Dados agrupados sem intervalos de classe - Variável discreta

- Se os dados estão apresentados na forma de uma variável discreta, eles já estão naturalmente ordenados.
- Identifica-se a medida que queremos obter com o percentil correspondente: P_i.
- Calcula-se $i\%$ de n, ou seja,

$$\frac{i \times n}{100}$$

  para localizar a posição do percentil i na série ordenada (rol).
- Em seguida utilizamos a frequência acumulada da série para localizar o elemento que ocupa esta posição.
- O valor deste elemento é o P_i.

Dados agrupados sem intervalos de classe - Variável discreta - Exemplo

Calcule D_4 para a série:

x_i	f_i
2	3
4	5
5	8
7	6
10	2

- O número do elementos da série é ?
- $D_4 = P_?$.

Dados agrupados sem intervalos de classe - Variável discreta - Exemplo

Calcule D_4 para a série:

x_i	f_i
2	3
4	5
5	8
7	6
10	2

- O número do elementos da série é $\sum f_i = 24$.
- Identificamos que $D_4 = P_{40}$.
- Calculamos 40% de 24, ou seja, $\frac{40 \times 24}{100} = 9,6$.
- Esta posição não inteira significa que o P_{40} é um valor compreendido entre o nono e o décimo elemento da série.

Dados agrupados sem intervalos de classe - Variável discreta - Exemplo

Construindo a frequência acumulada:

x_i	f_i	f_{ac}
2	3	3
4	5	8
5	8	16
7	6	22
10	2	24

- O nono elemento é 5, e o décimo elemento também é 5.
- Assim,

$$D_4 = P_{40} = \frac{5+5}{2} = 5.$$

- **Interpretação:** 40% dos valores desta série são valores menores ou iguais a 5 e 60% dos valores desta série são valores maiores ou

Dados agrupados com intervalos de classe - Variável contínua

- Se os dados estão apresentados na forma de uma variável contínua, eles já estão naturalmente ordenados e o número de elementos da série é $n = \sum f_i$.
- Para se obter uma fórmula geral para o cálculo dos percentis, vamos generalizar a fórmula da mediana:

$$\tilde{x}_c = l_{m_d} + \frac{\frac{n}{2} - fac_{ant}}{f_{m_d}} h,$$

- Identificando $\tilde{x}_c = P_{50}$, podemos obter a fórmula particular para o P_{50}.
- Note que a classe que contém a mediana é a mesma classe que contém o P_{50}.

Dados agrupados com intervalos de classe - Variável contínua

- Portanto, identificamos o limite inferior da classe que contém a mediana (l_{m_d}) com o limite inferior da classe que contém o P_{50}, que é l_{50}.
- O termo $\frac{n}{2}$ pode ser representado na linguagem do P_{50} como
$$\frac{50 \times n}{100}.$$
- A frequência simples da classe mediana f_{m_d} é a mesma frequência simples da classe que contém P_{50}, que chamamos de f_{50}.
- A frequência acumulada da classe anterior à classe mediana fac_{ant} é a frequência acumulada da classe anterior da classe que contém P_{50}. Este termo não se modifica, assim como h, que é a amplitude da classe.

Dados agrupados com intervalos de classe - Variável contínua

Assim, a fórmula da mediana, adaptada para a linguagem do P_{50} pode ser escrita por:

$$P_{50} = l_{50} + \frac{\frac{50 \times n}{100} - fac_{ant}}{f_{50}} h.$$

Dados agrupados com intervalos de classe - Variável contínua

Substituindo 50 pelo índice i, generalizamos a fórmula para o cálculo de qualquer percentil:

$$P_i = l_i + \frac{\frac{i \times n}{100} - fac_{ant}}{f_i} h,$$

em que

- P_i: percentil i $(i = 1, 2, \ldots, 99)$;
- l_i: limite inferior da classe que contém o percentil i;
- n: número de elementos da série;
- fac_{ant}: frequência acumulada da classe anterior à classe que contém o P_i;
- f_i: frequência simples da classe que contém o percentil i;
- h: amplitude do intervalo de classe

Quartis - Dados agrupados

- Quando os dados são agrupados, para determinar os quartis usamos a mesma técnica do cálculo da mediana, substituindo $\frac{\sum n_i}{2}$ por

$$\frac{q \sum n_i}{4},$$

onde q é o número de ordem do quartil.

- Então,
  - 1^o quartil $\Rightarrow \frac{\sum n_i}{4}$
  - 2^o quartil $\Rightarrow \frac{\sum n_i}{2}$
  - 3^o quartil $\Rightarrow \frac{3 \sum n_i}{4}$

Quartis - Dados agrupados sem intervalos de classe

Número de filhos	frequência (n_i)	frequência acumulada (N_i)
0	4	4
1	5	9
2	7	16
3	3	19
5	1	20
Total	20	

$$\frac{\sum ni}{4} = \frac{20}{4} = 5$$

a menor frequência acumulada que supera esse valor é 9, que corresponde ao valor 1 da variável. Logo,

$$Q_1 = 1$$

Analogamente, $\frac{3\sum ni}{4} = \frac{3\times 20}{4} = 15 \Longrightarrow Q_3 = 2$

Quartis - Dados agrupados com intervalos de classe

Temos,

$$
\begin{aligned}
Q_1 &= \ell^* + \frac{\left[\frac{\sum n_i}{4} - N_{anterior}\right] h^*}{n^*} \\
Q_2 &= \ell^* + \frac{\left[\frac{\sum n_i}{2} - N_{anterior}\right] h^*}{n^*} \longmapsto \text{mediana} \\
Q_3 &= \ell^* + \frac{\left[\frac{3\sum n_i}{4} - N_{anterior}\right] h^*}{n^*}
\end{aligned}
$$

Quartis - Exemplo

Exemplo: Considere a distribuição e calcule Q_1 e Q_3:

Estaturas (cm)	frequência (n_i)	frequência acumulada (N_i)
150 - 154	4	4
154 - 158	9	13
158 - 162	11	24
162 - 166	8	32
166 - 170	5	37
170 - 174	3	40
Total	40	

Exemplo - Primeiro quartil

$$\frac{\sum n_i}{4} = \frac{40}{4} = 10$$

- $\ell^* = 154 \longrightarrow$ limite inferior da classe considerada;
- $N_{anterior} = 4 \longrightarrow$ frequência acumulada da classe anterior à classe considerada;
- $n^* = 9 \longrightarrow$ frequência simples da classe considerada;
- $h^* = 4 \longrightarrow$ amplitude da classe.

$$\begin{aligned} Q_1 &= 154 + \frac{(10-4) \times 4}{9} \\ &= 154 + 2,66 \\ &= 156,7\ cm \end{aligned}$$

Exemplo - Terceiro quartil

$$\frac{3\sum n_i}{4} = \frac{3 \times 40}{4} = 30$$

- $\ell^* = 162 \longrightarrow$ limite inferior da classe considerada;
- $N_{anterior} = 24 \longrightarrow$ frequência acumulada da classe anterior à classe considerada;
- $n^* = 8 \longrightarrow$ frequência simples da classe considerada;
- $h^* = 4 \longrightarrow$ amplitude da classe.

$$\begin{aligned} Q_3 &= 162 + \frac{(30-24) \times 4}{8} \\ &= 162 + 3 \\ &= 165\ cm \end{aligned}$$

Decis

São os valores que dividem o conjunto de dados em 10 partes iguais. (Representação gráfica)

- Quando os dados são agrupados, para determinar os decis usamos a mesma técnica do cálculo da mediana, substituindo $\frac{\sum n_i}{2}$ por

$$\frac{d \sum n_i}{10},$$

onde d é o número de ordem do decil, $d = 1, 2, 3, 4, 5, 6, 7, 8, 9$.

Percentis

São os valores que dividem o conjunto de dados em 100 partes iguais. (Representação gráfica)

- Quando os dados são agrupados, para determinar os percentis usamos a mesma técnica do cálculo da mediana, substituindo $\frac{\sum n_i}{2}$ por
$$\frac{p \sum n_i}{100},$$
onde d é o número de ordem do decil, $p = 1, 2, 3, \ldots, 99$.

Exemplo (Decil e Percentil)

Determine o 4º decil e o 72º percentil da seguinte distribuição:

Classes	frequência (n_i)	frequência acumulada (N_i)
4 - 9	8	8
9 - 14	12	20
14 - 19	17	37
19 - 24	3	40
Total	40	

Exemplo (Decil e Percentil)

Determine o 4^o decil e o 72^o percentil da seguinte distribuição:

Classes	frequência (n_i)	frequência acumulada (N_i)
4 - 9	8	8
9 - 14	12	20
14 - 19	17	37
19 - 24	3	40
Total	40	

- Cálculo do D_4: $\frac{d\sum n_i}{10} = \frac{4\times 40}{10} = 16$
  - $\ell^* = 9 \longrightarrow$ limite inferior da classe D_4;
  - $N_{anterior} = 8 \longrightarrow$ frequência acumulada da classe anterior à classe D_4;
  - $n^* = 12 \longrightarrow$ frequência simples da classe D_4;
  - $h^* = 5 \longrightarrow$ amplitude da classe D_4.

Exemplo (Decil e Percentil)

$$D_4 = 9 + \frac{(16 - 8)\,5}{12} = 12,33$$

- Cálculo do P_{72}:

$$\frac{p \sum n_i}{100} = \frac{72 \times 40}{100} = 28,8$$

  - $\ell^* = 14 \longrightarrow$ limite inferior da classe P_{72};
  - $N_{anterior} = 20 \longrightarrow$ frequência acumulada da classe anterior à classe P_{72};
  - $n^* = 17 \longrightarrow$ frequência simples da classe P_{72};
  - $h^* = 5 \longrightarrow$ amplitude da classe P_{72}.

$$P_{72} = 14 + \frac{(28,8 - 20)\,5}{17} = 16,59$$