Capítulo 1

Medidas de Posição e Dispersão

1.1 Medidas de Posição

Frequentemente é necessário resumir dados por meio de um único número que, a seu modo, descreve todo o conjunto. Por exemplo, em um experimento podemos estar interessados em um valor que descreva o centro dos dados, ou o valor que é ultrapassado por apenas 30% dos dados. As medidas estatísticas que descrevem tais características são chamadas de medidas de localização. Dentre elas, as medidas que indicam o centro dos dados são chamadas de medidas de posição.

1- Média Aritmética ($\bar{X}$): é o quociente da soma dos valores de uma seqüência $x_1, x_2, ..., x_n$ pelo número total de elemento, isto é,

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

onde n é o número de elementos da seqüência.

1.1- Média para dados agrupados em distribuição de freqüências.

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

onde n é o número de elementos da seqüência, x_i são os valores da seqüência e f_i são as freqüências dos valores na seqüência.

Exemplo: A tabela mostra o peso (kg) de 37 mercadorias de um depósito.

Classes	f_i	x_i	fac_i
5⊢10	4	7.5	4
10⊢15	7	12.5	11
15⊢20	11	17.5	22
20⊢25	18	22.5	40
25⊢30	6	27.5	46
Total	46	—	—

Fonte: Dados fictícios.

A média aritmética será : $\bar{X} = \frac{880}{46} = 19,13kg.$

1.2- Propriedades sobre a média

a) $Y_i = X_i \pm c \Rightarrow \bar{Y} = \bar{X} \pm c$

b) $Y_i = X_i \times c \Rightarrow \bar{Y} = \bar{X} \times c$

c) $Y_i = \frac{X_i}{c} \Rightarrow \bar{Y} = \frac{\bar{X}}{c}$

Obs: A média é utilizada quando desejamos obter a medida de posição que possui a maior estabilidade.

2- Moda (Mo) : é o valor que ocorre com maior freqüência na seqüência de valores em estudo.

Exemplo: X={ 2, 2, 2, 3, 4, 4, 4, 4, 5, 5, 6}

Logo, Mo = 4 (elemento com maior freqüência).

2.1 - Moda para dados agrupados em distribuição de freqüências

$$Mo = L_i + \frac{a(f_i - f_{i-1})}{2f_i - (f_{i+1} + f_{i-1})},$$

onde L_i é o limite inferior da classe modal; a é a amplitude da classe modal; f_i é a freqüência absoluta da classe modal; f_{i+1} é a freqüência absoluta da classe posterior à classe modal; f_{i-1} é a freqüência absoluta da classe anterior à classe modal.

Exemplo: Considerando a tabela do exemplo anterior, encontre a moda.

Classe modal: $20 \vdash 25 \Rightarrow Mo = 20 + \frac{5(18-11)}{2(18)-(11+6)} = 21,84kg.$

3- Mediana (Me ou Md): é o valor que divide um conjunto de dados ordenados em duas partes iguais, com o mesmo número de elementos de cada lado. Verificamos que, estando ordenados os valores de uma série e sendo n o número de elementos da seqüência, o valor mediano será:

- o termo de ordem $\frac{n+1}{2}$, se n for ímpar;
- a média aritmética dos termos de ordem $\frac{n}{2}$ e $\frac{n+2}{2}$, se n for par.

Exemplo: Seja a sequência { 4, 5, 7, 8, 11, 12, 15}

Temos que n é ímpar a mediana será encontrada na 4ª posição . Logo, $Me = 8$

Exemplo : Seja a sequência {5, 6, 9, 12, 13, 17}

Temos que n é par, então a mediana será a média aritmética dos termos que ocupam a 3ª e 4ª posições da seqüência em estudo. Logo, $Me = \frac{9+12}{2} = 10,5$.

3.1- Mediana para dados agrupados em distribuição de freqüências

1º) Identificar a classe mediana. Calcula-se $\frac{n}{2}$ e toma-se a freqüência acumulada imediatamente superior ao valor encontrado.

2º) Utiliza-se a fórmula:

$$Me = L_i + \frac{a}{f_i}\left(\frac{n}{2} - fac_{i-1}\right),$$

onde L_i é o limite inferior da classe mediana; a é a amplitude da classe mediana; f_i é a freqüência absoluta da classe mediana; fac_{i-1} é a freqüência acumulada da classe anterior à classe mediana.

Exemplo: Considerando a tabela anterior, calcule a mediana.

Classe mediana: $20 \vdash 25 \Rightarrow Me = 20 + \frac{5}{18}\left(\frac{46}{2} - 22\right) = 20,28kg$

4- Simetria e Assimetria

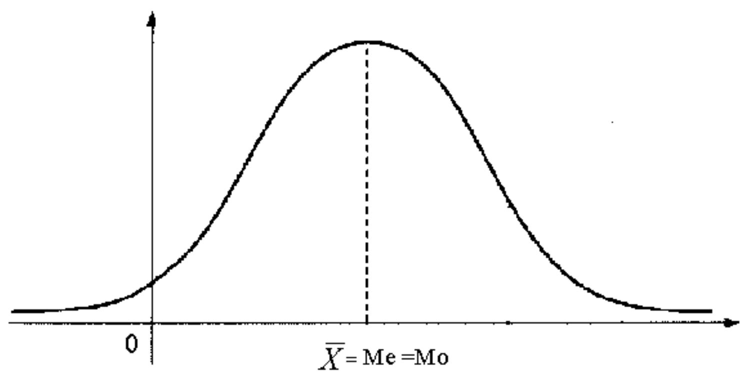
Assimetria é o grau de afastamento de uma distribuição da unidade de simetria.

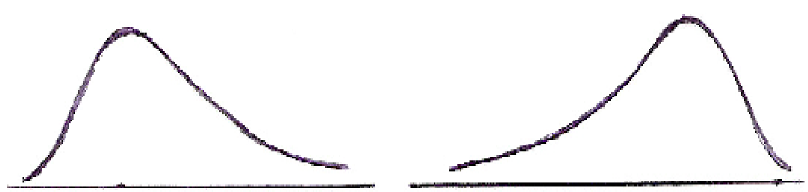
Temos as seguintes curvas de simetria:

Se $\bar{X} = Me = Mo \Rightarrow$ curva simétrica.

Se $Mo \ < \ Me \ < \ \bar{X} \Rightarrow$ curva assimétrica positiva.

Se $Mo \ > Me \ > \bar{X} \Rightarrow$ curva assimétrica negativa.





Curva Assimétrica Positiva
Moda < Mediana < Média

Curva Assimétrica Negativa
Moda > Mediana > Média

4- Quantis (Separatrizes)

Quantis: são os valores dos Quartis ($Q_k, \ k = 1, 2, 3$), Decis ($D_k, \ k = 1, 2, ..., 9$) e Percentis ($P_k, \ k = 1, 2, ..., 99$) que dividem um conjunto de dados ORDENADOS em partes iguais conforme veremos a seguir.

Para localizarmos cada medida (quartil, decil, percentil), temos que localizar a posição em que se encontra o valor da medida desejada. Conforme vimos em Mediana, temos a seguinte regra para localizar a posição e, consequentemente, a medida desejada.

4.1- Quartil ($Q_k,\ k = 1, 2, 3$)

São os valores que dividem uma seqüência ordenada em quatro partes iguais, deixando 25% dos dados em cada parte.

- Se n for ímpar a posição é obtida por $\frac{(n+1)k}{4}$;
- Se n for par, as posições são obtidas por $\frac{nk}{4}$ e $\frac{(n+2)k}{4}$. Neste caso, as medidas desejadas são obtidas fazendo a média dos valores que ocupam as posições encontradas.

4.2- Decil ($D_k,\ k = 1, 2, ..., 9$)

São os valores que dividem uma seqüência ordenada em dez partes iguais, deixando 10% dos dados em cada parte.

- Se n for ímpar a posição é obtida por $\frac{(n+1)k}{10}$;
- Se n for par, as posições são obtidas por $\frac{nk}{10}$ e $\frac{(n+2)k}{10}$. Neste caso, as medidas desejadas são obtidas fazendo a média dos valores que ocupam as posições encontradas.

4.3- Percentil ($P_k,\ k = 1, 2, ..., 99$)

São os valores que dividem uma seqüência ordenada em cem partes iguais, deixando 1% dos dados em cada parte.

- Se n for ímpar a posição é obtida por $\frac{(n+1)k}{100}$;
- Se n for par, as posições são obtidas por $\frac{nk}{100}$ e $\frac{(n+2)k}{100}$. Neste caso, as medidas desejadas são obtidas fazendo a média dos valores que ocupam as posições encontradas.

Exemplo: Seja a sequência { 2, 2, 3, 3, 3, 4, 5, 7, 7, 11, 12, 12, 12, 13, 13, 14, 15}, vamos obter o Q_1 e D_6. Temos que $n = 17$ é ímpar.

Então, para Q_1 teremos a posição $\frac{(n+1)k}{4} = \frac{(18)2}{4} = 8^{a}$.

Logo, $Q_1 = 7$.

Para D_6, teremos a posição $\frac{(n+1)k}{10} = \frac{(18)6}{10} = 10,8 \approx 11^{a}$.

Logo, $D_6 = 12$.

4.4- Cálculo do quartil, decil e percentil para dados agrupados em distribuição de freqüências.

I) Quartil

1º) Identificar a classe k- quartil. Calcula-se $\frac{kn}{4}$ e toma-se a freqüência acumulada imediatamente superior ao valor encontrado.

2º) Aplicar a fórmula: $Q_k = L_i + \frac{a}{f_i}\left(\frac{kn}{4} - fac_{i-1}\right)$.

II) Decil

1º) Identificar a classe k- decil. Calcula-se $\frac{kn}{10}$ e toma-se a freqüência acumulada imediatamente superior ao valor encontrado.

2º) Aplicar a fórmula: $D_k = L_i + \frac{a}{f_i}\left(\frac{kn}{10} - fac_{i-1}\right)$.

III) Percentil

1º) Identificar a classe k-percentil. Calcula-se $\frac{kn}{100}$ e toma-se a freqüência acumulada imediatamente superior ao valor encontrado.

2º) Aplicar a fórmula: $P_k = L_i + \frac{a}{f_i}\left(\frac{kn}{100} - fac_{i-1}\right)$,

onde: L_i é o limite inferior da classe k-quantil; a é a amplitude da classe k-quantil; f_i é a freqüência absoluta da classe k-quantil; fac_{i-1} é a freqüência acumulada da classe anterior à classe k-quantil.

Exemplo: Considerando a tabela abaixo, encontrar Q_1, D_4 e P_{73}.

Classes	f_i	x_i	fac_i
5⊢10	4	7.5	4
10⊢15	7	12.5	11
15⊢20	11	17.5	22
20⊢25	18	22.5	40
25⊢30	6	27.5	46
Total	46	—	—

Resolução:

a) Temos que $\frac{kn}{4} = \frac{46}{4} = 11,5$, logo a classe Q_1 será: $15 \vdash 20$.

$$Q_1 = 15 + \frac{5}{11}(11,5 - 11) = 15,23.$$

b) Temos que $\frac{kn}{10} = \frac{4\times 46}{10} = 18,4$, logo a classe D_4 será: $15 \vdash 20$.

$$D_4 = 15 + \frac{5}{11}(18,4 - 11) = 18,36.$$

c) Temos que $\frac{kn}{100} = \frac{73\times 46}{100} = 33,58$, logo a classe P_{73} será: $20 \vdash 25$.

$$P_{73} = 20 + \frac{5}{18}(33,58 - 22) = 23,22.$$

5- Medida de Curtose

5.1- Curtose: é o grau de achatamento de uma distribuição em relação a distribuição padrão (curva normal).

5.2- Coeficiente de curtose

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}.$$

Se $C = 0,263 \Rightarrow$ a curva será mesocúrtica;

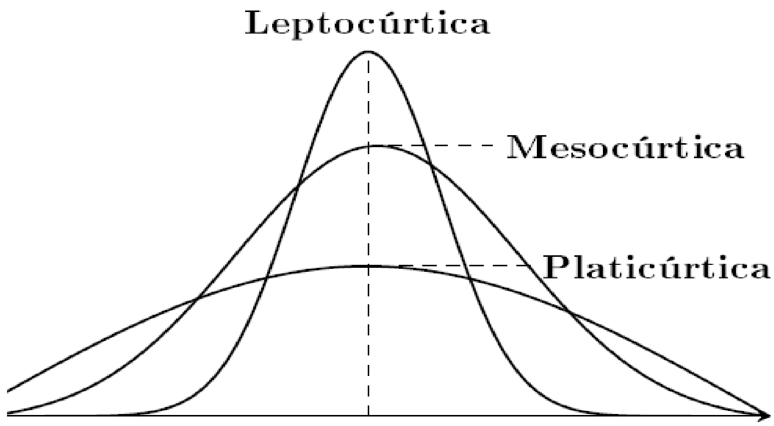
Se $C < 0,263 \Rightarrow$ a curva será leptocúrtica;

Se $C > 0,263 \Rightarrow$ a curva será platicúrtica.

Exemplo: Sabendo-se que uma distribuição apresenta as seguintes medidas,

$Q_1 = 15,23, \quad Q_3 = 23,47, \quad P_{10} = 10,43, \quad P_{90} = 26,17$ temos:

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{23,47 - 15,23}{2\times(26,17 - 10,43)} = 0,2617 \approx 0,262.$$



Como $C = 0,262 < 0,263 \quad \Rightarrow \quad$ a curva é leptocúrtica.

1.1.1 Medidas de Dispersão ou Variabilidade

Resumir um conjunto de dados por uma única medida de posição faz que percamos informações sobre a variabilidade do conjunto de dados. Por exemplo, suponhamos que 3 grupos formados com 5 alunos fizeram um teste de estatística, obtendo-se as notas:

Grupo 1: 2, 4, 5, 6, 8 $\Rightarrow \quad \bar{X}_1 = 5.$

Grupo 2: 2, 2, 5, 4, 7 $\Rightarrow \quad \bar{X}_2 = 5.$

Grupo 3: 5, 5, 5, 5, 5 $\Rightarrow \quad \bar{X}_3 = 5.$

Os grupos possuem a mesma média aritmética $\bar{X}_1 = \bar{X}_2 = \bar{X}_3 = 5$. Esse resultado nada informa sobre a variabilidade em cada grupo. Assim, é conveniente buscar medidas que informem a variabilidade de um conjunto de dados, segundo algum critério.

Def.: As medidas de dispersão medem o grau de variação dos elementos de uma seqüência de valores com relação a média aritmética dessa seqüência.

1- Amplitude Total (A_T): É a diferença entre o maior e o menor valor de uma seqüência, isto é:

$$A_T = X_{max} - X_{min}.$$

Exemplo: Seja $X = \{25, 30, 32, 37, 42, 47\}$ então, $A_T = 47 - 25 = 22$.

Obs: A amplitude total não é uma medida dispersão muito utilizada devido ao fato de levar em consideração apenas os valores extremos da seqüência.

2- Desvio Médio (DM):

Desvio é a diferença entre cada valor de uma seqüência e a média aritmética da mesma.

$$d_i = x_i - \bar{X}, \quad i = 1, 2, ..., n.$$

Desvio médio é a soma dos módulos dos desvios dividida pelo número de elementos da seqüência.

$$DM = \frac{\sum_{i=1}^{k} \mid d_i \mid f_i}{n} = \frac{\sum_{i=1}^{k} \mid x_i - \bar{X} \mid f_i}{n}, \text{sendo } k \text{ o número de classes.}$$

Exemplo: Seja $X : \{25, 30, 32, 37, 42, 47\}$ onde $\bar{X} = 35,5$.

3- Variância (S^2) **e Desvio-Padrão** (S)

Trabalhando com a **população:**

$$S^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{X})^2 f_i}{n} \quad ou \quad S^2 = \frac{1}{n}\left[\sum_{i=1}^{k} x_i^2 f_i - \frac{(\sum_{i=1}^{k} x_i f_i)^2}{n}\right].$$

Trabalhando com **amostra:**

$$S^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{X})^2 f_i}{n-1} \quad ou \quad S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - \frac{(\sum_{i=1}^{k} x_i f_i)^2}{n}\right].$$

Nota-se que a variância é uma soma de quadrados, o que pode causar problemas de interpretação. Dessa forma, se faz necessário definir uma outra medida de dispersão, que é a raiz quadrada da variância denominada desvio-padrão, voltando a unidade de medida original. Importante lembrar que $S^2 \geq 0$ e $S \geq 0$

Então, o **desvio-padrão** é dado por : $S = \sqrt{S^2}$.

Exemplo: Considere a tabela abaixo e calcule a variância e o desvio-padrão.

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$
05⊢10	4	7.5	30	225
10⊢15	7	12.5	87,5	1093,75
15⊢20	11	17.5	192,5	3368,75
20⊢25	18	22.5	405	9112,5
25⊢30	6	27.5	165	4537,5
Total	46	—	880	18337,5

Temos que :

$$S^2 = \frac{1}{n}\left[\sum_{i=1}^{k} x_i^2 f_i - \frac{(\sum_{i=1}^{k} x_i f_i)^2}{n}\right] = \frac{1}{46}\left[18337,5 - \frac{(880)^2}{46}\right] \approx 32,66\ cm^2.$$

e o desvio-padrão será: $S = \sqrt{32,66} \approx 5,72\ cm.$

Obs: Quanto menor o desvio-padrão, mais homogênio é um conjunto de dados.

4- Coeficiente de Variação (CV)

Trata-se de uma **medida relativa de dispersão**, útil para a comparação em termos relativos do grau de concentração em torno da média de seqüências distintas. É dado por:

$$CV = \frac{S}{\bar{X}} \times 100.$$

Obs.: A grande utilidade do coeficiente de variação é permitir a comparação das variabilidades de diferentes conjuntos de dados.

Exemplo: Em uma empresa existem dois tipos de parafusos, onde as medidas quanto aos comprimentos são as seguintes:

Parafuso A: $\bar{X} = 5\ cm$ e $S = 1\ cm$

Parafuso B: $\bar{X} = 8\ cm$ e $S = 1,5\ cm$

Daí,

$$CV_A = \frac{1}{5} \times 100 = 20\% \qquad e \qquad CV_B = \frac{1,5}{8} \times 100 = 18,75\%.$$

Podemos concluir que o parafuso B possui menor dispersão em torno do comprimento médio.

5- Exercícios

1- Dada a seqüência: 1, 1, 2, 2, 3, 4, 6, 6, 6, 7 calcular a média, a mediana e a moda populacional.

2- Para as tabelas de distribuição de freqüência abaixo calcular: (a) média (b) mediana (c) moda (d) 1º quartil (e) 4º decil (f) 68º percentil (g) classifique a curva quanto a assimetria (h) medida de curtose.

(a)

Altura(cm)	160⊢164	164⊢168	168⊢172	172⊢176	176⊢180	180⊢184
f_i	6	11	15	22	25	10

(b)

Classes	0⊢25	25⊢50	50⊢75	75⊢100	100⊢125	125⊢150
f_i	9	12	18	26	25	15

(c)

Peso(kg)	40⊢43	43⊢46	46⊢49	49⊢52	52⊢55	55⊢58
f_i	7	13	19	24	35	28

3) Achar o 3º quartil, o 8º decil e o 81º percentil da distribuição:

Altura(cm)	0⊢1	1⊢2	2⊢3	3⊢4	4⊢5	5⊢6
f_i	15	27	31	23	10	12

4) Estudar a distribuição abaixo, com respeito à assimetria e à curtose.

Altura(cm)	50⊢100	100⊢150	150⊢200	200⊢250	250⊢300	300⊢350
f_i	5	16	21	28	19	8

5) Calcular a média, a moda e a mediana da distribuição:

Altura(cm)	6	9	14	17	25	30
f_i	7	11	18	22	20	10

6- Para as distribuições do exercício 1, encontrar:

a) desvio médio; b) Variância; c) desvio-padrão; d) coeficiente de variação.

7- Os dados a seguir referem-se aos graus dos funcionários da UFRRJ em uma recente avaliação: 3;4;4;4;4;4;1;1;1;1;1;1;1;2;3;8;8;8;7;7;8;9;6;5;5;5;5;5;10;0;2;3;4;1;9; 9.

Pede-se: a) grau médio b) grau modal c) grau mediano d) amplitude total.

8- Para a distribuição do exercício 2:

a) Calcule a variância populacional;

b) Calcular o coeficiente de curtose.

9- Para as distribuições dos exercícios 4 e 5, calcular:

a) Variância b) desvio-padrão c) desvio médio d) coeficiente de variação.

10- Responda CERTO ou ERRADO para cada uma das afirmativas abaixo. Procedendo os acertos que se façam necessários

a)A soma dos desvios ou afastamentos tomados a partir da média aritmética nunca será nula.

b) Se a média aritmética de um conjunto de observações é igual a zero, podemos afirmar que todas observações são iguais a zero.

c) Em uma distribuição só existe um único valor para a moda.

d) Em uma distribuição a soma de todas as freqüências absolutas será sempre igual a unidade.

e) Em uma distribuição ao somarmos ou diminuirmos todas as observações por uma mesma constante, a média, a moda e a mediana ficarão aumentadas ou diminuídas da referida constante.

f) Em uma distribuição ao multiplicarmos ou dividirmos todas as observações por uma mesma constante, a média aritmética, a moda, a mediana e o desvio médio da nova distribuição formada ficará multiplicada ou dividida pela referida constante.

g) Se em uma distribuição os valores da média aritmética, da moda e da mediana forem iguais a uma mesma constante, podemos afirmar que todas as observações são iguais a referida constante.

h) O valor de uma medida de dispersão nunca poderá ser igual ou inferior a zero.

i) Separatriz é um valor que divide uma distribuição em duas partes quaisquer.

j) Em uma distribuição sempre o valor do 1º Quartil, da mediana, do 5º Decil e do 50º Percentil serão iguais.

k) A amplitude total é uma medida de dispersão, porém, a sua utilização como tal não é aconselhável devido ao fato de que em seu calculo são utilizados valores extremos.

l) Quando o valor de uma medida de dispersão é igual a zero, significa que o fenômeno em estudo é totalmente homogêneo.

m) Se em um conjunto de observações a média aritmética, a mediana e a moda são iguais a uma mesma constante, a variância, o desvio padrão, o desvio médio e a amplitude total do conjunto também serão iguais a zero.