3 MEDIDAS DE TENDÊNCIA CENTRAL E DE VARIABILIDADE DE DADOS

Na pesquisa agropecuária, os pesquisadores utilizam a Estatística Experimental para obter, analisar e interpretar dados experimentais, obtidos de experimentos, visando à elucidação de princípios biológicos bem como a solução de problemas agropecuários.

Na elucidação de tais princípios e na solução de tais problemas, o pesquisador define quais as características que irá utilizar para avaliar os tratamentos, de modo que possa atingir os objetivos da pesquisa. Por exemplo, na avaliação de variedades de milho e na avaliação de raças bovinas de leite, o pesquisador pode definir as seguintes características: resistência à lagarta do cartucho, período de maturação da espiga e rendimento de grãos (kg/ha), no caso do milho, e resistência a carrapato, consumo alimentar e rendimento de leite (kg/dia), no caso de bovino de leite, para avaliar seus tratamentos. Cada característica é medida nas parcelas e é denominada de **variável**.

Uma variável pode ser **discreta** ou **contínua**. Variável discreta é aquela que somente pode ter certos valores da amplitude de variação, ou seja, valores inteiros que se originam de contagens. Por exemplo, número de plantas doentes por parcela, número de sementes por fruto, número de ovos por galinha em determinado período, número de carrapatos por cavalo, etc.. Variável contínua é aquela que pode assumir qualquer valor dentro da amplitude de variação, ou seja, valores decimais que se originam de medições. Altura e rendimento de grãos de plantas de milho, peso e produção de leite de vacas leiteiras são exemplos desse tipo de variável.

No linguajar estatístico, uma **população** é um conjunto de medições, de uma única variável, efetuadas sobre todos os indivíduos pertencentes a uma classe. No nosso caso, por exemplo, o rendimento de grãos (kg/ha) de todos os campos de milho no Brasil, cultivados com a variedade CENTRALMEX, constituiu uma população. Da mesma forma, o rendimento de leite (kg/dia) de todas as vacas holandesas criadas no Estado de Alagoas constitui uma população. As medições individuais de uma variável recebem o nome de **elemento**.

Uma **amostra** é um conjunto de medições que constitui parte de uma população. A partir da amostra obtêm-se informações e fazem-se inferências acerca da população. Por esta razão, é importante que a amostra seja representativa da população.

As populações são descritas mediante características denominadas **parâmetros**. Os parâmetros são valores fixos; por exemplo, a média aritmética de todos os elementos de uma população é um parâmetro. As amostras são descritas pelas mesmas características, mas recebem a denominação de **estatístico**. A média de uma amostra é um estatístico. Calculam-se os estatísticos das amostras para estimarem-se os parâmetros da população. Obviamente, os estatísticos variam de amostra para amostra enquanto que os parâmetros têm apenas um valor.

3.1 Organização de Dados

Diferentes valores de uma variável apresentam distintas freqüências de incidência em sua população. Para caracterizar convenientemente uma população, os dados provenientes de uma amostra grande, como, por exemplo, os dados brutos de altura de planta (cm) de sorgo granífero da TABELA 3.1 e os dados brutos de peso corporal (g) de um lote misto de frangos de corte com 15 dias de idade da TABELA 3.2, são organizados mediante a construção de uma **tabela de freqüência**, um **histograma de freqüência** ou um **polígono de freqüência**.

TABELA 3.1 – DADOS BRUTOS DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

90,60	79,30	95,90	75,35	99,10	46,00
87.90	87,60	78,30	79,80	104,80	68,00
74,92	86,75	93,78	75,50	57,13	84,18
100,80	99,80	65,90	74,65	95,40	58,65
94,20	71,80	85,00	73,70	81,60	66,20
84,80	82,50	81,30	106,90	64,20	48,20
63,90	76,45	59,50	83,90	80,80	110,00
79,20	68,70	82,60	70,30	81,30	77,51
68,70	89,10	77,60	93,79	108,00	82,00
74,35	89,70	98,45	71,75	55,10	56,20
74,10	64,50	90,80	78,88	75,80	78,61
88,16	88,00	55,80	71,35	60,30	71,80
70,15	79,20	79,90	96,80	75,65	73,05
78,67	79,10	73,10	69,90	74,00	75,60
85,00	67,00	76,50	64,05	71,30	52,40

TABELA 3.2 – DADOS BRUTOS DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE

485,6	482,0	476,7	473,8	482,8	495,5
482,1	478,5	484,8	522,0	459,0	498,3
470,7	479,5	469,0	472,9	468,9	439,1
490,2	499,1	464,4	418,0	462,6	502,8
455,0	482,0	442,9	460,3	449,0	499,2
482,8	481,8	501,6	468,9	452,4	470,0
463,1	527,8	506,8	494,8	503,8	488,6
469,0	487,3	452,8	469,0	502,0	469,5
444,4	484,7	459,0	438,7	530,0	499,1
442,9	460,3	488,9	494,2	429,5	462,0
501,6	468,9	507,4	484,7	499,1	471,8
500,8	484,0	506,8	481,9	475,7	527,8
456,3	481,6	427,8	486,4	469,4	466,5
459,2	473,0	436,9	453,2	484,2	474,2
476,9	478,9	488,0	507,1	451,7	485,0
460,5	454,3	475,9	467,3	467,2	466,9
469,1	487,1	453,1	465,9	468,2	453,2
439,7	436,9	507,5	487,2	469,4	459,7
477,9	458,2	502,8	489,5	458,0	477,4

475,3	477,9	469,0	499,3	452,9	477,2

3.1.1 Tabela de freqüência

A tabela de freqüência proporciona ao pesquisador um meio eficaz de organização dos dados para estudo do comportamento de variáveis de interesse.

Na construção de uma tabela dessa natureza, devem-se levar em conta certas considerações importantes:

a) O intervalo de classe será de amplitude uniforme e de tamanho que se manifestem as linhas características da distribuição. Assim, o intervalo de classe não deve ser tão grande para não se cometer um erro considerável ao supor que o ponto médio do intervalo é o valor médio da classe. Não deve ser tão pequeno para não aparecerem demasiadas classes com freqüência zero ou muito pequenas.

b) Se possível, é conveniente fazer com que o ponto médio da classe seja um número inteiro.

c) As freqüências das classes podem ser absoluta, relativa ou relativa acumulada. Cabe ao pesquisador escolher a que mais lhe convier. A freqüência absoluta da uma classe corresponde a quantidade de valores de uma determinada variável que pertence a referida classe. Esse tipo de freqüência informa apenas o número absoluto de valores de um determinado intervalo de classe. Já a freqüência relativa de uma classe corresponde a freqüência absoluta da referida classe dividida pela soma de todas as freqüências absolutas, sendo expressa em porcentagem. Ela é útil quando se quer conhecer à proporção de valores situados em um determinado intervalo de classe ou quando se querem comparar conjuntos de dados que contenham números desiguais de observações. Enquanto que a freqüência relativa acumulada de uma classe corresponde à soma da freqüência relativa da referida classe e todas as outras freqüências relativas anteriores. Esse tipo de freqüência é útil quando se querem comparar conjuntos de dados que contenham números desiguais de observações.

Para a construção de uma tabela de freqüência, primeiramente define-se o número de classes, normalmente por meio da seguinte Fórmula de STURGES (citada por IPARDES, 2000):

$$k = 1 + 3{,}32 \times \log N$$

onde:
k = número de classes;
N = número total de observações.

Sem considerar a fórmula acima para se definir o número de classes, SPIEGEL (1993) recomenda, como regra geral, que o número de classes esteja entre cinco e 20. Por outro lado, MAGALHÃES e LIMA (2005), sem adotarem nenhuma regra formal quanto ao número de classes, utilizam, em geral, de cinco a oito classes.

Em seguida, determina-se a amplitude total dos dados, que é a diferença entre o maior e o menor valor da série.

De posse desses valores, define-se o intervalo de classe, dividindo a amplitude total pelo número de classes. Em seguida, são estabelecidos os limites inferiores e superiores das classes, onde o limite inferior da segunda classe é igual ao limite superior da primeira, e assim sucessivamente, observando-se que todos os dados devem estar entre o limite inferior da primeira classe e o limite superior da última classe.

Como exemplo têm-se as tabelas de freqüência de altura de planta (cm) de sorgo granífero (TABELA 3.3) e de peso corporal (g) de um lote misto de frangos de corte com 15 dias de idade (TABELA 3.4) contendo os três tipos de freqüências (absoluta, relativa e relativa acumulada).

TABELA 3.3 – TABELA DE FREQÜÊNCIAS ABSOLUTA, RELATIVA E RELATIVA ACUMULADA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

Intervalo de Classe	Ponto Médio	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Relativa Acumulada (%)
46,00 – 54,00	50,00	3	3,33	3,33
54,00 – 62,00	58,00	7	7,78	11,11
62,00 – 70,00	66,00	11	12,22	23,33
70,00 – 78,00	74,00	24	26,67	50,00
78,00 – 86,00	82,00	22	24,44	74,44
86,00 – 94,00	90,00	11	12,22	86,66
94,00 – 102,00	98,00	8	8,89	95,55
102,00 – 110,00	106,00	4	4,45	100,00

TABELA 3.4 – TABELA DE FREQÜÊNCIAS ABSOLUTA, RELATIVA E RELATIVA ACUMULADA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE

Intervalo de Classe	Ponto Médio	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Relativa Acumulada (%)
418,0 – 432,0	425,0	3	2,50	2,50
432,0 – 446,0	439,0	8	6,67	9,17
446,0 – 460,0	453,0	17	14,17	23,34
460,0 – 474,0	467,0	30	25,00	48,34
474,0 – 488,0	481,0	32	26,66	75,00
488,0 – 502,0	495,0	17	14,17	89,17
502,0 – 516,0	509,0	9	7,50	96,67
516,0 – 530,0	523,0	4	3,33	100,00

A TABELA 3.3 fornece um quadro global de como os dados de altura de planta de sorgo granífero estão distribuídos pelos intervalos de classe. Nota-se que as observações variam de 46,00 até 110,00 cm, com relativamente poucas medidas nas extremidades do intervalo e uma grande proporção dos valores situados entre 62,00 e 94,00 cm. Os intervalos 70,00 – 78,00 cm e 78,00 – 86,00 cm contém as maiores freqüências, ou seja, 24 plantas de sorgo granífero que corresponde a 26,67%, e 22 plantas de sorgo granífero que corresponde a 24,44%, respectivamente. Por outro lado, metade das plantas de sorgo granífero (50,00%) tem uma altura menor ou igual a 78,00 cm.

A TABELA 3.4 também fornece um quadro global de como os dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade estão distribuídos pelos intervalos de classe. As observações variam de 418,0 até 530,0 g, com relativamente poucas medidas nas extremidades do intervalo e uma grande proporção dos dados situados entre 446,0 e 502,0 g. Os intervalos 460,0 – 474,0 g e 474,0 – 488,0 g contém as maiores freqüências, ou seja, 30 frangos de corte que corresponde a 25,00%, e 32 frangos de corte que corresponde a 26,66%, respectivamente. Por outro lado, aproximadamente metade do lote misto de frangos de corte (48,34%) tem um peso corporal menor ou igual a 474,0 g.

Pelo visto, as TABELAS 3.3 e 3.4 proporcionam um entendimento muito melhor dos dados que as TABELAS 3.1 e 3.2, fornecendo informações importantes que auxiliam a entender a distribuição de altura de plantas de sorgo granífero e a distribuição de peso corporal de um lote misto de frangos de corte com 15 dias de idade.

3.1.2 Histograma de freqüência

O histograma de freqüência, tipo de gráfico mais comumente usado, também proporciona ao pesquisador um meio eficaz de organização dos dados para estudo do comportamento de variáveis de interesse. Embora freqüentemente forneçam menor grau de detalhe que as tabelas de freqüências, são mais fáceis de ler, proporcionando ao pesquisador um ganho no entendimento dos dados.

Esse tipo de gráfico consiste em um conjunto de retângulos que tem as bases sobre um eixo horizontal (eixo dos X) com centro no ponto médio e as larguras iguais às amplitudes dos intervalos das classes; e o eixo vertical (eixo dos Y), as áreas proporcionais às freqüências das classes, podendo ser as freqüências absolutas ou relativas.

Para a construção de um histograma de freqüência, inicialmente traçam-se as escalas dos eixos. A escala do eixo vertical deve começar do zero; se isso não é feito, as comparações visuais entre os intervalos podem ficar distorcidas. Uma vez que os eixos tenham sido desenhados, uma barra vertical centrada no ponto médio é colocada sobre cada intervalo. A altura da barra demarca a freqüência associada com o intervalo.

Como exemplo têm-se os histogramas de freqüências absoluta e relativa de altura de planta (cm) de sorgo granífero (FIGURAS 3.1 e 3.2) e de peso corporal (g) de um lote misto de frangos de corte com 15 dias de idade (FIGURAS 3.3 e 3.4).

FIGURA 3.1 – HISTOGRAMA DE FREQÜÊNCIA ABSOLUTA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

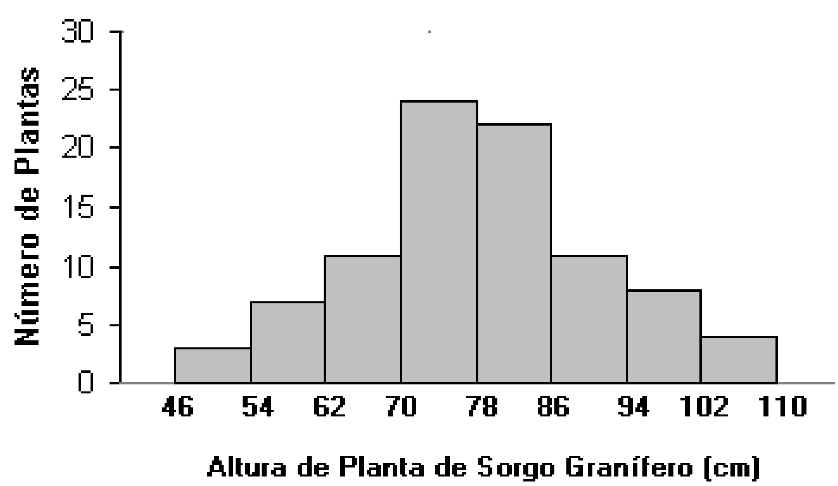


FIGURA 3.2 – HISTOGRAMA DE FREQÜÊNCIA RELATIVA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

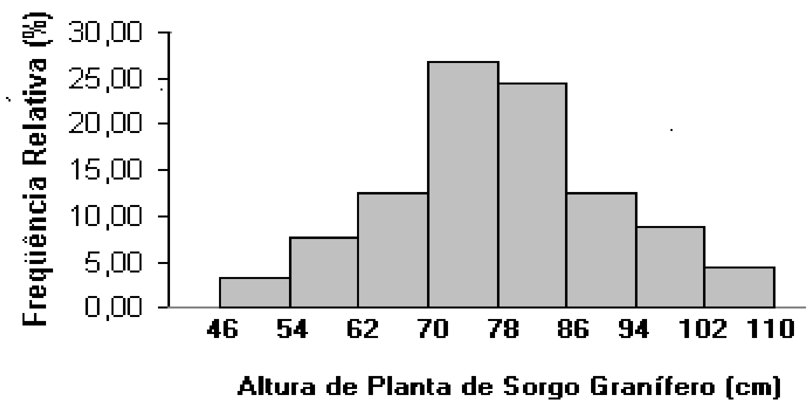


FIGURA 3.3 – HISTOGRAMA DE FREQÜÊNCIA ABSOLUTA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE

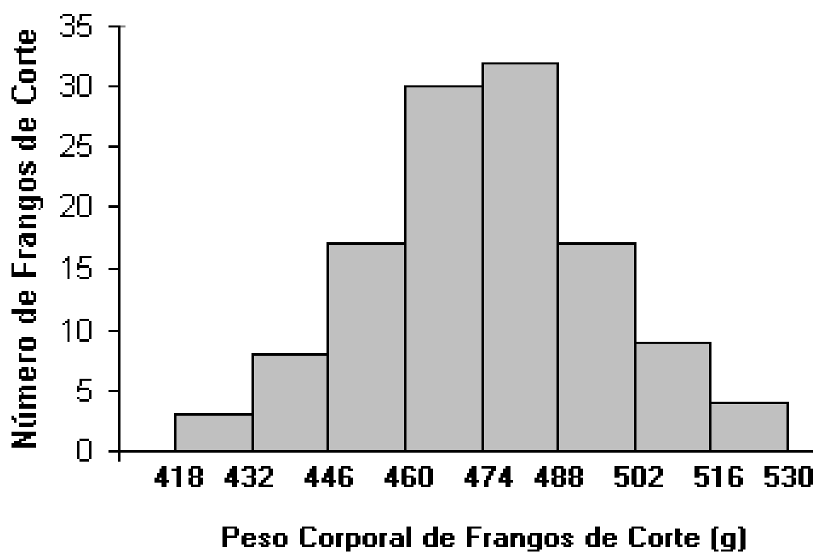
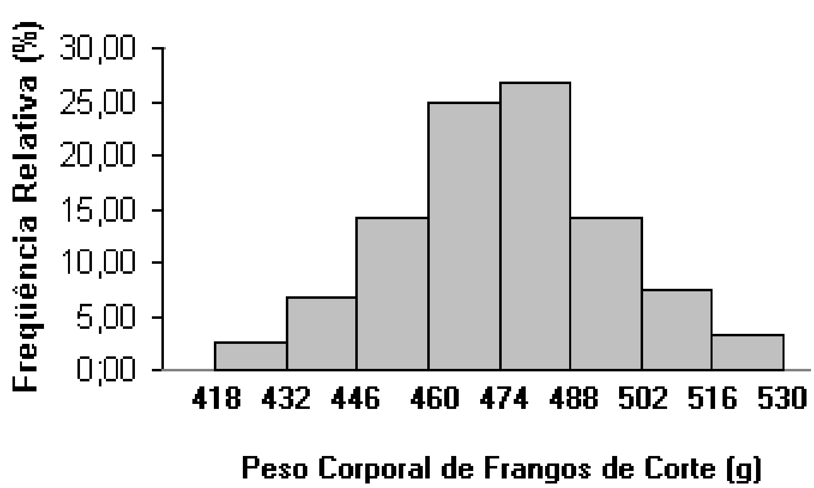


FIGURA 3.4 – HISTOGRAMA DE FREQÜÊNCIA RELATIVA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE



As FIGURAS 3.1 e 3.2 fornecem as mesmas informações da TABELA 3.3 para as freqüências absoluta e relativa da altura de planta de sorgo granífero, enquanto que as FIGURAS 3.3 e 3.4 fornecem as mesmas informações da TABELA 3.4 para as freqüências absoluta e relativa do peso corporal de um lote misto de frangos de corte com 15 dias de idade. Pelo visto, as FIGURAS 3.1 e 3.2 e as FIGURAS 3.3 e 3.4 proporcionam um entendimento melhor dos dados que as TABELAS 3.3 e 3.4, facilitando o entendimento da distribuição de altura de plantas de sorgo granífero e da distribuição de peso corporal de um lote misto de frangos de corte com 15 dias de idade.

3.1.3 Polígono de freqüência

O polígono de freqüência, gráfico de linha comumente usado, é muito semelhante ao histograma de freqüência, pois usa os mesmos dois eixos que um histograma de freqüência e transmitem essencialmente as mesmas informações quando são usadas as freqüências absolutas ou relativas. A diferença básica entre o histograma e o polígono de freqüência está no fato de este utilizar os pontos médios das classes, enquanto o histograma considera os limites reais das classes. Por outro lado, os polígonos de freqüência, por poderem ser facilmente superpostos, são superiores aos histogramas quando se quer comparar dois ou mais conjuntos de dados.

Para a construção de um polígono de freqüência, tanto para freqüência absoluta como para freqüência relativa, basta apenas unir os pontos médios de cada classe de um histograma de freqüência, conforme FIGURAS 3.5 e 3.6, para altura de planta (cm) de sorgo granífero e FIGURAS 3.7 e 3.8, para peso corporal (g) de um lote misto de frangos de corte com 15 dias de idade.

FIGURA 3.5 – POLÍGONO DE FREQÜÊNCIA ABSOLUTA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

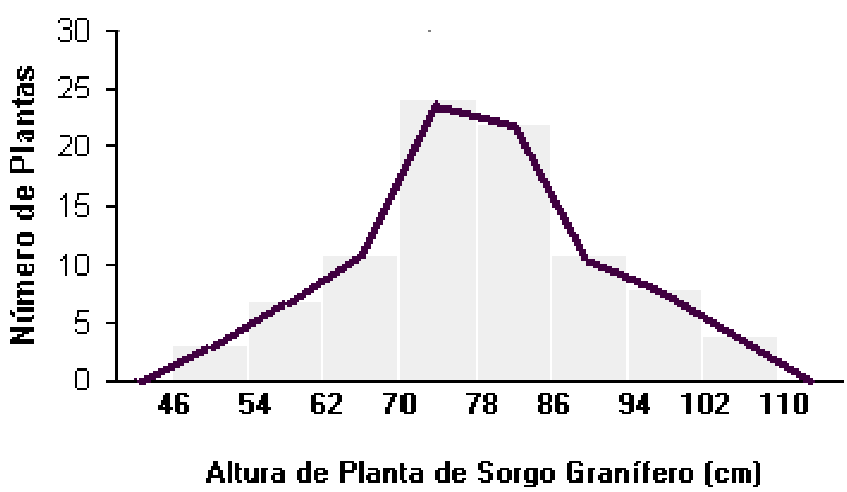


FIGURA 3.6 – POLÍGONO DE FREQÜÊNCIA RELATIVA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

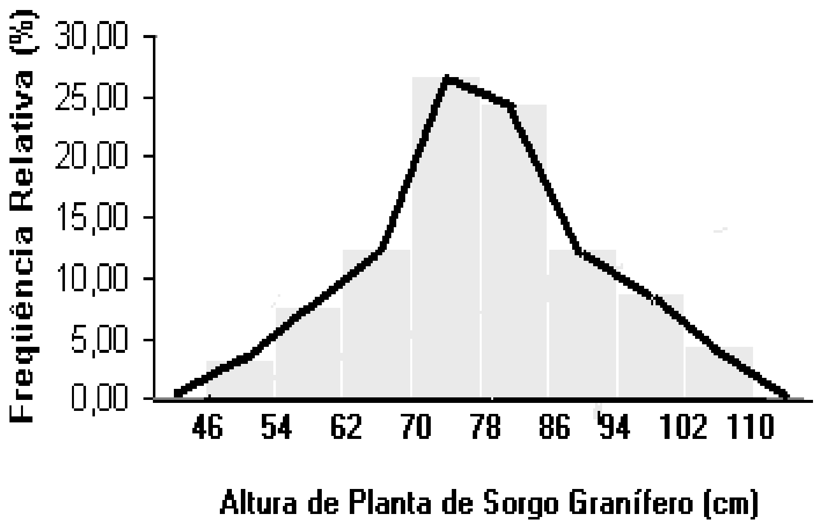


FIGURA 3.7 – POLÍGONO DE FREQÜÊNCIA ABSOLUTA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE

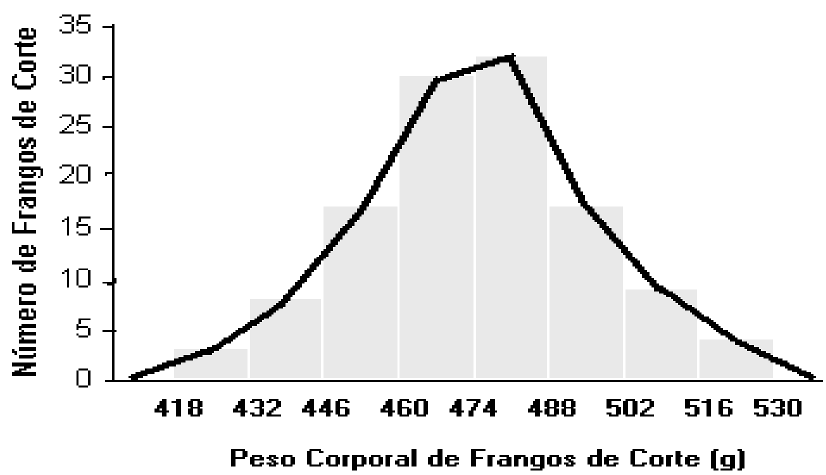
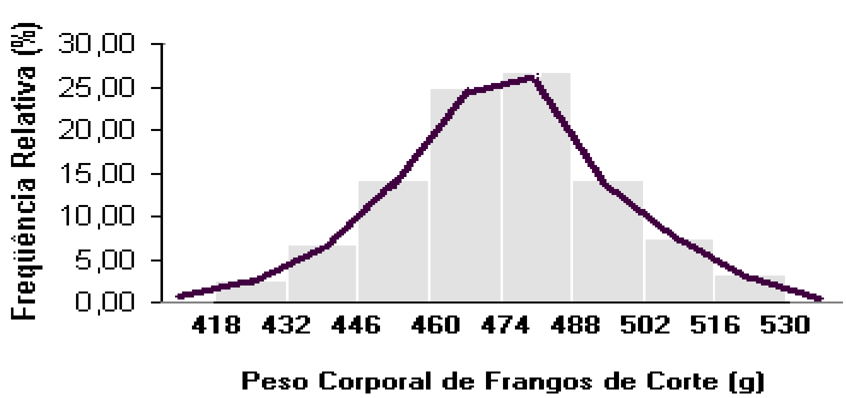


FIGURA 3.8 – POLÍGONO DE FREQÜÊNCIA RELATIVA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE



As FIGURAS 3.5 e 3.6 fornecem essencialmente as mesmas informações das FIGURAS 3.1 e 3.2 para as freqüências absoluta e relativa da altura de planta de sorgo granífero.

As FIGURAS 3.7 e 3.8 também fornecem essencialmente as mesmas informações das FIGURAS 3.3 e 3.4 para as freqüências absoluta e relativa do peso corporal de um lote misto de frangos de corte com 15 dias de idade.

Pelo visto, as FIGURAS 3.5 e 3.6 e as FIGURAS 3.7 e 3.8 proporcionam o mesmo entendimento dos dados da distribuição de altura de plantas de sorgo granífero e da distribuição de peso corporal de um lote misto de frangos de corte com 15 dias de idade que as FIGURAS 3.1 e 3.2 e as FIGURAS 3.3 e 3.4, respectivamente.

Também, para a construção de um polígono de freqüência pode-se usar a freqüência relativa acumulada, o qual é chamado de polígono de freqüência relativa acumulada ou Ogiva de Galton. Embora seu eixo horizontal seja o mesmo de um polígono de freqüência padrão, o seu eixo vertical utiliza-se das freqüências relativas acumuladas. Um ponto é colocado no limite superior de cada intervalo de classe; a altura do ponto representa a freqüência relativa acumulada associada ao intervalo de classe. Os pontos são então conectados por linhas retas. Como os polígonos de freqüência, os polígonos de freqüência relativa acumulada podem ser usados para comparar conjuntos de dados.

Como exemplos têm-se os polígonos de freqüência relativa acumulada de altura de planta (cm) de sorgo granífero (FIGURA 3.9) e de peso corporal (g) de um lote misto de frangos de corte com 15 dias de idade (FIGURA 3.10).

FIGURA 3.9 – POLÍGONO DE FREQÜÊNCIA RELATIVA ACUMULADA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

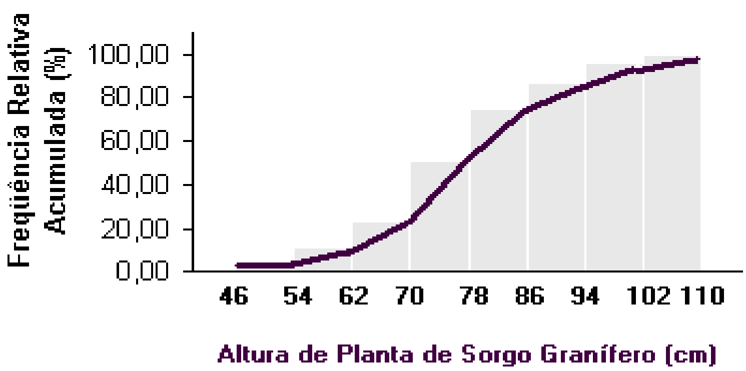
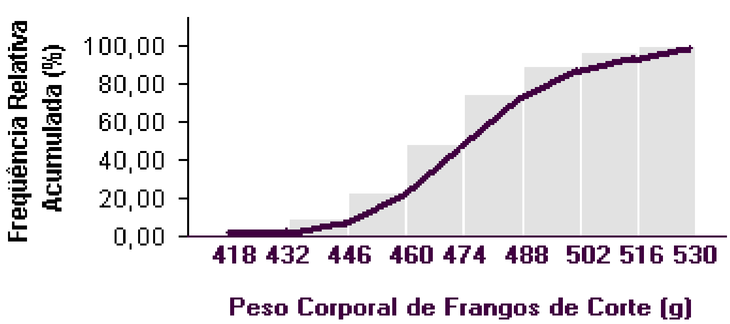


FIGURA 3.10 – POLÍGONO DE FREQÜÊNCIA RELATIVA ACUMULADA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE



As FIGURAS 3.9 e 3.10 fornecem as mesmas informações das TABELAS 3.3 e 3.4 para, respectivamente, as freqüências relativas acumuladas de altura de planta de sorgo granífero e de peso corporal de um lote misto de frangos de corte com 15 dias de idade, porém proporcionam um entendimento melhor dos dados que as referidas tabelas.

3.1.4 Curva normal

Se fossem construídos gráficos a partir de freqüências, por exemplo, do número de frutos por planta de 200 progênies de pimentão, de leituras refractométricas de diversas cebolas, da altura de planta de sorgo granífero, do peso corporal de frangos de corte, da produção de leite de vacas leiteiras, etc., os mesmos mostrariam diversas características importantes em comum. Todas as curvas teriam seu ponto mais alto próximo ao meio, representando a classe mais comum. Estas poderiam desviar-se bastante, simetricamente, sobre qualquer de seus lados em direção às classes raras. A maioria dos dados biológicos apresenta curva deste tipo, conhecida como **curva normal**, representadas pelas FIGURAS 3.11 e 3.12.

FIGURA 3.11 – CURVA NORMAL DA DISTRIBUIÇÃO DE FREQÜÊNCIA DE ALTURA DE PLANTA (cm) DE SORGO GRANÍFERO

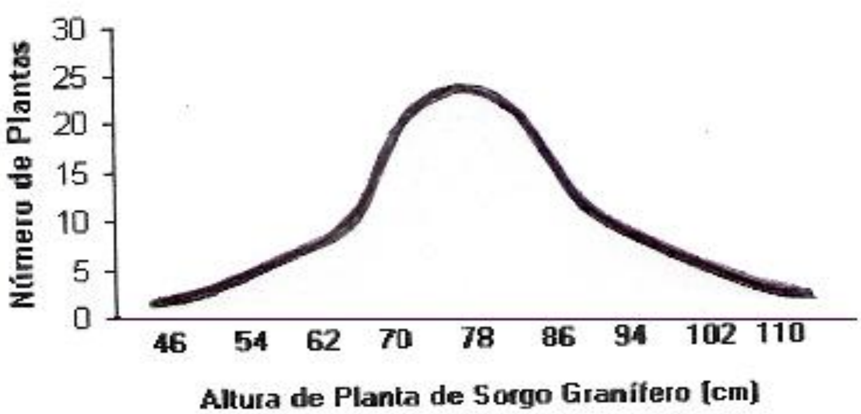
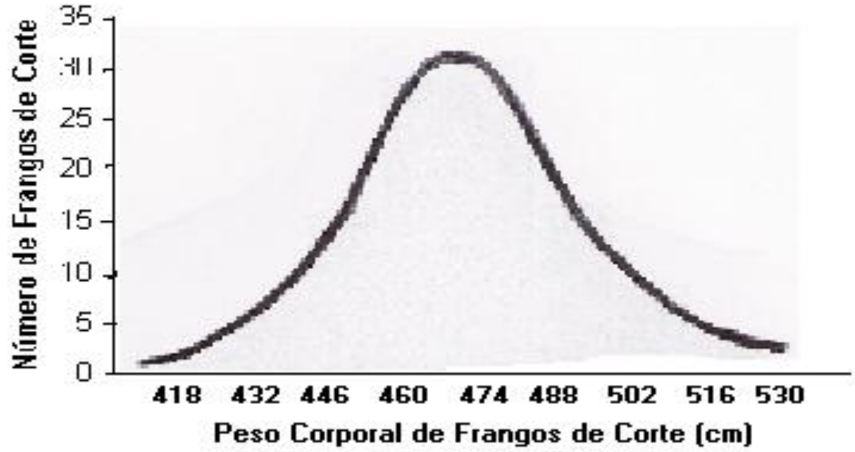


FIGURA 3.12 – CURVA NORMAL DA DISTRIBUIÇÃO DE FREQÜÊNCIA DE PESO CORPORAL (g) DE UM LOTE MISTO DE FRANGOS DE CORTE COM 15 DIAS DE IDADE



As curvas de distribuição normal podem diferir quanto à posição do ponto médio (o ponto de maior freqüência) e à dispersão dos dados, conforme FIGURAS 3.11 e 3.12, porém todas podem ser descritas, somente, mediante os parâmetros **média** e **desvio padrão.** Os métodos de estimá-los serão descritos nas seções seguintes.

3.2 Medidas de Tendência Central

Após serem os dados tabulados, é necessário encontrar valores típicos que possam representar a distribuição como um todo. Esses valores tendem a se localizar em um ponto central e reproduzirá as características da população, quanto mais homogêneos forem os seus componentes.

Esses valores são chamados de **medidas de tendência central** ou **medidas de posição**.

Entre as medidas de tendência central de uma distribuição de freqüência, as mais conhecidas são: a **média**, a **mediana** e a **moda**.

3.2.1 Média

A média é a mais importante das medidas de tendência central. Entre os vários tipos de médias, a **média aritmética,** ou simplesmente média, é a que mais nos interessa, do ponto de vista estatístico, por ser a mais representativa de uma amostra de dados. Ela apresenta as seguintes características:

a) É medida exata e rigorosamente definida;

b) Como medida de tendência central é de fácil compreensão e descreve todos os dados da série;

c) Serve de apoio a cálculos posteriores como o das probabilidades, desvio padrão, coeficiente de variação, etc.;

d) É a medida de tendência central de maior emprego no campo da análise quantitativa.

A média aritmética pode ser **simples** ou **ponderada**. Quando nada se especifica, significa estar-se tratando de média simples.

Numa série de dados não agrupados, isto é, dados que não estejam relacionados com distribuições de freqüências, a média aritmética simples é a razão entre o somatório dos valores da amostra ($\sum Xi$) e o número de observações (N).

Assim, numa amostra de dados $X_1, X_2, \ldots X_N$, tem-se:

$$\hat{m} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

$$= \frac{\sum X_i}{N}$$

Deve-se distinguir, neste caso, a média verdadeira, que é obtida quando se tomam todos os dados de uma população, e a média estimada, que é obtida a partir de dados de uma amostra.

Exemplo 1: Calcular a média aritmética simples a partir de dados da TABELA 3.5.

TABELA 3.5 – DADOS DE PRODUTIVIDADE (kg/ha) DE ALGODÃO HERBÁCEO, VARIEDADE ALLEN – 333/57, NO MUNICÍPIO DE VIÇOSA-AL, NO ANO DE 1977

Área	Produtividade (kg/ha)
1	273,0
2	660,0
3	675,0
4	355,0
5	315,0
6	453,0

FONTE: FERREIRA (1977).

A média será:

$$\hat{m} = \frac{\sum X}{N}$$

$$= \frac{273{,}0 + 660{,}0 + ... + 453{,}0}{6}$$

$$= \frac{2.731{,}0}{6} \cong \mathbf{455{,}17\ kg/ha}$$

O valor $\hat{m} \cong 455{,}17$ kg/ha é uma estimativa de produtividade da população de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, que é desconhecida.

Exemplo 2: Calcular a média aritmética simples a partir de dados da TABELA 3.6.

TABELA 3.6 – DADOS DE PESO AO NASCER (kg) DE BEZERROS MACHOS DA RAÇA CHAROLESA

Bezerro	Peso ao Nascer (kg)
1	47,0
2	41,0
3	34,0
4	45,0
5	45,0
6	46,0
7	25,0
8	48,0
9	37,0
10	47,0
11	40,0
12	40,0

FONTE: GOMES (1985).

A média será:

$$\hat{m} = \frac{\sum X}{N}$$

$$= \frac{47{,}0 + 41{,}0 + ... + 40{,}0}{12}$$

$$= \frac{495{,}0}{12} = \mathbf{41{,}25\ kg}$$

Também, o valor $\hat{m} = 41{,}25$ kg é uma estimativa de peso ao nascer da população de bezerros machos da raça Charolesa que é desconhecida.

Numa série de dados grupados em classes, portanto numa distribuição de freqüência, a média aritmética simples é a razão entre o somatório dos produtos dos pontos médios pelas freqüências [$\sum$ (*Pm x f*)] e o somatório das freqüências ($\sum f$).

Assim, tem-se:

$$\hat{m} = \frac{\sum (Pm\ x\ f)}{\sum f}$$

Exemplo 3: Calcular a média aritmética simples a partir de dados da TABELA 3.3.

A média será:

$$\hat{m} = \frac{\sum (Pm\ x\ f)}{\sum f}$$

$$= \frac{(50{,}0\ x\ 3) + (58{,}0\ x\ 7) + ... + (106{,}0\ x\ 4)}{3 + 7 + ... + 4}$$

$$= \frac{(150{,}0) + (406{,}0) + ... + (424{,}0)}{90}$$

$$= \frac{7.060{,}0}{90} \cong \mathbf{78{,}44\ cm}$$

O valor $\hat{m} \cong 78{,}44$ cm é uma estimativa de altura de planta da população de sorgo granífero que é desconhecida. Esse valor, que se localiza em um ponto central, representa a amostra de 90 dados da distribuição de altura de planta de sorgo granífero.

Exemplo 4: Calcular a média aritmética simples a partir de dados da TABELA 3.4.

A média será:

$$\hat{m} = \frac{\sum (Pm \ x \ f)}{\sum f}$$

$$= \frac{(425{,}0 \ x \ 3) + (439{,}0 \ x \ 8) + ... + (523{,}0 \ x \ 4)}{3 + 8 + ... + 4}$$

$$= \frac{(1.275{,}0) + (3.512{,}0) + ... + (2.092{,}0)}{120}$$

$$= \frac{56.978{,}0}{120} \cong \mathbf{474{,}82 \ g}$$

O valor $\hat{m} \cong 474{,}82$ g é uma estimativa de peso corporal de uma população mista de frangos de corte com 15 dias de idade que é desconhecida. Esse valor, que também se localiza em um ponto central, representa a amostra de 120 dados da distribuição de peso corporal de um lote misto de frangos de corte com 15 dias de idade.

Em certos casos não próprios de distribuições de freqüências, em que os dados não possuem identidade de significação, devem-se equiparar os dados entre si para obtenção da média aritmética. Esse tipo de média se chama, especificamente, média aritmética ponderada, ou às vezes, simplesmente, média ponderada. A ponderação é a única forma que proporciona um resultado capaz de traduzir a realidade.

Ponderar, significa pesar. Isto quer dizer que se devem pesar os dados para se obter a média, que será uma razão entre o somatório dos produtos de cada valor pelo peso respectivo [$\sum$ (P x X)] e o somatório dos pesos ($\sum P$).

Assim, tem-se:

$$\hat{m}p = \frac{\sum (P \ x \ X)}{\sum P}$$

Exemplo 5: Calcular a média aritmética ponderada a partir de dados da TABELA 3.7.

TABELA 3.7 – DADOS DE STAND FINAL E DE NÚMERO DE FRUTOS DE ABACAXI (*Ananas comosus* (L.) Merrill, VARIEDADE PÉROLA, EM ÁREAS DE 42 m^2, NO MUNICÍPIO DE ARAPIRACA-AL, NO ANO DE 1985

Lote	Stand Final	Número de Frutos
1	129,0	73,0
2	139,0	101,0
3	138,0	102,0
4	132,0	87,0
5	129,0	79,0
6	112,0	69,0

FONTE: FERREIRA e MARTINS, 1985.

A média será:

$$\hat{m}p = \frac{\sum(P\ x\ X)}{\sum P}$$

$$= \frac{(129{,}0\ x\ 73{,}0) + (139{,}0\ x\ 101{,}0) + ... + (112{,}0\ x\ 69{,}0)}{129{,}0 + 139{,}0 + ... + 112{,}0}$$

$$= \frac{(9.417{,}0) + (14.039{,}0) + ... + (7.728{,}0)}{779{,}0}$$

$$= \frac{66.935{,}0}{779{,}0} \cong \mathbf{85{,}92\ frutos}$$

O valor $\hat{m}p \cong 85{,}92$ frutos é uma estimativa de número de frutos por lote de 42 m^2 da população de abacaxi, variedade PÉROLA, no Município de Arapiraca-AL, que é desconhecida. Esse valor representa melhor a amostra de seis lotes de 42 m^2 da distribuição de número de frutos da população de abacaxi, variedade PÉROLA, no Município de Arapiraca-AL, do que a média aritmética simples ($\hat{m} \cong 85{,}17$ frutos), pois é levado em conta o número de plantas por lote no cálculo da média do número de frutos de abacaxi, enquanto que na média aritmética simples isso não ocorre. Sabe-se que numa mesma área, quanto maior o número de plantas maior será o número de frutos. Portanto, o valor $\hat{m}p \cong 85{,}92$ frutos traduz melhor a realidade.

Exemplo 6: Calcular a média aritmética ponderada a partir de dados da TABELA 3.8.

TABELA 3.8 – DADOS DE NÚMERO DE POEDEIRAS ISA BROWN POR PARCELA E DE NÚMERO DE OVOS PRODUZIDOS DURANTE UM PERÍODO DE 60 DIAS

Parcela	Número de Aves	Produção de Ovos
1	8,0	468,0
2	7,0	410,0
3	7,0	416,0
4	6,0	351,0
5	8,0	460,0

A média será:

$$\hat{m}p = \frac{\sum(P\ x\ X)}{\sum P}$$

$$= \frac{(8{,}0 \; x \; 468{,}0) + (7{,}0 \; x \; 410{,}0) + ... + (8{,}0 \; x \; 460{,}0)}{8{,}0 + 7{,}0 + ... + 8{,}0}$$

$$= \frac{(3.744{,}0) + (2.870{,}0) + ... + (3.680{,}0)}{36{,}0}$$

$$= \frac{15.312{,}0}{36{,}0} \cong \mathbf{425{,}33\ ovos}$$

O valor $\hat{m}p \cong 425{,}33$ ovos também é uma estimativa de número de ovos por parcela da população de poedeiras Isa Brown durante um período de 60 dias, que é desconhecida. Esse valor representa melhor a amostra de cinco parcelas da distribuição de número de ovos por parcela de poedeiras Isa Brown durante um período de 60 dias, do que a média aritmética simples ($\hat{m} \cong 421{,}0$ ovos), pois é levado em conta o número de aves por parcela no cálculo da média do número de ovos de poedeiras Isa Brown, enquanto que na média aritmética simples isso não ocorre. Sabe-se que numa mesma área, quanto maior o número de galinhas poedeiras maior será o número de ovos. Portanto, o valor $\hat{m}p \cong 425{,}33$ ovos traduz melhor a realidade.

3.2.2 Mediana

A mediana de um conjunto ordenado de dados é o valor que ocupa exatamente o centro da série ou a média aritmética dos dois valores centrais, sendo insensível ao valor de cada observação, o que pode ser uma vantagem, quando a distribuição dos dados for assimétrica. Esta medida de tendência central serve para representar e analisar uma série de dados grupados ou não, dividindo a série em duas partes iguais, isto é, forma uma dicotomia de área.

Numa série de dados não agrupados, a mediana é facilmente localizável, tanto quanto as demais medidas de tendência central. Neste caso específico, como foi dito, a mediana (m_e) ficará no centro da série.

Considerando os dados do Exemplo 1, a mediana será:

$$m_e = \frac{X_3 + X_4}{2}$$

$$= \frac{355{,}0 + 453{,}0}{2}$$

$$= \frac{808{,}0}{2} = \mathbf{404{,}0\ kg/ha}$$

O valor m_e = 404,0 kg/ha é uma estimativa de produtividade da população de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, que é desconhecida. Esse valor foi bem inferior ao valor da média aritmética ($\hat{m} \cong$ 455,17 kg/ha), tendo uma diferença de 51,17 kg/ha. Como houve uma variação muito grande entre os dados de produtividade de algodão herbáceo, onde o maior valor (675,0 kg/ha) foi, aproximadamente, 2,5 vezes maior que o menor valor (273,0 kg/ha) e a média aritmética é sensível a esse tipo de variação, a mediana, nesse caso, seria a medida de tendência central que traduz melhor a realidade por ser mais robusta, ou seja, muito menos sensível a esse tipo de variação.

Considerando, também, os dados do Exemplo 2, a mediana será:

$$m_e = \frac{X_6 + X_7}{2}$$

$$= \frac{41{,}0 + 45{,}0}{2}$$

$$= \frac{86{,}0}{2} = \mathbf{43{,}0\ kg}$$

O valor m_e = 43,0 kg é uma estimativa de peso ao nascer da população de bezerros machos da raça Charolesa que é desconhecida. Esse valor foi ligeiramente superior ao valor da média aritmética ($\hat{m}$ = 41,25 kg), tendo uma diferença de apenas 1,75 kg por animal. Nesse caso, tanto a mediana como à média aritmética traduzem a realidade.

Numa série de dados grupados em classes, a mediana (m_e) é obtida através da seguinte fórmula:

$$m_e = Li + \left(\frac{\frac{N}{2} - \sum f'}{fm_e} \right) \text{x Ic}$$

onde:
Li = limite inferior da classe mediana;
N = total de freqüência;
$\sum f'$ = soma de todas as freqüências das classes inferiores à mediana;
fm_e = freqüência da classe mediana;
Ic = amplitude do intervalo da classe mediana.

Considerando os dados do Exemplo 3, a mediana será:

$$m_e = Li + \left(\frac{\frac{N}{2} - \sum f'}{fm_e} \right) \text{x Ic}$$

$$= 74{,}0 + \left(\frac{\frac{90}{2} - 45}{23} \right) \text{x } 8{,}0$$

$$= 74{,}0 + \left(\frac{45 - 45}{23} \right) \text{x } 8{,}0$$

$$= 74{,}0 + \left(\frac{0}{23}\right) \text{x } 8{,}0$$

$$= 74{,}0 + (0) \text{ x } 8{,}0$$

$$= 74{,}0 + 0 = \mathbf{74{,}0\ cm}$$

O valor m_e = 74,0 cm é uma estimativa de altura de planta da população de sorgo granífero que é desconhecida. Esse valor foi ligeiramente inferior ao valor da média aritmética ($\hat{m} \cong$ 78,44 cm), tendo uma diferença de apenas 4,44 cm por planta. Nesse caso, tanto a mediana como à média aritmética traduzem a realidade.

Considerando, também, os dados do Exemplo 4, a mediana será:

$$m_e = Li + \left(\frac{\frac{N}{2} - \sum f'}{fm_e}\right) \text{x Ic}$$

$$= 467{,}0 + \left(\frac{\frac{120}{2} - 58}{31}\right) \text{x } 14{,}0$$

$$= 467{,}0 + \left(\frac{60 - 58}{31}\right) \text{x } 14{,}0$$

$$= 467{,}0 + \left(\frac{2}{31}\right) \text{x } 14{,}0$$

$$= 467{,}0 + (0{,}0645) \text{ x } 14{,}0$$

$$= 467{,}0 + 0{,}903 \cong \mathbf{467{,}90\ g}$$

O valor $m_e \cong$ 467,90 g é uma estimativa de peso corporal da população de um lote misto de frangos de corte com 15 dias de idade que é desconhecida. Esse valor foi ligeiramente inferior ao valor da média aritmética ($\hat{m} \cong$ 474,82 g), tendo uma diferença de apenas 6,92 g por frango. Também, nesse caso, tanto a mediana quanto a média aritmética traduzem a realidade.

3.2.3 Moda

A moda de um conjunto de dados é o valor que ocorre com maior freqüência, ou seja, é o valor mais comum. A moda pode não existir e, mesmo que exista, pode não ser única.

Numa série de dados não grupados, quando todos os valores da série ocorrem com a mesma freqüência, como no Exemplo 1, a moda (m_o) não existe. Quando a série possuir apenas um valor como sendo o mais freqüente, este será a moda, denominando-se unimodal. Contudo, quando a série possuir mais de um valor como sendo os mais freqüentes, ela pode possuir mais de uma moda, denominando-se bimodal, trimodal, etc..

Exemplo 7: Calcular a moda a partir dos dados da TABELA 3.9.

TABELA 3.9 – DADOS DE EMERGÊNCIA DE PLÂNTULAS, DE EMERGÊNCIA DA 1ª VAGEM E DE MATURAÇÃO DE VAGENS DE CULTIVARES DE SOJA (*Glicine max* (L.) Merrill) NO MUNICÍPIO DE VIÇOSA-AL, NO ANO DE 1984

Cultivar	Emergência de Plântulas (em dias)	Emergência da 1ªVagem (em dias)	Maturação de Vagens (em dias)
BOSSIER	6	45	93
BR – 2	5	36	85
FOSCARIN – 31	8	36	95
IAC – 2	6	42	97
IAC – 4	6	41	99
IAC – 6	5	42	112
IAC – 9	6	44	101
IAC – 10	6	42	101
IAC – 12	4	39	93
PARANÁ	5	35	85
PÉROLA	4	37	97
PLANALTO	4	37	109
PRATA	4	35	90
TROPICAL	6	54	117
UFV – 1	5	40	99
UFV – 4	5	37	95
UFV – 5	6	41	99
VIÇOJA	7	36	93

FONTE: FERREIRA e OLIVEIRA (1985).

No caso da emergência de plântulas, a moda será:

$$m_o = \textbf{6 dias}$$

No caso da emergência da 1ª vagem, as modas serão:

$$m_o = \textbf{36 dias}$$

$$m_o = \textbf{37 dias}$$

$$m_o = \textbf{42 dias}$$

E, no caso da maturação de vagens, as modas serão:

$$m_o = \textbf{93 dias}$$

$$m_o = \textbf{99 dias}$$

No caso da emergência de plântulas, o valor m_o = 6,0 dias é uma estimativa da população de soja no Município de Viçosa-AL, no ano de 1984, que é desconhecida. Esse valor foi bastante próximo do valores da média aritmética ($\hat{m} \cong$ 5,44 dias) e da mediana (m_e = 5,5 dias). Assim sendo, qualquer uma dessas medidas de tendência central traduz a realidade quanto à emergência de plântulas de soja em Viçosa-AL.

Também, no caso da emergência da 1ª vagem, os valores: m_o = 36,0 dias, m_o = 37,0 dias e m_o = 42,0 dias são estimativas da população de soja no Município de Viçosa-AL, no ano de 1984, que são desconhecidas. Esses valores foram bastante diferentes do valor da média aritmética ($\hat{m} \cong$ 39,94 dias) e do valor da mediana (m_e = 39,5 dias). Por outro lado, a média aritmética e a mediana apresentaram valores bem próximos. Desse modo, apenas a média aritmética e a mediana, nesse caso, como medidas de tendência central, traduzem melhor a realidade quanto à emergência da 1ª vagem de soja em Viçosa-AL.

Ainda, no caso de maturação de vagens, os valores: m_o = 93,0 dias e m_o = 99,0 dias são estimativas da população de soja no Município de Viçosa-AL, no ano de 1984, que são desconhecidas. Esses valores foram bastante diferentes do valor da média aritmética ($\hat{m} \cong$ 97,78 dias) e do valor da mediana (m_e = 97,0 dias). Por outro lado, a média aritmética e a mediana apresentaram valores bem próximos. Desse modo, como no caso anterior, apenas a média aritmética e a mediana, como medidas de tendência central, traduzem melhor a realidade quanto à maturação de vagens de soja em Viçosa-AL.

Considerando os dados do Exemplo 2, as modas serão:

m_0 = **40,0 kg**

m_0 = **45,0 kg**

m_0 = **47,0 kg**

Os valores: m_o = 40,0 kg, m_o = 45,0 kg e m_o = 47,0 kg são estimativas de peso ao nascer da população de bezerros machos da raça Charolesa que são desconhecidas. Esses valores foram bastante diferentes do valor da média aritmética ($\hat{m}$ = 41,25 kg) e do valor da mediana (m_e = 43,0 kg). Por outro lado, a média aritmética e a mediana apresentaram valores bem próximos. Então, dessa forma, tanto a média aritmética quanto à mediana traduzem melhor a realidade em relação ao peso ao nascer de bezerros machos da raça Charolesa.

Numa série de dados grupados em classes, chama-se classe modal a classe que possui a maior freqüência. Neste caso, existem vários processos para se determinar à moda (m_o). Contudo, serão vistos os mais utilizados:

a) **Processo de KING** – A moda (m_o) é calculada através da seguinte fórmula:

$$m_o = Li' + \left(\frac{fp}{fa + fp}\right) \text{x Ic}$$

onde:
Li’= limite inferior da classe modal;
fp = freqüência posterior à classe modal;

fa = freqüência anterior à classe modal;
Ic = amplitude do intervalo da classe modal.

Considerando os dados do Exemplo 3, a moda será:

$$m_o = Li' + \left(\frac{fp}{fa + fp}\right) \text{x Ic}$$

$$= 70{,}0 + \left(\frac{22}{11+22}\right) \text{x } 8{,}0$$

$$= 70{,}0 + \left(\frac{22}{33}\right) \text{x } 8{,}0$$

$$= 70{,}0 + (0{,}6667) \text{ x } 8{,}0$$

$$= 70{,}0 + 5{,}3336 \cong \mathbf{75{,}33\ cm}$$

Considerando, também, os dados do Exemplo 4, a moda será:

$$m_o = Li' + \left(\frac{fp}{fa + fp}\right) \text{x Ic}$$

$$= 474{,}0 + \left(\frac{17}{30+17}\right) \text{ x } 14{,}0$$

$$= 474{,}0 + \left(\frac{17}{47}\right) \text{ x } 14{,}0$$

$$= 474{,}0 + (0{,}3617) \text{ x } 14{,}0$$

$$= 474{,}0 + 5{,}0638 \cong \mathbf{479{,}06\ g}$$

b) **Processo de CZUBER** - A moda (m_o) é calculada através da seguinte fórmula:

$$m_o = Li' + \left[\frac{fm_o - fa}{2\ x\ fm_o - (fa + fp)}\right] \text{x Ic}$$

onde:
Li'= limite inferior da classe modal;
fp = freqüência posterior à classe modal;
fa = freqüência anterior à classe modal;
fm_o= freqüência da classe modal;
Ic = amplitude do intervalo da classe modal.

Considerando os dados do Exemplo 3, a moda será:

$$m_o = Li' + \left[\frac{fm_o - fa}{2\ x\ fm_o - (fa + fp)}\right] x\ Ic$$

$$= 70{,}0 + \left[\frac{24 - 11}{2\ x\ 24 - (11 + 22)}\right] x\ 8{,}0$$

$$= 70{,}0 + \left[\frac{13}{48 - (33)}\right] x\ 8{,}0$$

$$= 70{,}0 + \left[\frac{13}{15}\right] x\ 8{,}0$$

$$= 70{,}0 + (0{,}8667)\ x\ 8{,}0$$

$$= 70{,}0 + 6{,}93360 = \mathbf{76{,}93\ cm}$$

Considerando, também, os dados do Exemplo 4, a moda será:

$$m_o = Li' + \left[\frac{fm_o - fa}{2\ x\ fm_o - (fa + fp)}\right] x\ Ic$$

$$= 474{,}0 + \left[\frac{32 - 30}{2\ x\ 32 - (30 + 17)}\right] x\ 14{,}0$$

$$= 474{,}0 + \left[\frac{2}{64 - (47)}\right] x\ 14{,}0$$

$$= 474{,}0 + \left[\frac{2}{17}\right] x\ 14{,}0$$

$$= 474{,}0 + (0{,}1176)\ x\ 14{,}0$$

$$= 474{,}0 + 1{,}6464 \cong \mathbf{475{,}65\ g}$$

Observe-se que há uma diferença entre os valores encontrados, por ambos os processos, tanto para altura de planta de sorgo granífero (Processo de King – $m_o \cong 75{,}33$ cm e Processo de Czuber – $m_o = 76{,}93$ cm) quanto para peso corporal de um lote misto de frangos de corte com 15 dias de idade (Processo de King – $m_o \cong 479{,}06$ g e Processo de Czuber – $m_o \cong 475{,}65$ g), mas que em termos de moda não tem importância.

Por outro lado, as estimativas da moda, pelo Processo de Czuber, para os dois tipos de distribuição de freqüência ficaram mais próximas das estimativas da média aritmética (Altura de planta de sorgo granífero: $\hat{m} \cong 78{,}44$ cm e $m_o \cong 76{,}93$ cm; Peso corporal de um lote misto de frangos de corte com 15 dias de idade: $\hat{m} \cong$ 474,82 g e $m_o \cong$ 475,65 g), enquanto que pelo Processo de King, apenas a estimativa da moda da Altura de planta de sorgo granífero (m_o = 75,33cm) ficou próxima da mediana (m_e = 74,0 cm).

Assim sendo, como as estimativas da média aritmética e da mediana para os dois tipos de distribuição em estudo foram muito próximas e que as estimativas da moda pelos dois processos estão em torno delas apenas para Altura de planta de sorgo granífero, enquanto que para o Peso corporal de um lote misto de frangos de corte com 15 dias de idade as estimativas da moda pelos dois processos foram muito próximas das estimativas da média aritmética, qualquer uma das medidas de tendência central traduz a realidade para altura de planta de sorgo granífero e para peso corporal de um lote misto de frangos de corte com 15 dias de idade.

Por fim, vale ressaltar que na pesquisa agropecuária, as medidas de tendência central são utilizadas, de um modo geral, isoladamente, cabendo ao pesquisador verificar qual delas é mais conveniente para auxiliar a análise dos seus dados. Entretanto, em determinadas situações, elas podem ser utilizadas em conjunto.

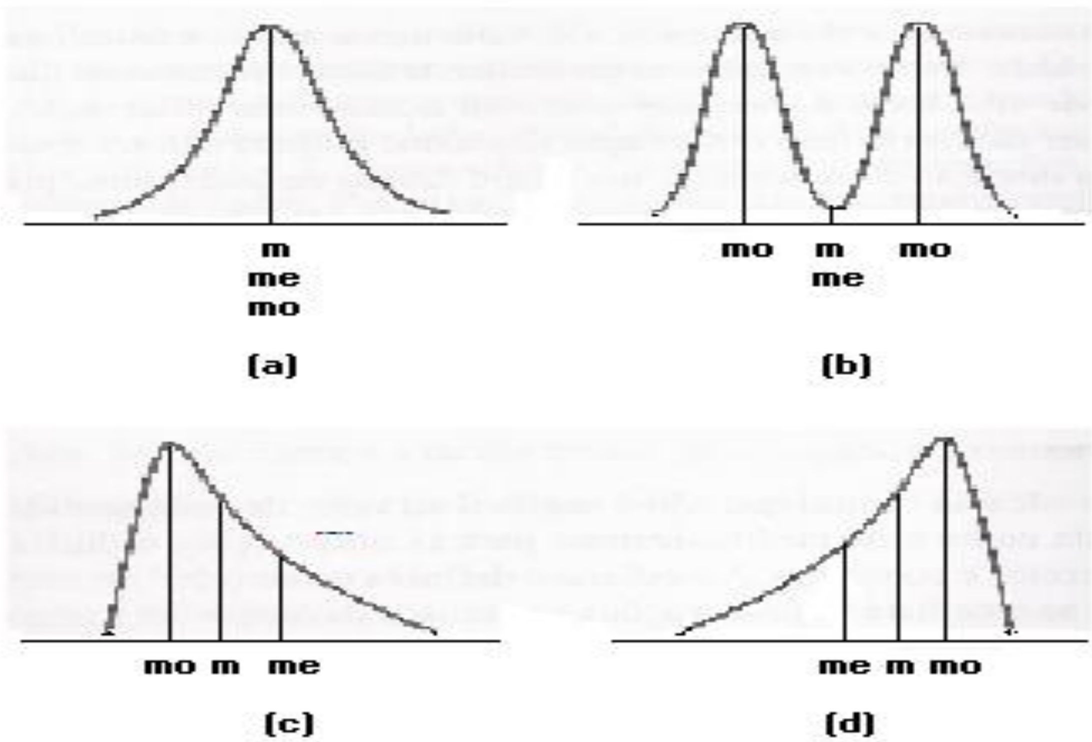
A melhor medida de tendência central para um determinado conjunto de dados depende freqüentemente da distribuição dos valores:

a) Se a distribuição de valores é simétrica e unimodal, a média, a mediana e a moda são aproximadamente as mesmas, onde, nesta situação, qualquer uma delas poderá ser usada convenientemente para analisar os dados, conforme FIGURA 3.13 (a).

b) Se a distribuição de valores é simétrica e bimodal, a média e a mediana são aproximadamente as mesmas, porém não convenientes para analisar os dados, pois se tratam de medidas improváveis de ocorrer, já que seus valores se encontrariam entre os dois picos, segundo FIGURA 3.13 (b). Uma distribuição bimodal indica freqüentemente que a população da qual os valores são tomados consiste realmente de dois subgrupos distintos que diferem na característica medida, onde a moda seria a medida de tendência central mais conveniente para analisar os dados ou então analisar os dois subgrupos separadamente.

c) Quando os dados são assimétricos, tanto à direita quanto à esquerda, a mediana é freqüentemente a melhor medida de tendência central. Por ser sensível às observações extremas, a média é puxada em direção dos valores atípicos e, conseqüentemente, poderia terminar excessivamente aumentada ou reduzida, em excesso. Quando os dados são assimétricos à direita, a média se encontra à direita da mediana [FIGURA 3.13 (c)]; e quando os dados são assimétricos à esquerda, a média se encontra à esquerda da mediana [FIGURA 3.13 (d)].

FIGURA 3.13 – TIPOS DE DISTRIBUIÇÃO DE VALORES DE UM DETERMINADO CONJUNTO DE DADOS: (a) SIMÉTRICA E UNIMODAL; (b) SIMÉTRICA E BIMODAL; (c) ASSIMÉTRICA À DIREITA; (d) ASSIMÉTRICA À ESQUERDA



3.3 Medidas de Variabilidade de Dados

Na seção anterior, foi visto que entre as medidas de tendência central, a média é a mais importante, do ponto de vista estatístico, por ser a mais representativa de uma amostra de dados. Contudo, ela não diz como os dados de uma amostra se distribuem em torno dela. Por exemplo, sejam as seguintes amostras de dados:

(1) 10, 10, 10, 10, 10 $\hat{m}$ = **10,0**

(2) 8, 10, 12, 9, 11 $\hat{m}$ = **10,0**

(3) 10, 3, 9, 17, 11 $\hat{m}$ = **10,0**

(4) 17, 15, 7, 3, 8 $\hat{m}$ = **10,0**

Ver-se que as amostras (1), (2), (3) e (4) têm a mesma média, mas observa-se que na amostra (1) todos os valores são iguais a 10, ou seja, igual a média aritmética, logo todos os valores estão concentrados na média, não existindo qualquer diferença entre cada valor e a média, consequentemente não existe variabilidade dos dados, o que na prática é improvável de ocorrer. Ao passo que nas outras amostras existem diferenças em relação à média. Assim pode-se dizer que na mostra (1) não existe variabilidade nos dados, havendo para todas as outras, sendo a amostra (4) a de maior variabilidade.

Portanto, além da média, necessita-se de uma medida estatística complementar para melhor caracterizar cada amostra apresentada.

As medidas estatísticas responsáveis pela variação ou dispersão dos valores de uma série são as **medidas de variabilidade** ou **medidas de dispersão,** onde se destacam, em nosso caso, a **amplitude total,** a **variância,** o **desvio padrão,** o **erro padrão da média** e o **coeficiente de variação**.

3.3.1 Amplitude total

A amplitude total (At) é a diferença entre os valores maior (ma) e menor (me) de um conjunto de dados de uma determinada variável.

Assim, numa amostra de dados X_1, X_2, ... , X_N, tem-se:

$$At = X_{ma} - X_{me}$$

Considerando todas as amostras com média $\hat{m} = 10$, do exemplo citado anteriormente, ver-se que a média $\hat{m} = 10$ não dá por si só, uma completa informação a respeito do comportamento dos dados. Entretanto, se for tomado a diferença entre o maior e o menor deles, dentro de cada amostra, isto é, a amplitude total, ter-se-á respectivamente:

$$At_{(1)} = X_{ma} - X_{me}$$

$$= 10 - 10 = \mathbf{0{,}0}$$

$$At_{(2)} = X_{ma} - X_{me}$$

$$= 12 - 8 = \mathbf{4{,}0}$$

$$At_{(3)} = X_{ma} - X_{me}$$

$$= 17 - 3 = \mathbf{14{,}0}$$

$$At_{(4)} = X_{ma} - X_{me}$$

$$= 17 - 3 = \mathbf{14{,}0}$$

De imediato conclui-se que as amostras (3) e (4) são as mais dispersas. No entanto, elas são bem distintas, faltando, consequentemente, alguma informação a mais, que permita diferenciá-las.

É por isso que a amplitude total, mesmo sendo fácil de calcular, é uma medida de dispersão de utilidade limitada, por depender somente dos valores extremos de um conjunto de dados, desprezando assim os valores intermediários, o que a torna insensível à dispersão dos demais valores entre o maior e o menor.

Considerando os dados do Exemplo 1, a amplitude total será:

$$At = X_{ma} - X_{me}$$

$$= 675{,}0 - 273{,}0 = \mathbf{402{,}0\ kg/ha}$$

O valor At = 402,0 kg/ha é uma estimativa de variabilidade dos dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, em torno da média aritmética ($\hat{m} \cong 455{,}17$ kg/ha). Houve uma variação muito grande nos dados de produtividade de algodão herbáceo, em relação à média aritmética.

Considerando, também, os dados do Exemplo 2, a amplitude total será:

$$At = X_{ma} - X_{me}$$

$$= 48{,}0 - 25{,}0 = \mathbf{23{,}0\ kg}$$

O valor At = 23,0 kg é uma estimativa de variabilidade dos dados de peso ao nascer de bezerros machos da raça Charolesa, em torno da média aritmética ($\hat{m}$ = 41,25 kg). Houve uma variação grande nos dados de peso ao nascer de bezerros machos, em relação à média aritmética.

Considerando, ainda, os dados do Exemplo 3, a amplitude total será:

$$At = X_{ma} - X_{me}$$

$$= 110{,}0 - 46{,}0 = \mathbf{64{,}0\ cm}$$

O valor At = 64,0 cm é uma estimativa de variabilidade dos dados de altura de planta de sorgo granífero, em torno da média aritmética ($\hat{m} \cong$ 78,44 cm). Houve uma variação muito grande nos dados de altura de planta, em relação à média aritmética.

Considerando, por fim, os dados do Exemplo 4, a amplitude total será:

$$At = X_{ma} - X_{me}$$

$$= 530{,}0 - 418{,}0 = \mathbf{112{,}0\ g}$$

O valor At = 112,0 g é uma estimativa de variabilidade dos dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade, em torno da média aritmética ($\hat{m} \cong$ 474,82 g). Houve uma variação relativamente grande nos dados de peso corporal, em relação à média aritmética.

3.3.2 Variância

A variância é uma medida de variabilidade que leva em conta todos os valores de um conjunto de dados. É, indiscutivelmente, a melhor medida de dispersão.

Numa amostra de dados não grupados, como por exemplo, numa amostra de dados X_1, X_2, ... , X_N, a variância (s^2) é obtida através da seguinte fórmula:

$$s^2 = \frac{SQD}{N-1}$$

onde:
SQD = soma dos quadrados dos desvios em relação à média aritmética;
N = número de observações.

É oportuno observar que o denominador da fórmula da variância acima é equivalente ao **número de graus de liberdade** envolvido.

O número de graus de liberdade é utilizado no cálculo da variância e de outras medidas de variabilidade, quando as mesmas são obtidas a partir de uma amostra de dados e a teoria prova que, quando a média verdadeira não é conhecida e faz-se o cálculo de s^2 a partir de uma estimativa $\hat{m}$, isto eqüivale exatamente à perda de uma das observações.

O número de graus de liberdade é conceituado como o número de valores num conjunto de dados que pode ser designado arbitrariamente. Por exemplo, suponha que um pesquisador vai distribuir, através de sorteio, dez vacas holandesas em um galpão contendo dez baias, para avaliar duas rações comerciais em relação à produção de leite. No primeiro sorteio, a chance de qualquer uma das dez vacas ocupar a baia nº 1 é a mesma, pois têm-se dez opções de escolha. Depois de sorteada a baia nº 1, passa-se ao segundo sorteio, onde a chance de qualquer uma das nove vacas ocupar a baia nº 2 é a mesma, pois têm-se nove opções de escolha. Depois de sorteada a baia nº 2, passa-se ao terceiro sorteio, onde a chance de qualquer uma das oito vacas ocupar a baia nº 3 é a mesma, pois têm-se oito opções de escolha, e assim sucessivamente. Quando só restarem duas baias, passa-se ao nono sorteio, onde a chance de qualquer uma das duas vacas ocupar a baia nº 9 é a mesma, pois têm-se duas opções de escolha. Porém, depois de sorteada a baia nº 9, a última vaca já não tem mais opção de escolha, ou seja, ela ficará na baia nº 10. Portanto, o número de opções é igual a 9, isto é, N – 1.

Considerando os dados das amostras do exemplo anterior, tem-se:

$$s^2_{\ (1)} = \frac{SQD}{N-1}$$

$$= \frac{(0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2}{5-1}$$

$$= \frac{(0) + (0) + (0) + (0) + (0)}{4}$$

$$= \frac{0}{4} = \mathbf{0{,}0}$$

$$s^2_{\ (2)} = \frac{SQD}{N-1}$$

$$= \frac{(-2)^2 + (0)^2 + (2)^2 + (-1)^2 + (1)^2}{5-1}$$

$$= \frac{(4) + (0) + (4) + (1) + (1)}{4}$$

$$= \frac{10}{4} = \mathbf{2{,}5}$$

$$s^2_{\ (3)} = \frac{SQD}{N-1}$$

$$= \frac{(0)^2 + (-7)^2 + (-1)^2 + (7)^2 + (1)^2}{5-1}$$

$$= \frac{(0) + (49) + (1) + (49) + (1)}{4}$$

$$= \frac{100}{4} = \mathbf{25{,}0}$$

$$s^2_{\ (4)} = \frac{\text{SQD}}{\text{N}-1}$$

$$= \frac{(7)^2 + (5)^2 + (-3)^2 + (-7)^2 + (-2)^2}{5-1}$$

$$= \frac{(49) + (25) + (9) + (49) + (4)}{4}$$

$$= \frac{136}{4} = \mathbf{34{,}0}$$

Um modo mais prático de calcular a SQD é o que se segue:

$$\text{SQD} = \sum X^2 - \frac{\left(\sum X\right)^2}{N}$$

Assim, a fórmula da variância fica:

$$s^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

Considerando o mesmo exemplo, tem-se:

$$s^2_{\ (1)} = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(10)^2 + (10)^2 + (10)^2 + (10)^2 + (10)^2 - \frac{(50)^2}{5}}{5-1}$$

$$= \frac{(100)+(100)+(100)+(100)+(100)-\frac{.(2.500)}{5}}{4}$$

$$= \frac{500-500}{4}$$

$$= \frac{0}{4} = \mathbf{0,0}$$

$$s^2{}_{(2)} = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(8)^2+(10)^2+(12)^2+(9)^2+(11)^2-\frac{(50)^2}{5}}{5-1}$$

$$= \frac{(64)+(100)+(144)+(81)+(121)-\frac{(2.500)}{5}}{4}$$

$$= \frac{510-500}{4}$$

$$= \frac{10}{4} = \mathbf{2,5}$$

$$s^2{}_{(3)} = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(10)^2+(3)^2+(9)^2+(17)^2+(11)^2-\frac{(50)^2}{5}}{5-1}$$

$$= \frac{(100)+(9)+(81)+(289)+(121)-\frac{(2.500)}{5}}{4}$$

$$= \frac{600-500}{4}$$

$$= \frac{100}{4} = \mathbf{25,0}$$

$$s^2_{(4)} = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(17)^2 + (15)^2 + (7)^2 + (3)^2 + (8)^2 - \frac{(50)^2}{5}}{5-1}$$

$$= \frac{(289) + (225) + (49) + (9) + (64) - \frac{(2.500)}{5}}{4}$$

$$= \frac{636 - 500}{4}$$

$$= \frac{136}{4} = \mathbf{34{,}0}$$

A vantagem deste método é que se trabalha diretamente com os dados originais, não havendo, pois, necessidade de calcular-se previamente a média e os desvios em relação a ela.

É interessante observar que as amostras (3) e (4) já referidas, embora não pudessem ser diferenciadas pela amplitude total, podem perfeitamente ser identificadas através da variância. Neste caso, observa-se que a amostra (4) é mais dispersa que a amostra (3).

Considerando os dados do Exemplo 1, a variância será:

$$s^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(273{,}0)^2 + (660{,}0)^2 + \cdots + (453{,}0)^2 - \frac{(2.731{,}0)^2}{6}}{6-1}$$

$$= \frac{(74.529{,}0) + (435.600{,}0) + \cdots + (205.209{,}0) - \frac{(7.458.361{,}0)}{6}}{5}$$

$$= \frac{(74.529{,}0) + (435.600{,}0) + \cdots + (205.209{,}0) - 1.243.060{,}167}{5}$$

$$= \frac{1.396.213{,}0 - 1.243.060{,}167}{5}$$

$$= \frac{153.152,833}{5} = \mathbf{30.630,5666\ (kg/ha)^2}$$

O valor s^2 = 30.630,5666 $(kg/ha)^2$ é uma estimativa de variabilidade dos dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, em torno da média aritmética ($\hat{m} \cong$ 455,17 kg/ha). Mesmo sendo uma unidade quadrática, verifica-se que houve uma variação relativamente grande nos dados de produtividade de algodão herbáceo em torno da média aritmética.

Considerando, também, os dados do Exemplo 2, a variância será:

$$s^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(47,0)^2 + (41,0)^2 + \cdots + (40,0)^2 - \frac{(495,0)^2}{12}}{12-1}$$

$$= \frac{(2.209,0) + (1.681,0) + \cdots + (1.600,0) - \frac{(245.025,0)}{12}}{11}$$

$$= \frac{(2.209,0) + (1.681,0) + \cdots + (1.600,0) - 20.418,75}{11}$$

$$= \frac{20.919,0 - 20.418,75}{11}$$

$$= \frac{500,25}{11} \cong \mathbf{45,4773\ kg^2}$$

O valor $s^2 \cong$ 45,4773 kg^2 é uma estimativa de variabilidade dos dados de peso ao nascer de bezerros machos da raça Charolesa, em torno da média aritmética ($\hat{m}$ = 41,25 kg). Mesmo sendo uma unidade quadrática, verifica-se que houve uma variação relativamente pequena nos dados de peso ao nascer de bezerros machos em torno da média aritmética.

Numa série de dados grupados em classes, a variância (s^2) é obtida através da seguinte fórmula:

$$s^2 = \frac{\sum\left(d^2\ x\ f\right)}{N-1}$$

onde:
d = desvio de cada ponto médio em relação à média aritmética da série ($Pm - \hat{m}$);
f = freqüência de cada classe;
N = número de observações.

Considerando os dados do Exemplo 3, a variância será:

$$s^2 = \frac{\sum \left(d^2 \ x \ f\right)}{N-1}$$

$$= \frac{\left[(50{,}0 - 78{,}44)^2 \, x \, 3 + (58{,}0 - 78{,}44)^2 \, x \, 7 + \cdots + (106{,}0 - 78{,}44)^2 \, x \, 4\right]}{90-1}$$

$$= \frac{\left[(-28{,}44)^2 \, x \, 3 + (-20{,}44)^2 \, x \, 7 + \cdots + (27{,}56)^2 \, x \, 4\right]}{89}$$

$$= \frac{\left[(808{,}8336) \, x \, 3 + (417{,}7936) \, x \, 7 + \cdots + (759{,}5536) \, x \, 4\right]}{89}$$

$$= \frac{\left[(2.426{,}5008) + (2.924{,}5552) + \cdots + (3.038{,}2144)\right]}{89}$$

$$= \frac{15.374{,}2240}{89} = \mathbf{172{,}7440899 \ cm^2}$$

O valor s^2 = 172,7440899 cm^2 é uma estimativa de variabilidade dos dados de altura de planta de sorgo granífero, em torno da média aritmética ($\hat{m} \cong$ 78,44 cm). Mesmo sendo uma unidade quadrática, verifica-se que houve uma variação pequena nos dados de altura de planta, em torno da média aritmética.

Considerando, também, os dados do Exemplo 4, a variância será:

$$s^2 = \frac{\sum \left(d^2 \ x \ f\right)}{N-1}$$

$$= \frac{\left[(425{,}0 - 474{,}82)^2 \ x \ 3 + (439{,}0 - 474{,}82)^2 \ x \ 8 + \cdots + (523{,}0 - 474{,}82)^2 \ x \ 4\right]}{120-1}$$

$$= \frac{\left[(-49{,}82)^2 \, x \, 3 + (-35{,}82)^2 \, x \, 8 + \cdots + (48{,}18)^2 \, x \, 4\right]}{119}$$

$$= \frac{\left[(2.482{,}0324) \, x \, 3 + (1.283{,}0724) \, x \, 8 + \cdots + (2.321{,}3124) \, x \, 4\right]}{119}$$

$$= \frac{\left[(7.446{,}0972) + (10.264{,}5792) + \cdots + (9.285{,}2496)\right]}{119}$$

$$= \frac{55.583{,}9680}{119} = \mathbf{467{,}0921681 \ g^2}$$

O valor s^2 = 467,0921681 g^2 é uma estimativa de variabilidade dos dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade, em torno da média aritmética ($\hat{m} \cong$ 474,82 g). Mesmo sendo uma unidade quadrática, verifica-se que houve uma variação muito pequena nos dados de peso corporal, em torno da média aritmética.

3.3.3 Desvio padrão

A variância, pela sua natureza, tem uma unidade quadrática. A sua raiz quadrada, que ainda é uma medida de variabilidade, é denominada desvio padrão.

O desvio padrão é uma medida de dispersão muito usada pelo fato de que permite a interpretação direta da variação dos dados, pois o mesmo apresenta a mesma unidade dos dados originais e, consequentemente, da média. O seu cálculo é muito importante, porque através dele o pesquisador estima a variação acidental que ocorre nos dados experimentais.

Numa série de dados não grupados, como por exemplo, numa amostra de dados X_1 , X_2 , ... , X_N , o desvio padrão (s) é obtido através das seguintes fórmulas:

$$s = \sqrt{\frac{SQD}{N-1}} = \sqrt{s^2}$$

ou

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}} = \sqrt{s^2}$$

Considerando os dados das amostras do exemplo anterior, tem-se:

$$s_{(1)} = \sqrt{s^2}$$

$$= \sqrt{0,0} = \mathbf{0,0000}$$

$$s_{(2)} = \sqrt{s^2}$$

$$= \sqrt{2,5} \cong \mathbf{1,5811}$$

$$s_{(3)} = \sqrt{s^2}$$

$$= \sqrt{25,0} = \mathbf{5,0000}$$

$$s_{(4)} = \sqrt{s^2}$$

$$= \sqrt{34,0} \cong \mathbf{5,8310}$$

Também, aqui, as amostras (3) e (4) podem perfeitamente ser identificadas através do desvio padrão, continuando a amostra (4) como sendo a mais dispersa.

Considerando os dados do Exemplo 1, o desvio padrão será:

$$s = \sqrt{s^2}$$

$$= \sqrt{30.630,5666} \cong \mathbf{175,0159\ kg/ha}$$

O valor s $\cong$ 175,0159 kg/ha é uma estimativa de variabilidade dos dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, em torno da média aritmética ($\hat{m} \cong$ 455,17 kg/ha). Houve uma variação relativamente grande nos dados de produtividade de algodão herbáceo, em relação à média aritmética.

Considerando, também, os dados do Exemplo 2, o desvio padrão será:

$$s = \sqrt{s^2}$$

$$= \sqrt{45,4773} \cong \mathbf{6,7437\ kg}$$

O valor s $\cong$ 6,7437 kg é uma estimativa de variabilidade dos dados de peso ao nascer de bezerros machos da raça Charolesa, em torno da média aritmética ($\hat{m}$ = 41,25 kg). Houve uma variação relativamente pequena nos dados de peso ao nascer de bezerros machos, em relação à média aritmética.

Numa série de dados grupados em classes, o desvio padrão (s) é obtido através da seguinte fórmula:

$$s = \sqrt{\frac{d^2 x\ f}{N-1}} = \sqrt{s^2}$$

Considerando os dados do Exemplo 3, o desvio padrão será:

$$s = \sqrt{s^2}$$
$$= \sqrt{172,7440899} \cong \mathbf{13,1432\ cm}$$

O valor s $\cong$ 13,1432 cm é uma estimativa de variabilidade dos dados de altura de planta de sorgo granífero, em torno da média aritmética ($\hat{m} \cong$ 78,44 cm). Houve uma variação pequena nos dados de altura de planta, em relação à média aritmética.

Considerando, também, os dados do Exemplo 4, o desvio padrão será:

$$s = \sqrt{s^2}$$

$$= \sqrt{467,0921681} \cong \mathbf{21,6123\ g}$$

O valor s $\cong$ 21,6123 g é uma estimativa de variabilidade dos dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade, em torno da média aritmética ($\hat{m} \cong$ 474,82 g). Houve uma variação muito pequena nos dados de peso corporal, em relação à média aritmética.

3.3.4 Erro padrão da média

Se ao invés de uma amostra tivessem várias, provenientes de uma mesma população, seriam obtidas também diversas estimativas da média, porém, distintas entre si.

A partir dessas diversas estimativas de média, pode-se estimar uma variância da média, considerando os desvios de cada média em relação à média de todas elas.

Entretanto, demonstra-se que a partir de uma única amostra pode-se estimar essa variância [$s^2(\hat{m})$], através da fórmula:

$$s^2(\hat{m}) = \frac{s^2}{N}$$

onde:
s^2 = variância de uma amostra de dados;
N = número de observações.

A sua raiz quadrada é denominada erro padrão da média, $s(\hat{m})$, ou seja:

$$s(\hat{m}) = \frac{s}{\sqrt{N}}$$

onde:
s = desvio padrão de uma amostra de dados;
N = número de observações.

O erro padrão da média dá uma perfeita idéia da precisão da média, isto é, quanto menor ele for, maior precisão terá a média.

Considerando os dados das amostras do exemplo anterior, tem-se:

$$s(\hat{m})_{(1)} = \frac{s}{\sqrt{N}}$$

$$= \frac{0,0}{\sqrt{5}}$$

$$= \frac{0,0}{2,236068} = \mathbf{0,0000}$$

$$s(\hat{m})_{(2)} = \frac{s}{\sqrt{N}}$$

$$= \frac{1,581139}{\sqrt{5}}$$

$$= \frac{1,581139}{2,236068} \cong \mathbf{0,7071}$$

$$s\,(\hat{m})_{(3)} = \frac{s}{\sqrt{N}}$$

$$= \frac{5,0}{\sqrt{5}}$$

$$= \frac{5,0}{2,236068} \cong \mathbf{2,2361}$$

$$s\,(\hat{m})_{(4)} = \frac{s}{\sqrt{N}}$$

$$= \frac{5,830952}{\sqrt{5}}$$

$$= \frac{5,830952}{2,236068} \cong \mathbf{2,6077}$$

Sempre que se cita uma média deve-se fazê-la acompanhar-se de seu erro padrão. Assim, no caso das amostras de (1) a (4) exemplificadas, quando acompanhadas de seus erros padrões ficam:

(1) 10,0 $\pm$ 0,0000

(2) 10,0 $\pm$ 0,7071

(3) 10,0 $\pm$ 2,2361

(4) 10,0 $\pm$ 2,6077

o que mostra a menor precisão da média, na amostra (4).

Considerando os dados do Exemplo 1, o erro padrão da média será:

$$s\,(\hat{m}) = \frac{s}{\sqrt{N}}$$

$$= \frac{175,0159}{\sqrt{6}}$$

$$= \frac{175,0159}{2,449490} \cong \mathbf{71,4499\ kg/ha}$$

O valor s ($\hat{m}$) $\cong$ 71,4499 kg/ha é uma estimativa de variabilidade dos dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977, em torno da média aritmética ($\hat{m} \cong$ 455,17 kg/ha). Houve uma variação muito grande entre a média aritmética dos dados de produtividade de algodão herbáceo e seu erro padrão, indicando uma precisão muito baixa da mesma.

Considerando, também, os dados do Exemplo 2, o erro padrão da média será:

$$s(\hat{m}) = \frac{s}{\sqrt{N}}$$

$$= \frac{6{,}7437}{\sqrt{12}}$$

$$= \frac{6{,}7437}{3{,}464102} \cong \mathbf{1{,}9467\ kg}$$

O valor s ($\hat{m}$) $\cong$ 1,9467 kg é uma estimativa de variabilidade dos dados de peso ao nascer de bezerros machos da raça Charolesa, em torno da média aritmética ($\hat{m}$ = 41,25 kg). Houve uma variação relativamente pequena entre a média aritmética dos dados de peso ao nascer de bezerros machos e seu erro padrão, indicando uma precisão relativamente alta da mesma.

Considerando, ainda, os dados do Exemplo 3, o erro padrão da média será:

$$s(\hat{m}) = \frac{s}{\sqrt{N}}$$

$$= \frac{13{,}1432}{\sqrt{90}}$$

$$= \frac{13{,}1432}{9{,}486833} \cong \mathbf{1{,}3854\ cm}$$

O valor s ($\hat{m}$) $\cong$ 1,3854 cm é uma estimativa de variabilidade dos dados de altura de planta de sorgo granífero, em torno da média aritmética ($\hat{m} \cong$ 78,44 cm). Houve uma variação pequena entre a média aritmética dos dados de altura de planta e seu erro padrão, indicando uma alta precisão da mesma.

Considerando, por fim, os dados do Exemplo 4, o erro padrão da média será:

$$s(\hat{m}) = \frac{s}{\sqrt{N}}$$

$$= \frac{21{,}6123}{\sqrt{120}}$$

$$= \frac{21{,}6123}{10{,}954451} \cong \mathbf{1{,}9729\ g}$$

O valor s ($\hat{m}$) ≅ 1,9729 g é uma estimativa de variabilidade dos dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade, em torno da média aritmética ($\hat{m}$ ≅ 474,82 g). Houve uma variação muito pequena entre a média aritmética dos dados de peso corporal e seu erro padrão, indicando uma precisão muito alta da mesma.

Como foi visto anteriormente, a média sempre deve vir acompanha de seu erro padrão. Assim, no caso dos Exemplos de 1 a 4, têm-se:

Exemplo 1: 455,17 kg/ha ± 71,45 kg/ha

Exemplo 2: 41,25 kg ± 1,95 kg

Exemplo 3: 78,44 cm ± 1,39 cm

Exemplo 4: 474,82 g ± 1,97 g

3.3.5 Coeficiente de variação

O coeficiente de variação (CV) é uma medida de variabilidade que mede percentualmente a relação entre o desvio padrão (s) e a média aritmética ($\hat{m}$), ou seja:

$$CV = \frac{100\ x\ s}{\hat{m}}$$

Como **s** e $\hat{m}$ são expressos na mesma unidade, o coeficiente de variação é um número abstrato, isto é, não tem unidade.

Esta medida de variabilidade pode ser empregada tanto em dados grupados como não grupados. Se o desvio padrão for calculado sobre a mediana ou sobre a moda (que é possível, mas não se usa), outros coeficientes poderão ser obtidos.

Considerando os dados das amostras do exemplo anterior, tem-se:

$$CV_{(1)} = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 0{,}0}{10}$$

$$= \frac{0{,}0}{10} = \mathbf{0{,}0\%}$$

$$CV_{(2)} = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 1{,}581139}{10}$$

$$= \frac{158{,}1139}{10} \cong \mathbf{15{,}81\%}$$

$$CV_{(3)} = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 5{,}0}{10}$$

$$= \frac{500{,}0}{10} = \mathbf{50{,}00\%}$$

$$CV_{(4)} = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 5{,}830952}{10}$$

$$= \frac{583{,}0952}{10} \cong \mathbf{58{,}31\%}$$

Aqui, também, as amostras (3) e (4) podem, perfeitamente, ser identificadas através do coeficiente de variação, mostrando novamente que a amostra (4) é a mais dispersa.

Considerando os dados do Exemplo 1, o coeficiente de variação será:

$$CV = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 175{,}0159}{455{,}17}$$

$$= \frac{17.501{,}59}{455{,}17} \cong \mathbf{38{,}45\%}$$

O valor CV $\cong$ 38,45% é uma estimativa de variabilidade dos dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977.

Considerando, também, os dados do Exemplo 2, o coeficiente de variação será:

$$CV = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 6{,}7437}{41{,}25}$$

$$= \frac{674,37}{41,25} \cong \mathbf{16,35\%}$$

O valor CV ≅ 16,35% é uma estimativa de variabilidade dos dados de peso ao nascer de bezerros machos da raça Charolesa.

Considerando, ainda, os dados do Exemplo 3, o coeficiente de variação será:

$$CV = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 13,1432}{78,44}$$

$$= \frac{1.314,32}{78,44} \cong \mathbf{16,76\%}$$

O valor CV ≅ 16,76% é uma estimativa de variabilidade dos dados de altura de planta de sorgo granífero.

Considerando, por fim, os dados do Exemplo 4, o coeficiente de variação será:

$$CV = \frac{100\ x\ s}{\hat{m}}$$

$$= \frac{100\ x\ 21,6123}{474,82}$$

$$= \frac{2.161,23}{474,82} \cong \mathbf{4,55\%}$$

O valor CV ≅ 4,55% é uma estimativa de variabilidade dos dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade.

O coeficiente de variação serve também para análise comparativa envolvendo unidades e séries diferentes. Por exemplo, considerando os dados dos Exemplos 1, 2, 3 e 4 têm-se:

Exemplo 1: Distribuição de produtividade de algodão herbáceo:

$$\hat{m} \cong 455,17 \text{ kg/ha}$$

$$s \cong 175,0159 \text{ kg/ha}$$

$$CV \cong 38,45\%$$

Exemplo 2: Distribuição de peso ao nascer de bezerros machos da raça Charolesa:

$$\hat{m} = 41,25 \text{ kg}$$

$$s \cong 6{,}7437 \text{ kg}$$

$$CV \cong 16{,}35\%$$

Exemplo 3: Distribuição de altura de planta de sorgo granífero:

$$\hat{m} \cong 78{,}44 \text{ cm}$$

$$s \cong 13{,}1432 \text{ cm}$$

$$CV \cong 16{,}76\%$$

Exemplo 4: Distribuição de peso corporal de um lote misto de frangos de corte:

$$\hat{m} \cong 474{,}82 \text{ g}$$

$$s \cong 21{,}6123 \text{ g}$$

$$CV \cong 4{,}55\%$$

Verifica-se, assim, que entre as distribuições comparadas, a distribuição de peso corporal de um lote misto de frangos de corte com 15 dias de idade é mais homogênea (menos dispersa), enquanto que a distribuição de produtividade de algodão herbáceo é a mais dispersa.

Uma desvantagem do coeficiente de variação é que ele deixa de ser útil quando a média esta próxima de zero.

O coeficiente de variação dá uma idéia de precisão do experimento, ou seja, quanto menor o coeficiente de variação maior será a precisão do experimento.

De um modo geral, quando o coeficiente de variação for inferior a 10%, diz-se que o experimento apresentou uma ótima precisão experimental; quando variar de 10 a 15%, diz-se que o experimento apresentou uma boa precisão experimental; quando estiver no intervalo de $> 15\% \leq 20\%$, diz-se que o experimento apresentou uma precisão experimental regular ou aceitável; quando estiver no intervalo de $> 20\% \leq 30\%$, diz-se que o experimento apresentou uma péssima precisão experimental; e quando for superior a 30%, diz-se que o experimento apresentou uma precisão experimental muito péssima. Por conta disso, espera-se que os coeficientes de variação dos ensaios agropecuários, principalmente aqueles conduzidos ao nível de campo, não ultrapassem a casa dos 20%, de modo que as conclusões obtidas de tais ensaios tenham credibilidade perante a comunidade científica.

Contudo, é preciso ressaltar que nem sempre um coeficiente de variação superior à casa dos 20% significa que as conclusões obtidas não tenham credibilidade perante a comunidade científica. Isso depende muito do tipo de experimento. Por exemplo, nos experimentos com consorciação de culturas, o coeficiente de variação é geralmente alto em comparação com os experimentos com culturas isoladas. Neste caso, os coeficientes de variação de 20 a 30% são racionais e aceitáveis perante a comunidade científica. Também em experimentos de campo na área de Entomologia, coeficientes de variação superiores a 20% são normais e aceitáveis, pois em função do comportamento dos insetos é muito raro obter coeficientes de variação baixos.

Por outro lado, nem sempre se consegue uma ótima precisão experimental, com CV < 5%, nos ensaios de laboratório, casa-de-vegetação ou galpão, visto que geralmente são mais precisos do que os ensaios de campo. Mais uma vez, isso depende muito do tipo de experimento. Por exemplo, dados de análise de solo não raro apresentam coeficientes de variação superiores a 20% e em alguns casos superiores a 30%, especialmente no caso de solos pobres, como os de cerrado.

Portanto, cabe ao pesquisador avaliar e justificar a precisão de seus dados experimentais baseando-se nesses fatos.

3.3.6 Intervalo de confiança para a média

Foi visto até agora que as médias obtidas das amostras dos Exemplos 1, 2, 3 e 4 representam suas médias populacionais, onde o único valor obtido de cada amostra estima esse parâmetro de interesse. Tal método de estimação é chamado de **estimação por ponto**, o qual é comumente usado. Contudo, como a média de uma amostra é um estatístico e os mesmos variam de amostra para amostra, o problema é que se tivessem duas ou mais amostras para cada um dos exemplos citados acima é muito provável que os resultados de suas médias não seriam iguais, havendo um grau de incerteza envolvido. Uma estimativa por ponto não fornece nenhuma informação sobre a variabilidade inerente do estimador, ou seja, não se sabe se a média estimada está próxima ou distante da média verdadeira.

Por outro lado, existe um outro método de estimação muito usado, conhecido como **estimação por intervalo**, que é freqüentemente preferido em relação ao método anterior, pois fornece um intervalo de valores razoável no qual se presume que esteja o parâmetro de interesse (a média verdadeira) com certo grau de confiança. Esse intervalo de valores é chamado **intervalo de confiança**.

O intervalo de confiança (IC) para a média é obtido através da seguinte fórmula:

$$IC = \hat{m} \pm t_{(5\%)} \times s(\hat{m})$$

onde:

$\hat{m}$ = estimativa da média;

$t_{(5\%)}$ = valor tabelado do teste t no nível de 5% de probabilidade (TABELA A.7);

$s(\hat{m})$ = erro padrão da média.

Considerando os dados do Exemplo 1, o intervalo de confiança da média será:

$$IC = \hat{m} \pm t_{(5\%)} \times s(\hat{m})$$
$$= 455{,}17 \pm 2{,}57 \times 71{,}4499$$

$$= 455{,}17 \pm 183{,}63 \;\therefore$$

IC = (**271,54 kg/ha**; **638,80 kg/ha**)

Os valores de IC = (271,54 kg/ha; 638,80 kg/ha) indicam o intervalo de confiança, com 95% de probabilidade, onde se encontra a média verdadeira para os dados de produtividade de algodão herbáceo, variedade ALLEN – 333/57, no Município de Viçosa-AL, no ano de 1977. Houve uma variação muito grande no intervalo de confiança dos dados de produtividade de algodão herbáceo, indicando uma precisão muito baixa da estimativa da média ($\hat{m} \cong$ 455,17 kg/ha).

Considerando, também, os dados do Exemplo 2, o intervalo de confiança da média será:

$$IC = \hat{m} \underline{+} t_{(5\%)} \times s(\hat{m})$$

$$= 41{,}25 \underline{+} 2{,}20 \times 1{,}9467$$

$$= 41{,}25 \underline{+} 4{,}28 \therefore$$

IC = (**36,97 kg**; **45,53 kg**)

Os valores de IC = (36,97 kg; 45,53 kg) indicam o intervalo de confiança, com 95% de probabilidade, onde se encontra a média verdadeira para os dos dados de peso ao nascer de bezerros machos da raça Charolesa. Houve uma variação relativamente pequena no intervalo de confiança dos dados de peso ao nascer de bezerros machos, indicando uma precisão relativamente alta da estimativa da média ($\hat{m}$ = 41,25 kg).

Considerando, ainda, os dados do Exemplo 3, o intervalo de confiança da média será:

$$IC = \hat{m} \underline{+} t_{(5\%)} \times s(\hat{m})$$

$$= 78{,}44 \underline{+} 1{,}99 \times 1{,}3854$$

$$= 78{,}44 \underline{+} 2{,}76 \therefore$$

IC = (**75,68 cm**; **81,20 cm**)

Os valores de IC = (75,68 cm; 81,20 cm) indicam o intervalo de confiança, com 95% de probabilidade, onde se encontra a média verdadeira para os dados de altura de planta de sorgo granífero. Houve uma variação pequena no intervalo de confiança dos dados de altura de planta de sorgo granífero, indicando uma alta precisão da estimativa da média ($\hat{m} \cong$ 78,44 cm).

Considerando, por fim, os dados do Exemplo 4, o intervalo de confiança da média será:

$$IC = \hat{m} \underline{+} t_{(5\%)} \times s(\hat{m})$$
$$= 474{,}82 \underline{+} 1{,}98 \times 1{,}9729$$

$$= 474{,}82 \underline{+} 3{,}91 \therefore$$

IC = (**470,91 g**; **478,82 g**)

Os valores de IC = (470,91 g; 478,82 g) indicam o intervalo de confiança, com 95% de probabilidade, onde se encontra a média verdadeira para os dados de peso corporal de um lote misto de frangos de corte com 15 dias de idade. Houve uma variação muito pequena no intervalo de confiança dos dados de peso corporal, indicando uma precisão muito alta da estimativa da média ($\hat{m} \cong 474{,}82$ g).

3.4 Exercícios

a) Num ensaio sobre competição de variedades de algodão herbáceo, foram obtidos os seguintes resultados de peso de 20 capulhos (gramas):

V1		V2		V3		V4		V5		V6		V7		V8	
78	75	100	85	102	85	72	88	98	85	88	102	98	100	102	100
90	70	65	92	95	80	85	80	70	88	83	88	90	85	92	102
90	88	78	90	102	98	98	85	85	80	138	85	95	95	88	85

Pede-se:

a.1) Determine, para cada variedade, o peso médio de 20 capulhos, o erro padrão da média, o coeficiente de variação e o intervalo de confiança da média.

a.2) Sem levar em conta a variedade, determine o peso médio de 20 capulhos, o erro padrão da média, a mediana, a moda, o coeficiente de variação e o intervalo de confiança da média.

b) Admitindo-se que seja de 18% o coeficiente de variação relativo ao peso de ovos de galinha, perguntam-se quantos ovos devem ser pesados para obter-se um erro padrão da média igual a 3% dela.

c) Numa amostra de 30 dados de pesos ao nascer de bezerros machos da raça nelore obteve-se a média $\hat{m} = 52$ kg, com um erro padrão da média s($\hat{m}$) = 3,2 kg. Pede-se o coeficiente de variação referente a estes dados.

d) A fim de se obter a produção média de algodão em uma fazenda, foi tomada ao acaso as produções de 20 pequenas parcelas de 100 m^2, cujo resultado, em gramas, foi o seguinte:

2.730	6.750	3.150	7.230
3.800	4.350	2.980	3.300
2.370	3.100	4.370	2.330
3.770	3.850	3.330	6.420
2.930	3.500	8.200	3.400

Pede-se:

d.1) A produção média, em kg/ha, com seu respectivo erro padrão.

d.2) O coeficiente de variação.

d.3) Admitindo-se que a área da fazenda destinada ao plantio de algodão seja de 180 ha, qual a produção esperada e seu erro padrão?

e) Na determinação da altura de planta de soja, em cm, foram analisadas 15 amostras, obtendo-se o resultado que se segue:

62,0	76,3	69,7	57,7	50,0
49,7	51,0	77,0	51,7	56,7
64,7	79,0	66,0	55,0	96,3

Pede-se:

e.1) Calcular a altura média de planta de soja, em cm, e o seu erro padrão.

e.2) Obter a mediana e a moda.

e.3) Determinar o coeficiente de variação.

e.4) Determinar o intervalo de confiança da média.

f) Considerando a série de dados a seguir, referente ao consumo acumulado de ração (g) de frangos de corte com 25 dias de idade:

1.530	1.750	1.350	1.430
1.400	1.350	1.680	1.360
1.370	1.400	1.370	1.330
1.570	1.780	1.330	1.420
1.330	1.500	1.500	1.300
1.730	1.750	1.550	1.530
1.800	1.350	1.580	1.600
1.370	1.400	1.370	1.630
1.770	1.800	1.330	1.420
1.630	1.500	1.500	1.500
1.530	1.750	1.550	1.630

Pede-se:

f.1) Construir uma tabela de freqüência, um histograma de freqüência e um polígono de freqüência.

f.2) Calcular o consumo médio acumulado de ração (g) de frangos de corte com 25 dias de idade e o seu erro padrão.

f.3) Obter a mediana e a moda.

f.4) Determinar a amplitude total, o coeficiente de variação e o intervalo de confiança da média.

g) Considerando a série de dados a seguir, referente ao número de sementes na espiga de progênies de meios irmãos de milho (PMI):

N$^{\underline{O}}$ de PMI	N$^{\underline{O}}$ de Sementes	N$^{\underline{O}}$ de PMI	N$^{\underline{O}}$ de Sementes	N$^{\underline{O}}$ de PMI	N$^{\underline{O}}$ de Sementes
1	313	18	412	35	392
2	596	19	358	36	370
3	350	20	627	37	599

4	440	21	392	38	409
5	426	22	354	39	486
6	476	23	522	40	519
7	326	24	348	41	416
8	385	25	474	42	344
9	490	26	410	43	430
10	418	27	412	44	551
11	457	28	411	45	573
12	394	29	482	46	602
13	344	30	495	47	407
14	483	31	405	48	355
15	399	32	370	49	431
16	523	33	405	50	372
17	413	34	433		

Pede-se:

g.1) Construir uma tabela de freqüência, um histograma de freqüência e um polígono de freqüência.

g.2) Calcular o número médio de sementes na espiga de milho e o seu erro padrão.

g.3) Obter a mediana e a moda.

g.4) Determinar a amplitude total e o coeficiente de variação.

h) Um estudo realizado com dois tipos de adubos orgânicos na cultura do capim elefante revelou os seguintes resultados de produção de matéria verde por ano (t/ha):

	AO1	AO2
Média	485,6	360,0
Mediana	242,4	359,1
Moda	210,0	359,8

Pede-se:

h.1) Interpretar e discutir os resultados obtidos desse estudo, considerando que uma cultura do capim elefante conduzida normalmente permite seis colheitas por ano, em torno de 60 t/ha por corte.