

Capítulo 5 – Representações gráficas para variáveis quantitativas

Introdução

Até o capítulo passado, você aprendeu a sintetizar dados a partir de um conjunto desordenado de dados, identificando a quantidade de classes, a amplitude das classes, o ponto médio, bem como as frequências absolutas, relativas, acumuladas e acumuladas relativas. Todos esses conceitos são importantes para que você compreenda os tópicos a serem abordados neste capítulo. Com eles, você conseguirá construir representações gráficas para variáveis quantitativas e interpretar informações apresentadas nos histogramas, polígonos de frequência, polígonos de frequência acumulada, gráficos de linha e gráficos de dispersão.

Além das tabelas de distribuição de frequência, os dados estatísticos podem ser representados de uma forma visual, isto é, por meio de gráficos. A principal vantagem da representação gráfica da distribuição de uma variável é possibilitar uma análise do comportamento dos dados de forma mais rápida e, assim, identificar algum padrão ou tendência. Além disso, muitos gestores preferem analisar relatórios por meio de gráficos a tabelas.

A estatística dispõe vários tipos de gráficos, que podem ser gráficos de linha ou de curva, e os diagramas de área, que incluem gráficos de colunas, barras e de setores (NAZARETH, 2003). Entre os gráficos que manipulam variáveis quantitativas, destacam-se os histogramas, os polígonos de frequência, os polígonos de frequência acumulada, gráficos de linha e os gráficos de dispersão. Então, agora, vamos aprender como construir cada um desses gráficos, bem como analisar as informações apresentadas.

Cabe ressaltar que os gráficos ilustrados neste capítulo foram elaborados no Microsoft Excel 2007, mas podem também ser construídos com o auxílio de software estatísticos.

Saiba mais

Statistica (<http://www.statsoft.com.br/>), *Stata* (<http://www.stata.com/>) e *Statistical Package for the Social Sciences* (<http://www.spss.com/statistics/>) são

ferramentas recomendadas para elaborar as representações gráficas discutidas neste capítulo. Além disso, essas ferramentas disponibilizam uma variedade de recursos para auxiliar a tomada de decisão em termos estatísticos. Visite os sítios antes de continuar seus estudos!

5.1 Histogramas

De acordo com Crespo (2002), um **histograma** é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidem com os pontos médios dos intervalos de classe. O número de retângulos é igual ao número de classes da distribuição.

Cada classe da distribuição é representada por um retângulo (ou coluna) e a sua altura corresponde à frequência da referida classe. Logo, quanto mais dados existirem em uma classe, mais alto será o retângulo. Já a largura dos retângulos representa a amplitude dos intervalos da classe. Assim a área de todo o histograma compreende a frequência total da distribuição.

Para construir um histograma, é necessário estabelecer as classes no eixo das abscissas de um sistema cartesiano, ou seja, no eixo x. Já as frequências ficam situadas no eixo das ordenadas, ou seja, no eixo y. Nos histogramas, podemos utilizar tanto a frequência absoluta quanto a frequência relativa, ou ambas, concomitantemente. A vantagem da inclusão da frequência relativa é ter uma informação mais percentual de uma classe em relação às demais.

Na figura 1, é possível visualizar o histograma construído a partir das informações das frequências absolutas apresentadas no quadro 5 do capítulo 4.

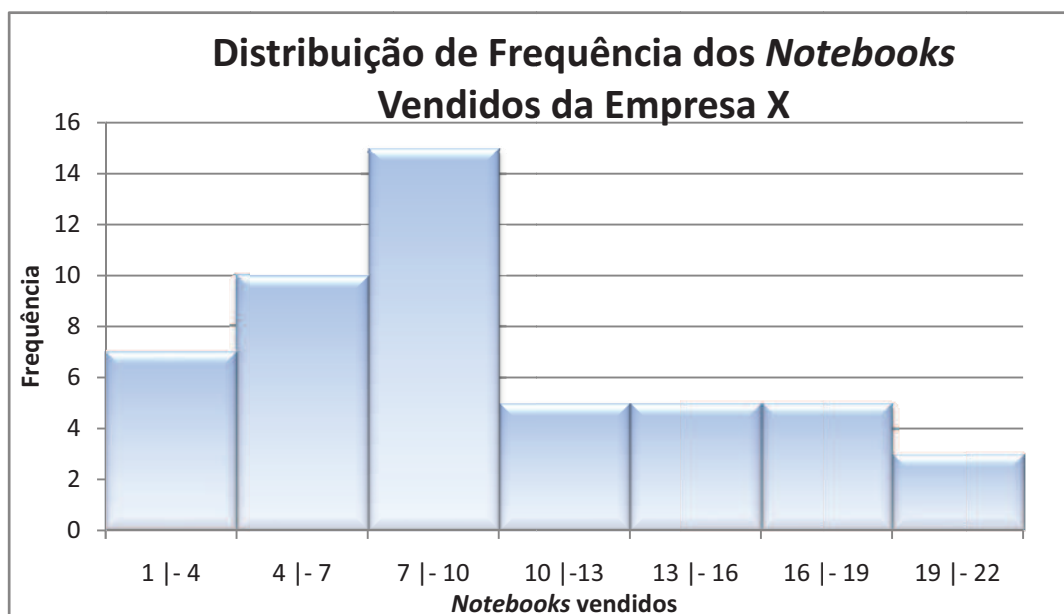


Figura 1: Histograma da frequência das vendas dos *notebooks*

As sete classes da distribuição encontram-se no eixo horizontal e correspondem ao intervalo de *notebooks* vendidos. Cada retângulo, por sua vez, representa a sua frequência absoluta. Assim, analisando esse histograma, podemos deduzir, por exemplo, que é muito mais comum a empresa X ter dias de venda entre 7 a 10 *notebooks* do que entre 19 a 22. Além disso, é possível identificar que há uma concentração maior de frequência com as classes situadas no início do gráfico. Nesse histograma, apresentamos ainda o título “Distribuição de Frequência dos *Notebooks* Vendidos da Empresa X” com o objetivo de registrar o tema das informações expostas. Isso não é uma regra, mas é recomendável, principalmente quando há diversos histogramas expostos ao mesmo tempo.

Os histogramas também podem ser construídos a partir de tabelas de distribuição de frequência cujas variáveis não possuem intervalos de classe, ou seja, variáveis quantitativas discretas. Como as variáveis discretas possuem um número reduzido de valores, no eixo horizontal, são dispostos todos os seus possíveis valores. Já os conceitos do eixo vertical e dos retângulos permanecem os mesmos.

Por exemplo, na figura 2, é ilustrado um histograma que apresenta a distribuição da frequência absoluta relativa de funcionários com relação ao número de computadores que possuem em casa.

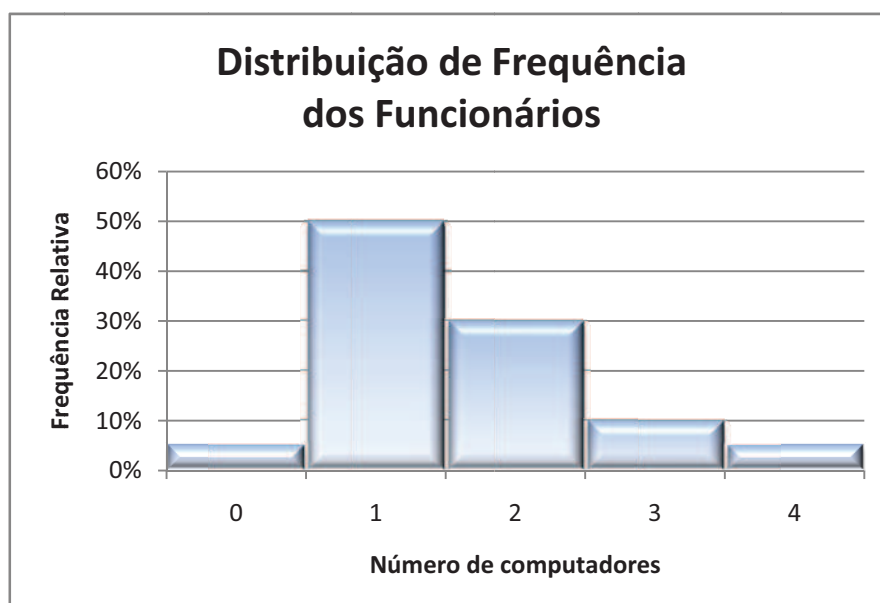


Figura 2: Histograma de frequência relativa dos funcionários

Nesse caso, podemos concluir que há uma concentração de funcionários que possuem entre 1 e 2 computadores em casa. Por outro lado, isso não ocorre para a quantidade de computadores igual a 0 ou 4. Nesse gráfico, inclusive, foram utilizadas as frequências relativas ao invés das absolutas.

Agora que você já conhece o histograma, vamos ao polígono de frequência.

5.2 Polígono de frequência

O polígono de frequência é um gráfico em linha, e as frequências são marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe (CRESPO, 2002).

A partir de um histograma, a construção de um polígono torna-se ainda mais fácil, pois é preciso somente encontrar o ponto médio do topo de cada retângulo e realizar uma pequena marcação ou um ponto. Em seguida, desenha-se uma reta reunindo os pontos identificados.

Na figura 3, é possível visualizar o polígono de frequência construído a partir das informações apresentadas no quadro 5 do capítulo 4.

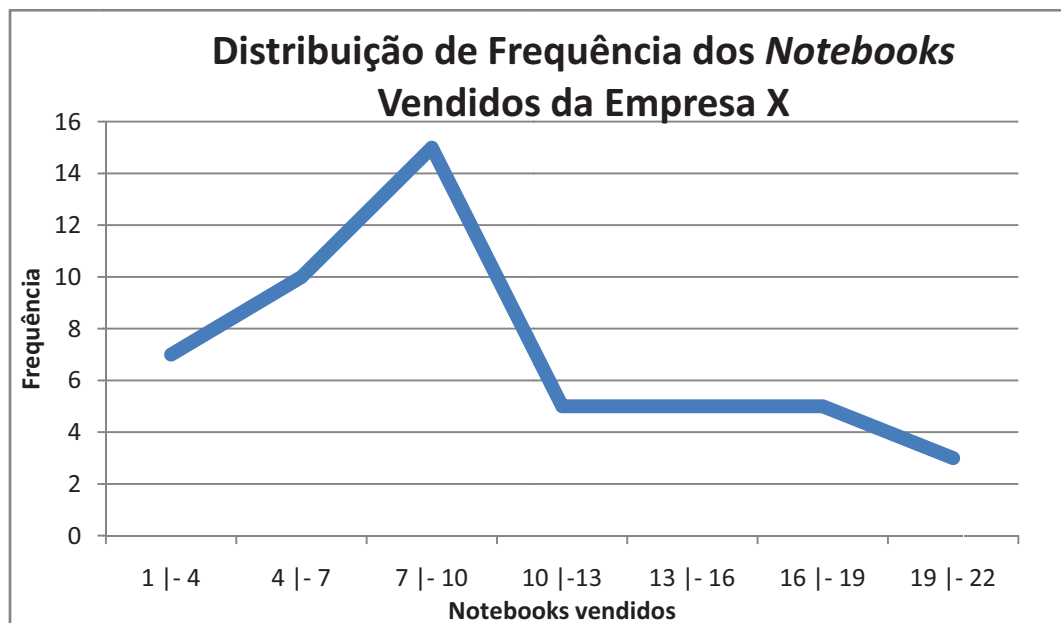


Figura 3: Polígono de frequência das vendas dos *notebooks*

Pode-se observar que esse gráfico contorna a altura dos retângulos apresentados na figura 1. É facilmente observável a estabilidade das classes 10 | -13, 13 | - 16 e 16 | - 19 para o valor de frequência igual a 5 devido ao traço horizontal.

Veja agora o polígono de frequência acumulada.

5.3 Polígono de frequência acumulada

O polígono de frequência acumulada, também conhecido como ogiva de Galton, é traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe (CRESPO, 2002).

Para construir esse tipo de gráfico, portanto, é preciso ter em mãos as frequências acumuladas das classes. Em cada classe, fazemos uma marcação (ponto) do valor da frequência e, depois, a cada dois pontos, unimos com uma reta.

Na figura 4, é possível visualizar o polígono de frequência acumulada construído a partir das informações apresentadas no quadro 5 do capítulo 4. Nesse gráfico, obtém-se uma visão mais acumulativa das frequências com o avanço das classes. Assim, por exemplo, podemos concluir que o número de *notebooks* vendidos entre 1 e 10 possui uma frequência acumulada maior do que 30.

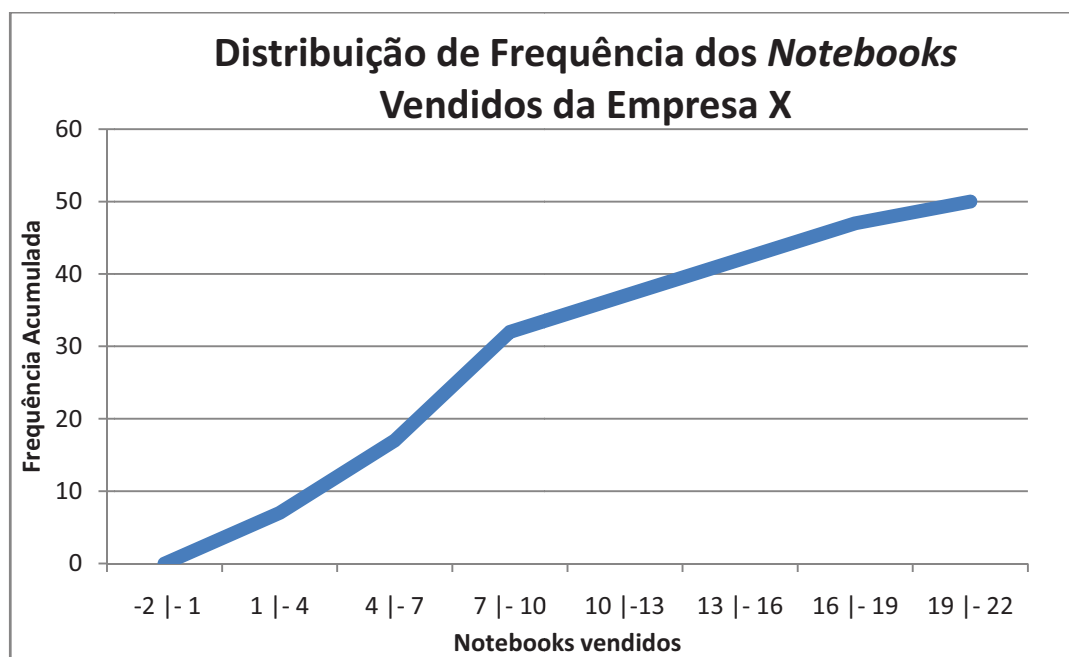


Figura 4: Polígono de frequência acumulada das vendas dos *notebooks*

Observe que nesse gráfico foi adicionada a classe -2 | -1 a fim de mostrar que o acúmulo inicia-se com 0 até a frequência acumulada total da distribuição, que é 50. Mas preste atenção! É importante manter a amplitude da classe que, nesse caso, é 4.

Como o histograma, o polígono de frequências e a ogiva produzem figuras regulares e irregulares, simétricas ou assimétricas, alongadas ou afiladas e nas posições mais díspares (como J ou U), frequentemente, uma análise é realizada com duas ou mais distribuições, colocando os respectivos gráficos sobrepostos ou lado-a-lado (MILONE, 2004). Nesse caso, teremos um gráfico combinado. Na figura 5, por exemplo, é possível visualizar um gráfico combinado, em que apresentamos um histograma e um polígono de frequência simultaneamente. Cabe salientar que dados aleatórios foram usados para exemplificar esse tipo de gráfico. Uma simples legenda também foi adicionada para informar os gráficos apresentados.

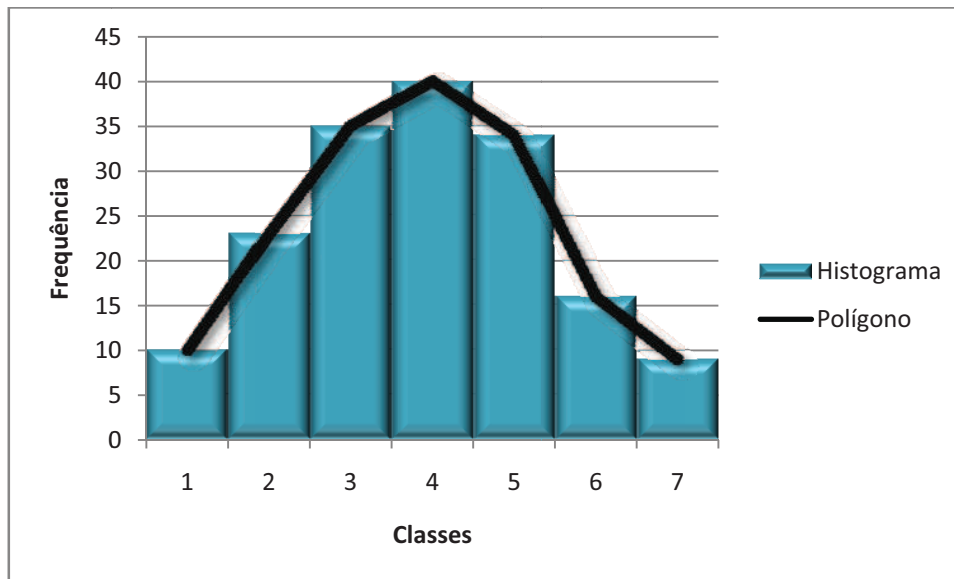


Figura 5: Polígono de frequência acumulada das vendas dos *notebooks*

Além disso, como, em geral, os dados coletados são de uma amostra, podemos imaginar as amostras mais amplas e a amplitude das classes ficando cada vez menor, e assim concluir que o polígono de frequência tende a se transformar em uma curva de frequência. Essa curva possibilita uma noção mais tendenciosa do comportamento da população ao invés da imagem real fornecida pelo polígono de frequência (CRESPO, 2002). Normalmente, as curvas de frequência têm as seguintes características: em forma de sino (simétrico ou assimétrico), em forma de J (normal ou invertido) e em forma de U. Por exemplo, a altura dos adultos é uma situação de curva em forma de sino, a distribuição de vencimentos é uma situação de curva em J, e as taxas de mortalidade por idade é uma situação de curva em U. Observando os gráficos da figura 5, podemos inferir que eles têm a forma de sino.

Agora apresentaremos o gráfico de linha e o de dispersão.

5.4 Gráfico de linha (ou em linha)

O **gráfico de linha** consiste em uma linha poligonal para representar uma série estatística. As séries são comumente conhecidas como séries temporais, cronológicas ou históricas. Essas séries são um conjunto de observações de uma mesma variável quantitativa (discreta ou contínua) obtidas

ao longo de um período de tempo. Assim esse gráfico representa a aplicação de funções em um sistema de coordenadas cartesianas (CRESPO, 2002).

Os gráficos de linha são bastante utilizados na identificação de tendências de variação (aumento ou diminuição) dos valores numéricos de uma dada informação (NAZARETH, 2003). Por exemplo, variação da cotação do dólar, índices de natalidade e índices de inflação são situações que, normalmente, são representadas por meio de gráficos de linha.

Para se construir esse tipo de gráfico, é necessário, inicialmente, definir o espaço cartesiano, sendo o eixo horizontal (x) destinado para a escala de tempo e o eixo vertical (y) destinado para representar as frequências. Em seguida, é preciso associar cada par ordenado (x,y) da série no espaço cartesiano, marcando, assim, um ponto. Após isso, traçar uma reta unindo todos os pontos.

Na figura 6, é possível visualizar um exemplo de gráfico de linha. Nesse gráfico, tem-se a variação da quantidade de vendas de *notebooks* entre os anos de 1999 e 2009. Pode-se observar que houve um acréscimo de mais de 100% entre os anos de 2001 e 2002. Ainda é possível verificar que entre os anos de 2007 para 2008 e de 2000 para 2001 houve um crescimento de igual tamanho.

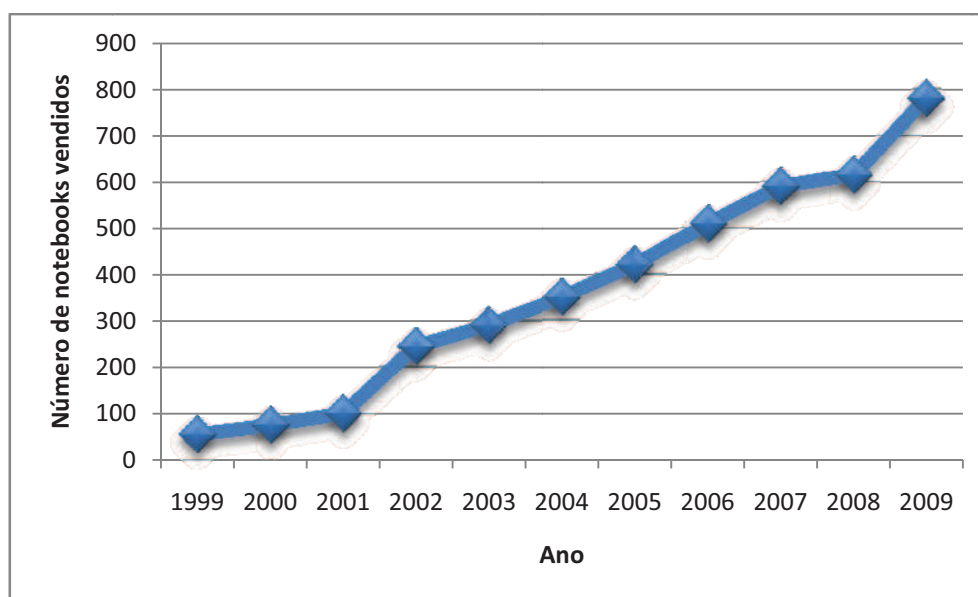


Figura 6: Gráfico de linha das vendas dos *notebooks*

Analisando esse gráfico, podemos concluir que a investigação de tendências temporais é facilitada com a exposição dos dados nesse formato visual. Desse modo, podemos, sobretudo, prever situações futuras de uma variável com base no seu comportamento anterior.

5.5 Gráfico de dispersão

O **gráfico de dispersão**, frequentemente utilizado para fins científicos, é uma representação da variação de uma variável quantitativa em função de outra variável (BUSSAB; MORETTIN, 2002). Assim, a partir de um gráfico de dispersão, é possível analisar a relação entre duas variáveis quantitativas por meio da plotagem dos valores dessas variáveis em um sistema de eixos. O eixo x é destinado para uma variável e o y para a outra variável.

Para se construir um gráfico de dispersão, é necessário associar as duas variáveis no espaço cartesiano, sendo uma para cada eixo. É importante lembrar que o detalhamento da escala dos eixos fica de acordo com a necessidade dos dados. Após isso, para cada par ordenado do conjunto, encontramos o ponto de encontro das duas variáveis e fazemos uma marcação (ponto ou quadrado).

Por exemplo, o quadro 1 apresenta dados sobre peso e altura de sete indivíduos hipotéticos. Considerando os valores dessas duas variáveis, o gráfico de dispersão correspondente seria o apresentado na figura 7.

Quadro 1: Dados sobre peso e altura de indivíduos

INDIVÍDUO	PESO (EM KG)	ALTURA (EM CM)
1	51	1,59
2	90	1,83
3	70	1,65
4	75	1,71
5	80	1,73
6	95	1,9
7	65	1,76

Nesse gráfico, o eixo horizontal representa a altura, e o vertical o peso dos indivíduos. O nível de detalhamento dos valores apresentados nos eixos x e y é variável, de acordo com a necessidade. Nesse caso, não há necessidade

de mostrar valores menores do que 1,55 cm no eixo horizontal, visto que não há ocorrência de indivíduos com estatura inferior a esse valor.

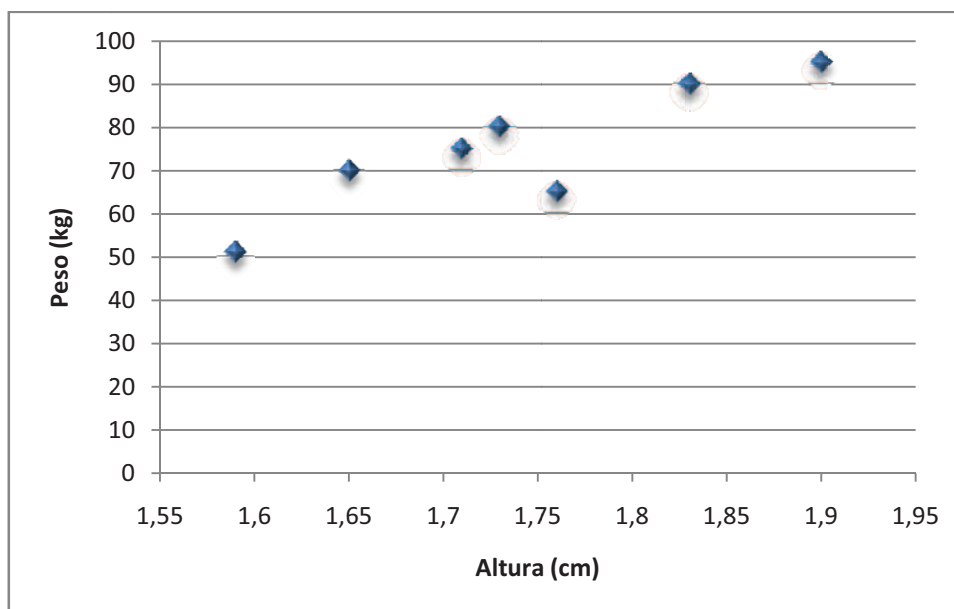


Figura 7: Gráfico de dispersão dos pesos e das alturas

De acordo com o gráfico exposto, podemos observar que há uma concentração de pessoas com altura entre 1,70 e 1,80 cm e com peso entre 60 e 80 kg. É possível ainda identificar rapidamente casos isolados, como o perfil da pessoa mais baixa e da mais alta.

Para concluir este capítulo, é importante ressaltar que os gráficos descritos são comumente utilizados para a tomada de decisões. Por isso precisamos conhecer os conceitos relacionados à sua construção para que possamos desenvolvê-los, bem como extrair informações.

Veremos, no próximo capítulo, as medidas de tendência central. Essas medidas são denominadas dessa forma pelo fato de os dados observados tenderem, em geral, a se agrupar em torno dos valores centrais. Assim aprenderemos como manipular essas medidas para as variáveis quantitativas. Até lá.

Resumo

O uso de representações gráficas para variáveis quantitativas é muito importante em um processo de tomada de decisão, pois possibilita a aquisição de informações de um modo visual e, muitas vezes, mais rápida. Ressaltamos que a presença ou não de classes em uma representação gráfica dependerá do objetivo do gráfico.

Entre os gráficos utilizados para representação de variáveis quantitativas, destacam-se os histogramas (conjunto de retângulos justapostos), os polígonos (gráficos de linha com frequência relativa ou acumulada), os gráficos de linha (linha poligonal para séries temporais) e os de dispersão (variação de duas variáveis quantitativas).

Atividades

1. De acordo com os dados da tabela de distribuição de frequência apresentada a seguir, construa o histograma, o polígono de frequência e o polígono de frequência acumulada correspondentes.

CLASSES DE IDADES	f_i
2 – 4	3
4 – 6	5
6 – 8	8
8 – 10	6
10 – 12	3

2. Sobre histogramas, analise as afirmações apresentadas a seguir e, depois, marque a alternativa que contém a sequência correta.

I – Os histogramas podem ser construídos tanto com frequências absolutas quanto com frequências relativas.

II – As alturas dos retângulos são proporcionais às frequências das classes. As larguras são iguais às amplitudes dos intervalos das classes.

III – Os histogramas são formas de descrição gráfica de dados qualitativos, organizados em classes de frequência.

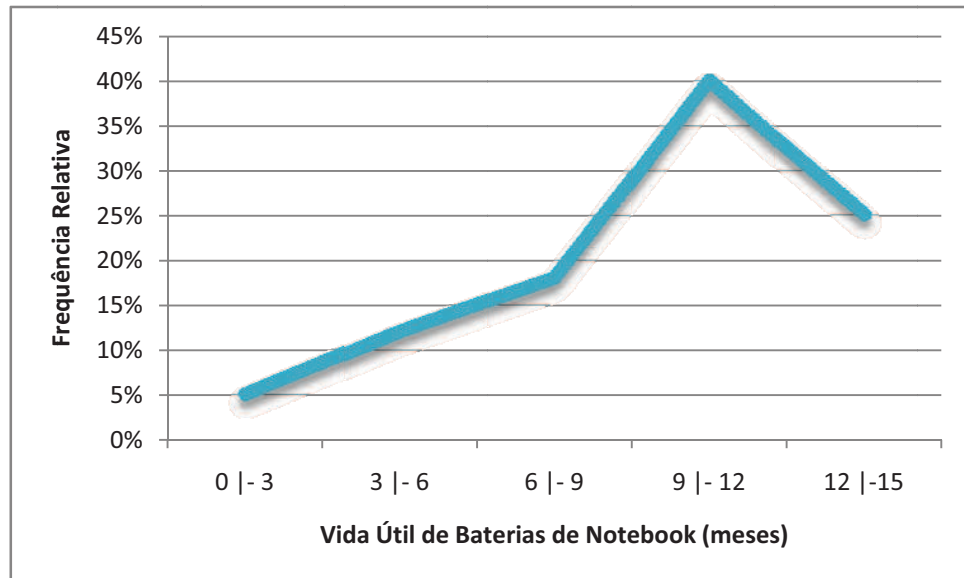
IV – Os histogramas são formas gráficas destinadas às representações de séries temporais ou cronológicas.

a) Somente (I) e (II) são verdadeiras.

b) Somente (II) e (III) são verdadeiras.

- c) Somente (I) e (III) são verdadeiras.
- d) Somente (I) e (IV) são verdadeiras.
- e) Somente (II) e (IV) são verdadeiras.

3. Leia as informações do seguinte gráfico e escolha a alternativa correta.



- a) Com base no gráfico, podemos afirmar que 40% das baterias de notebooks têm vida útil de 12 meses.
- b) Com base no gráfico, podemos afirmar que 25% das baterias de notebooks têm vida útil entre 6 e 9 meses.
- c) Com base no gráfico, podemos afirmar que 10% das baterias de notebooks têm vida útil entre 0 e 3 meses.
- d) Com base no gráfico, podemos afirmar que 75% das baterias de notebooks têm vida útil entre 0 e 12 meses.
- e) Com base no gráfico, podemos afirmar que 65% das baterias de notebooks têm vida útil entre 3 e 9 meses.

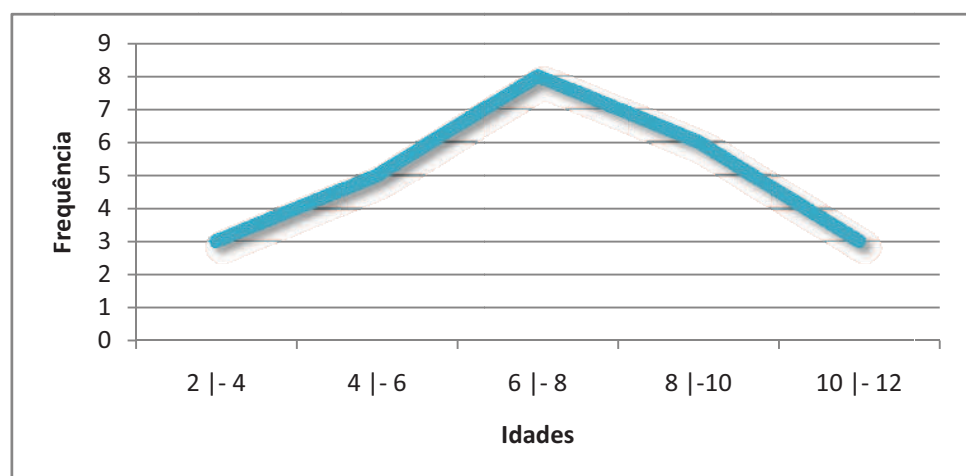
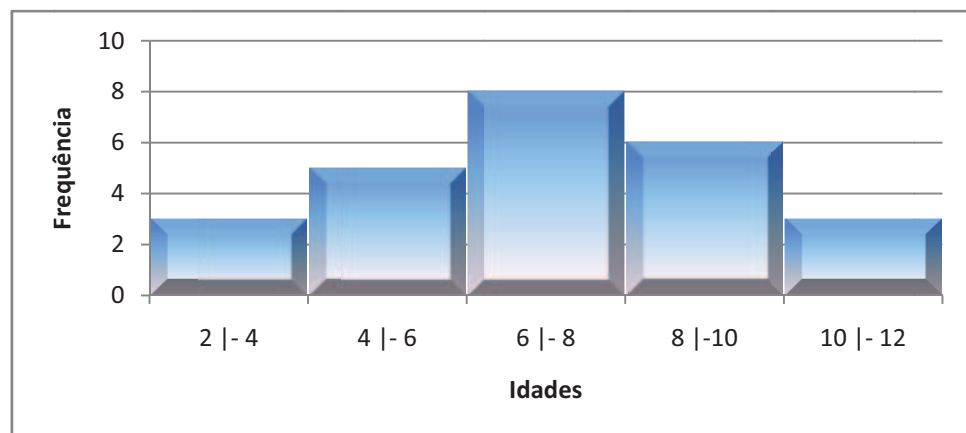
4. Os principais passos para a construção de um gráfico de linhas são:

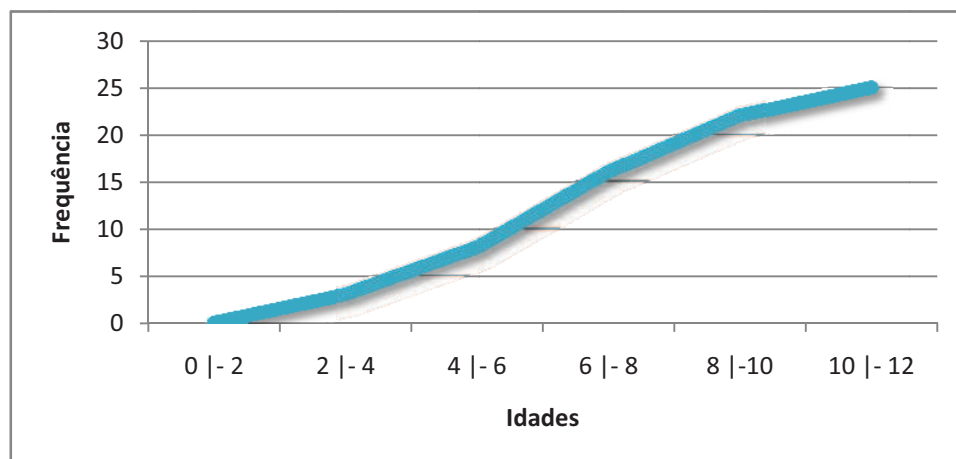
- a) traçar uma reta horizontal para cada par ordenado da série e, depois, unir todas as retas.

- b) marcar um ponto para cada par ordenado da série no espaço cartesiano e, depois, traçar a reta unindo todos os pontos.
- c) criar um retângulo para cada par ordenado da série e, depois, esboçar uma reta reunindo-os.
- d) desenhar uma linha vertical em cada valor da série apresentada no eixo das abscissas (tempo) e, depois, traçar uma reta associando todas as linhas.
- e) marcar no plano cartesiano cada par (x,y) e, depois, alinhá-los à esquerda.

Comentário das atividades

A **atividade 1** quis verificar a habilidade para construir histogramas, polígonos de frequência e polígonos de frequência acumulada. De acordo com os dados apresentados, os gráficos a serem construídos, respectivamente, são:





Na **atividade 2**, a afirmação (I) é verdadeira, porque histogramas podem ser elaborados por frequências absolutas ou relativas (ver seção 5.1). A afirmação (II) é verdadeira, porque a altura dos retângulos está relacionada com o valor da frequência da classe, e a largura representa a amplitude. A afirmação (III) é falsa, porque histogramas são formas de descrição gráfica de dados quantitativos e não qualitativos. A afirmação (IV) é falsa, porque gráficos de linhas são indicados para representação de séries temporais ou cronológicas. Logo a resposta certa é a alternativa **(a)**, porque apresenta as afirmações (I) e (II) como verdadeiras.

Para responder à **atividade 3**, devemos analisar as porcentagens correspondentes das classes. A alternativa **(d)** é a correta, porque 75% das baterias de notebooks têm vida útil entre 0 e 12 meses, ou seja, 25% baterias de notebooks têm vida útil entre 12 e 15 meses. A alternativa **(a)** é incorreta, porque 40% das baterias de notebooks têm vida útil entre 9 e 12 meses, e não somente de 12 meses. A alternativa **(b)** é incorreta, porque 25% das baterias de notebooks têm vida útil entre 12 e 15 meses, e não entre 6 e 9 meses. A alternativa **(c)** é incorreta, porque 5%, e não 10%, das baterias de notebooks têm vida útil entre 0 e 3 meses. A alternativa **(e)** é incorreta, porque cerca de 30% das baterias têm vida útil entre 3 e 9.

Na **atividade 4**, a alternativa **(b)** é correta, porque, para cada par ordenado da série, um ponto é marcado no espaço cartesiano e, depois, deve-se traçar a reta unindo todos os pontos. A alternativa **(a)** é incorreta,

porque nenhuma reta é traçada para cada par ordenado. A alternativa **(c)** é incorreta, porque nenhum retângulo é desenhado. A alternativa **(d)** é incorreta, porque nenhuma linha vertical é desenhada para cada valor da série. A alternativa **(e)** é incorreta, porque não é necessário alinhar os pontos marcados à esquerda ou à direita.

Referências

BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística básica**. São Paulo: Saraiva, 2002.
CRESPINO, Antônio Arnot. **Estatística fácil**. São Paulo: Saraiva, 2002.
MILONE, Giuseppe. **Estatística geral e aplicada**. São Paulo: Pioneira Thomson, 2004.
NAZARETH, Helenalda. **Curso básico de estatística**. São Paulo: Ática, 2003.