

Estatística descritiva

2.1 Distribuições de frequência e seus gráficos

Distribuições de frequência → Gráficos de distribuições de frequência

Quando um dado tem muitas entradas, pode ser difícil de ver padrões. Nesta seção, você aprenderá como organizar conjuntos de dados agrupando os dados em intervalos chamados de classes e formando uma distribuição de frequência. Você também aprenderá como usar as distribuições de frequência para a construção de gráficos.

Definição

A distribuição de frequência é uma tabela que mostra classes ou intervalos das entradas de dados com uma contagem do número de entradas em cada classe. A frequência f de uma classe é o número de entrada de dados em uma classe.

Exemplo de uma distribuição de frequência

Classe	Frequência f
1-5	5
6-10	8
11-15	6
16-20	8
21-25	5
26-30	4

Limites inferiores da classe:

1, 6, 11, 16, 21, 26

Limites superiores da classe:

1, 6, 11, 16, 21, 26

Largura da classe: 5

Instruções

Construindo uma distribuição de frequência com base em um conjunto de dados

1. Decida o número de classes para serem incluídas na distribuição de frequência. O número de classes deve estar entre 5 e 20; caso contrário, pode ser difícil detectar os padrões.
2. Encontre a largura da classe como a seguir. Determine a amplitude dos dados, divida a amplitude pelo número de classes e *arredonde para o próximo número que seja conveniente*.
3. Encontre os limites de classe. Você pode usar a entrada de dados mínima como o limite inferior da primeira classe. Para encontrar os limites inferiores restantes, adicione a largura da classe ao limite inferior da classe precedente. Então, encontre o limite superior da primeira classe. Lembre-se que as classes não se sobrepõem. Encontre os limites superiores de classe restantes.
4. Faça uma marca de contagem para cada entrada de dados em ordem da classe apropriada.
5. Conte as marcas para encontrar a frequência total f para cada classe.

Exemplo 1

Construindo uma distribuição de frequência com base em um conjunto de dados

O conjunto de dados amostrais a seguir lista o número de minutos que 50 usuários de Internet gastam na rede durante sua mais recente sessão. Construa uma distribuição de frequência para as sete classes.

50	40	41	17	11	7	22	44	28	21	19	23	37	51	54	42	86
41	78	56	72	56	17	7	69	30	80	56	29	33	46	31	39	20
18	29	34	59	73	77	36	39	30	62	54	67	39	31	53	44	

Solução

1. O número de classes (7) é dado no problema.

2. A entrada de dados mínima é 7 e a entrada de dados máxima é 86, então, a amplitude é $86 - 7 = 79$. Divida a amplitude pelo número de classes e arredonde para encontrar a largura da classe.

$$\begin{array}{l} \text{Largura da classe} = \frac{79}{7} \\ \approx 11,29. \end{array} \quad \begin{array}{l} \frac{\text{amplitude}}{\text{número de classes}} \\ \text{arredonde para 12} \end{array}$$

3. A entrada mínima de dados é um limite inferior conveniente para a primeira classe. Para encontrar os limites inferiores das seis classes restantes, adicione a largura de classe 12 ao limite inferior de cada classe anterior. O limite superior é 18, que é um a menos do que o limite inferior da segunda classe. Os limites superiores das outras classes são $18 + 12 = 30$, $30 + 12 = 42$, e assim por diante. Os limites inferiores e superiores para todas as sete classes são mostrados.

4. Faça as marcas de contagem para cada entrada de dados na classe apropriada. Por exemplo, a entrada de dado 51 está na classe 43–54, então faça uma marca de contagem nesta classe. Continue até que você tenha feito uma marca para cada uma das 50 entradas de dados.

5. O número de marcas de contagem para uma classe é a frequência para esta classe.

Distribuição de frequência para uso de Internet (em minutos)

Minutos on-line

Número de usuários

Classe	Marca	Frequência, f
7-18	HHH I	6
19-30	HHH HHH	10
31-42	HHH HHH III	13
43-54	HHH III	8
55-66	HHH	5
67-78	HHH I	6
79-90	II	2
		$\Sigma f = 50$

Note que a soma das frequências é idêntica ao número da amostra.

Depois de construir uma distribuição de frequência padrão tal como a do Exemplo 1, você pode incluir diversas características adicionais que ajudarão a fornecer um melhor entendimento dos dados. Essas características (pontos médios, frequência relativa e frequência acumulada) podem ser incluídas como colunas adicionais em sua tabela.

Definição

O **ponto médio** de uma classe é a soma dos limites inferiores e superiores da classe dividida por dois. O ponto médio é, às vezes, chamado de *marca da classe*.

$$\text{Ponto médio} = \frac{(\text{limite inferior da classe}) + (\text{limite superior da classe})}{2}$$

A **frequência relativa** de uma classe é a porção ou porcentagem de dados que está em determinada classe. Para encontrar a frequência relativa de uma classe, divida a frequência f pelo tamanho n da amostra.

$$\begin{aligned}\text{Frequência relativa} &= \frac{\text{Frequência da classe}}{\text{Tamanho da amostra}} \\ &= \frac{f}{n}\end{aligned}$$

A **frequência acumulada** de uma classe é a soma da frequência para aquela classe e todas as anteriores. A frequência acumulada da última classe é igual ao tamanho n da amostra.

Os pontos médios restantes, as frequências relativas e acumuladas são mostradas na distribuição de frequência expandida que vem a seguir.

Distribuição de frequência para o uso da Internet (em minutos)

Minutos on-line	Classe	Frequência f	Ponto médio	Frequência relativa	Frequência acumulada	Porção de usuários
	7-18	6	12,5	0,12	6	
	19-30	10	24,5	0,2	16	
	31-42	13	36,5	0,26	29	
	43-54	8	48,5	0,16	37	
	55-66	5	60,5	0,1	42	
	67-78	6	72,5	0,12	48	
	76-90	2	84,5	0,04	50	
Número de usuários		$\Sigma f = 50$		$\Sigma \frac{f}{n} = 1$		

Gráficos de distribuições de frequência

Às vezes, é mais fácil identificar padrões de um conjunto de dados olhando o gráfico da distribuição de frequência. Um desses gráficos é o histograma de frequência.

Definição

Um **histograma de frequência** é um diagrama de barras que representa a distribuição de frequência de um conjunto de dados. Um histograma tem as seguintes propriedades:

1. A escala horizontal é quantitativa e mede os valores dos dados.
2. A escala vertical mede as frequências das classes.
3. As barras consecutivas devem estar encostadas umas nas outras.

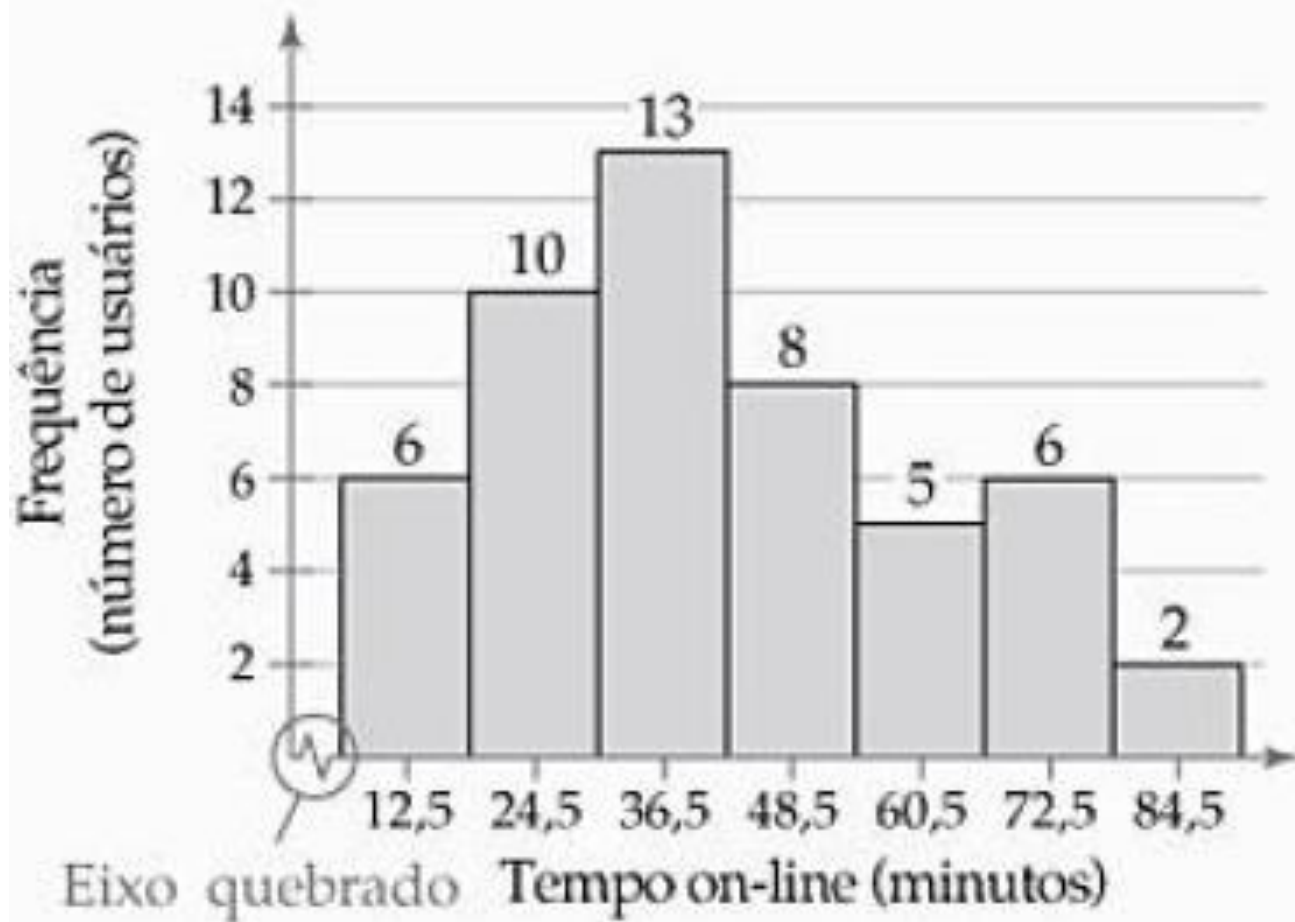
Em virtude de as barras consecutivas no histograma estarem encostadas, elas devem começar e terminar nas fronteiras da classe ao invés de em seus limites. As **fronteiras das classes** são números que separam as classes *sem* formar lacunas entre elas. Você pode marcar a escala horizontal tanto nos pontos médios quanto nas fronteiras das classes, conforme o Exemplo 3.

Classe	Fronteiras de classe	Frequência f
7-18	6,5-18,5	6
19-30	18,5-30,5	10
31-42	30,5-42,5	13
43-54	42,5-54,5	8
55-66	54,5-66,5	5
67-78	66,5-78,5	6
79-90	78,5-90,5	2

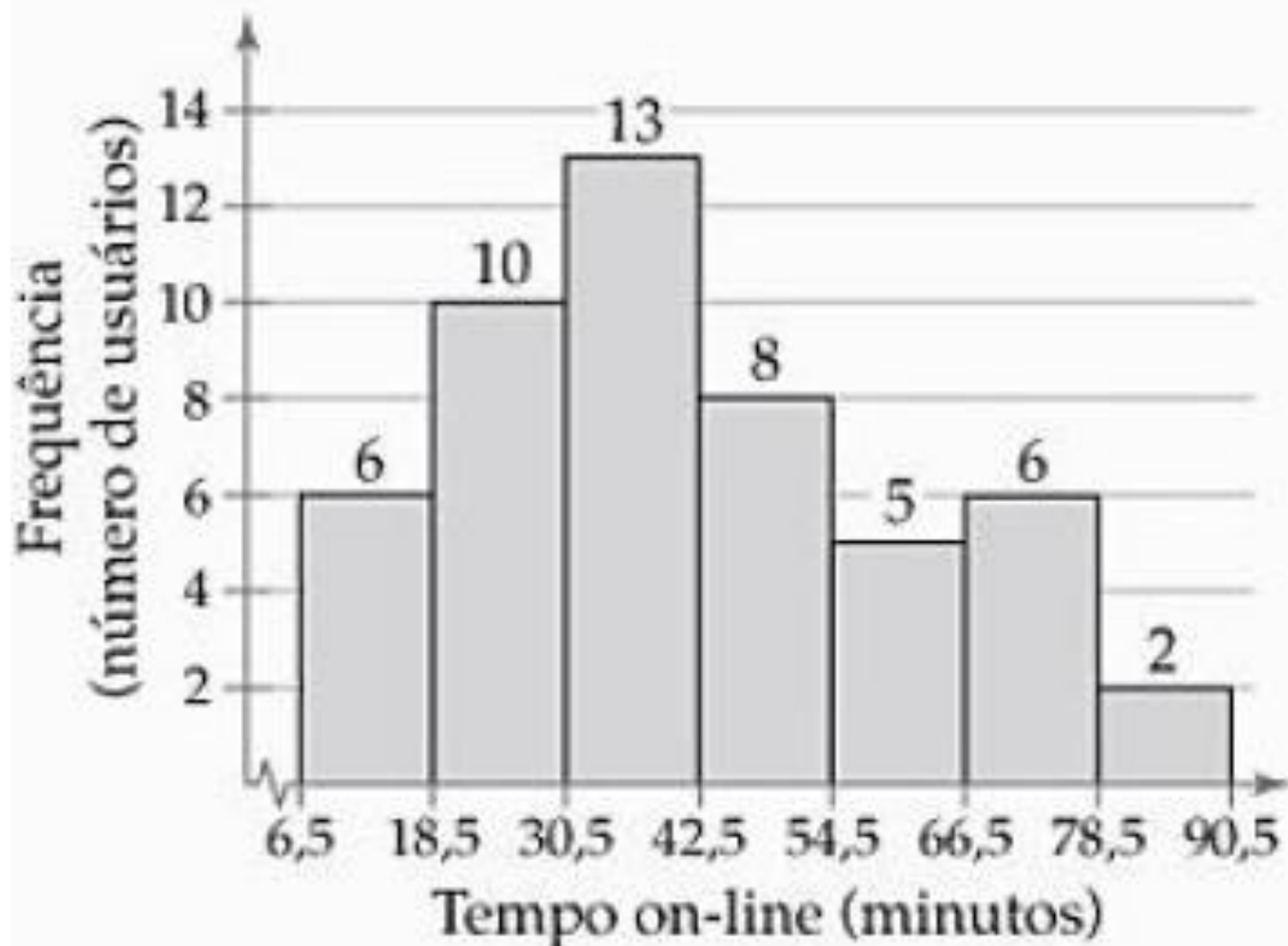
Fronteira inferior da primeira classe = $7 - 0,5 = 6,5$

Fronteira superior da primeira classe = $18 + 0,5 = 18,5$

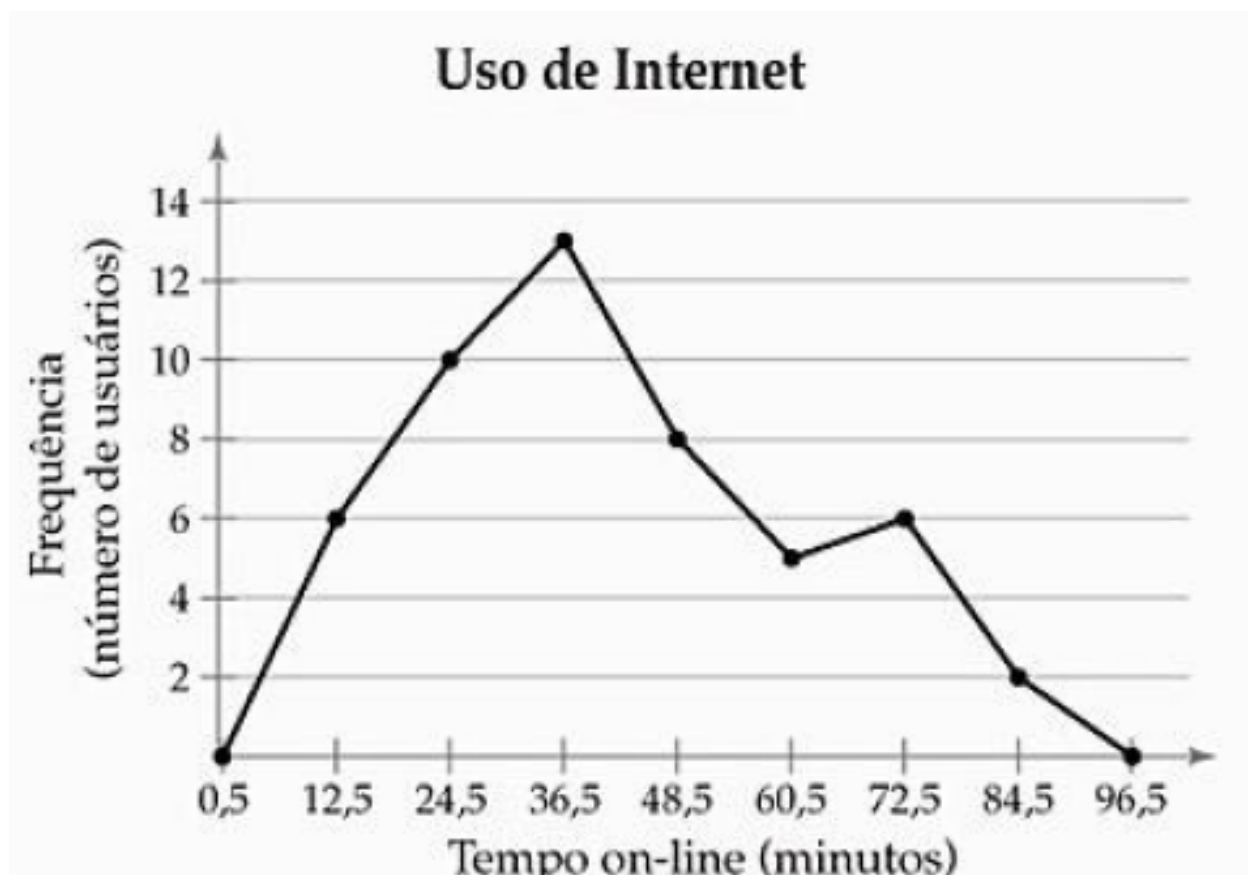
Uso da Internet (rotulado com pontos médios das classes)



Uso da Internet (rotulado com fronteiras de classes)



Outra maneira de representar graficamente a distribuição de frequência é usar um polígono de frequência. Um **polígono de frequência** é um gráfico de linhas que enfatiza as mudanças contínuas nas frequências.



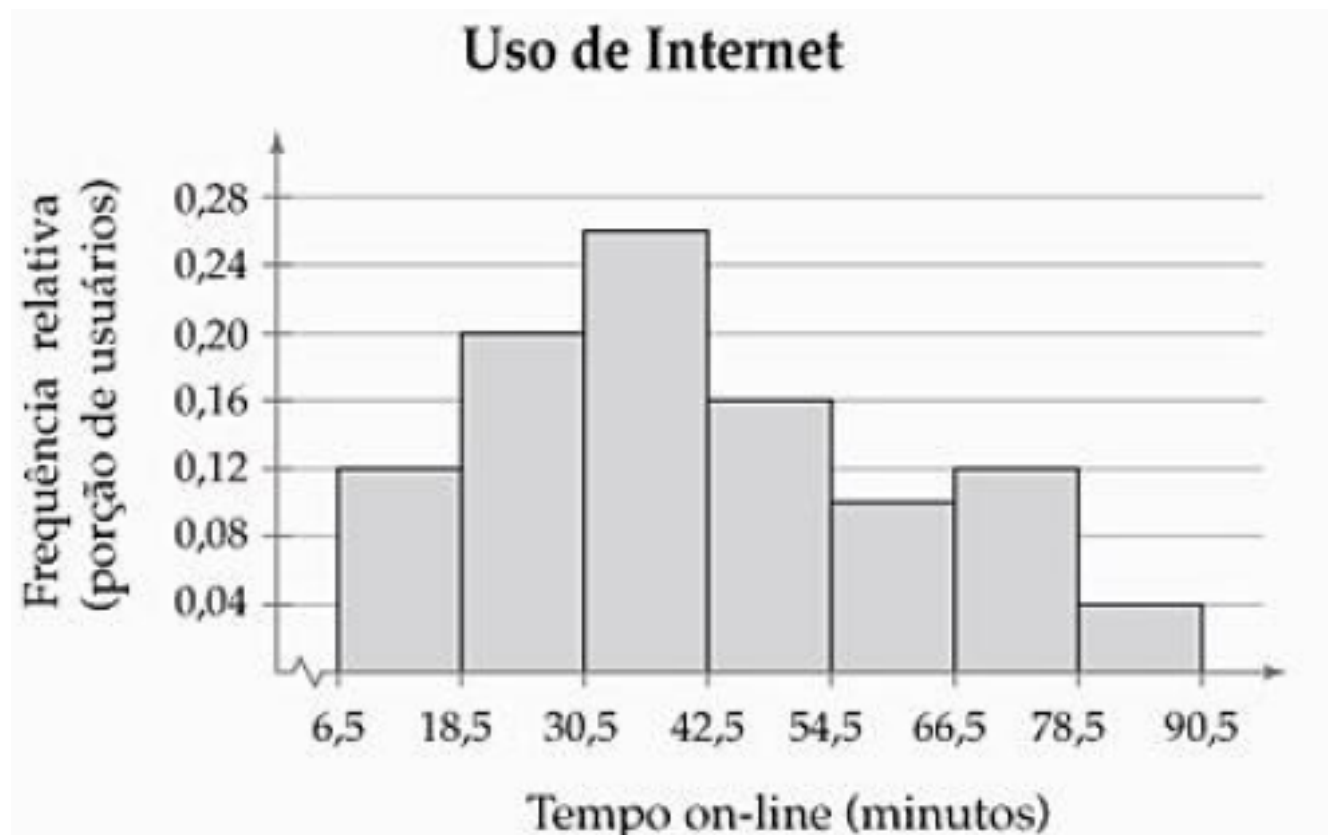
Interpretação

Podemos ver que a frequência de usuários aumenta até 36,5 minutos e depois decai.

Um **histograma de frequência relativa** tem a mesma forma e a mesma escala horizontal do histograma de frequência correspondente. A diferença é que a escala vertical mede as frequências *relativas* e não as frequências.

Solução

O histograma de frequência relativa é mostrado. Note que a forma do histograma é a mesma forma do histograma de frequência construído no Exemplo 3. A única diferença é que a escala vertical mede as frequências relativas.



Se você quiser descrever o número de entradas de dados que são iguais a ou menores que certo valor, você pode facilmente fazê-lo construindo um gráfico de frequência acumulada.

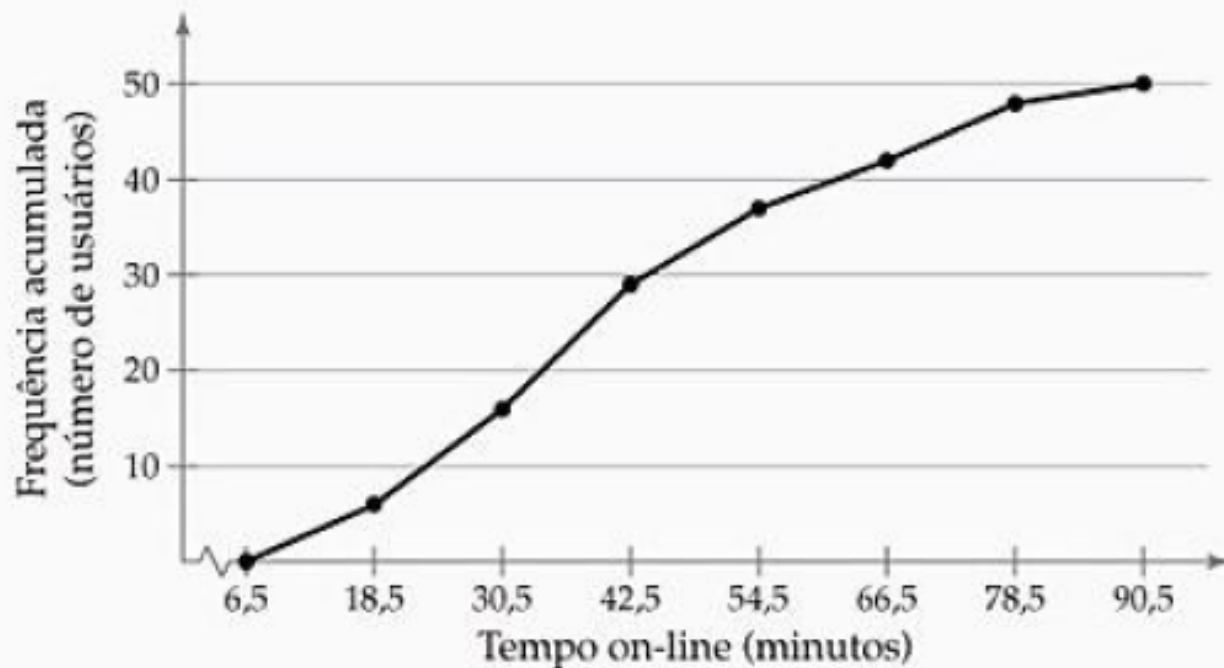
Definição

Um **gráfico de frequência acumulada** ou **ogiva** (pronuncia-se o'jiva) é um gráfico de linhas que mostra as frequências acumuladas de cada classe em sua fronteira da classe superior. As fronteiras superiores são marcadas no eixo horizontal e as frequências acumuladas são marcadas no eixo vertical.

Construindo uma ogiva

Desenhe uma ogiva para a distribuição de frequência do Exemplo 2. Estime quantos usuários gastam 60 minutos ou menos on-line durante sua última sessão. Também use o gráfico para estimar quando ocorre o maior aumento no uso.

Uso da Internet



Maior fronteira de classe	f	Frequência acumulada
18,5	6	6
30,5	10	16
42,5	13	29
54,5	8	37
66,5	5	42
78,5	6	48
90,5	2	50

Interpretação

Com base na ogiva, podemos ver que por volta de 40 usuários gastam 60 minutos ou menos durante sua última sessão. É evidente que o maior aumento no uso ocorre entre 30,5 e 42,5 minutos, já que o segmento de linha é mais inclinado entre essas duas fronteiras de classe.