



Resumindo os números da entrega

Transcrição

Agora que usamos um exemplo simples para compreender os conceitos, vamos para o caso dos números das entregas. Temos uma tabela com informações semelhantes as que levantamos anteriormente. Vamos preenche-las com os números das colunas de `Distancia` , `Tempo` e `Velocidade` . Para o valor mínimo, usaremos a função `Min` , para os quartis, usaremos `QUARTILE` e especificaremos a qual nos referimos (primeiro, segundo ou terceiro). O valor máximo encontraremos com a função `MAX` , a média, com a `AVERAGE` , e para o desvio padrão, utilizaremos `STDEV` .

	Distancia (km)	Tempo (minutos)	Velocidade (km/h)
Valor Mínimo	10	8.8	14.8
1o Quartil	35	47.5	40
2o Quartil/Mediana	61	77.5	49
3o Quartil	92	117.2	55
Valor Máximo	120	336.2	76.0
Média	63.48	86.2	47.8
Desvio Padrão	31.6	53.1	11.6

Analisando os valores, veremos que a média e mediana da `Distancia` estão próximas. Isto significa que o histograma está bem distribuído, sem deslocamentos para esquerda ou direita. Vemos que 25% das pessoas estão a uma distância entre 92 e 120 km. Talvez, seja o caso de criar um ponto de distribuição em outro local. Também, temos uma quantidade de 25% de entregas com um distância inferior a 35 km.

Passando para a coluna `Tempo` , percebemos que o valor mínimo é 8 minutos. mas o 1º quartil foi para 47 minutos. Até o 3º quartil, ou seja 75% dos casos, j

chegaremos a 117 minutos. O histograma será deslocado para esquerda. A diferença do valor do 3º quartil para o valor máximo é superior ao dobro. Também é possível notar que 50% das entregas são feitas entre 1h e 2h. Significa que em uma jornada de 8 horas, um entregador fará de 4 a 8 entregas por dia. Observando o histograma, é preferível dizer que a média de tempo está mais próxima do valor da mediana, porque o gráfico tende para a esquerda. Mas existem alguns casos extremos, nos quais a entrega demorou muito tempo.

Ao observarmos os dados da terceira coluna, referente à velocidade. Vemos que o valor mínimo foi 14,8 km/h. Também percebemos que o 1º, o 2º, e o 3º quartil estão próximos, com apenas 15 km/h de variação. Percebemos esta concentração dos dados no histograma. O valor máximo é 76 km/h.

O **desvio padrão** nos dá uma ideia de espalhamento das informações. Quando ele será útil? Quando repetirmos o experimento e tivermos uma tabela semelhante, com novos valores de desvio padrão. Dependendo da distribuição, pode resultar em alguns percentuais. Como estamos analisando os dados empiricamente, o desvio é usado para comparação. Se repetíssemos a experiência, observaríamos se houve ou não um aumento no desvio padrão. No caso de um aumento, significa que os valores dos dados estão mais dispersos. Se houve uma diminuição, significa que os valores estão mais concentrados e há mais uniformidade nos dados.

Com esta tabela, cumprimos a missão de resumir os dados. Podemos incluir estas informações da tabela em um relatório, e conseguimos representar de uma forma numérica o que foi apresentado nos histogramas.