



Atividade: Comparação das diferentes categorias na maratona

Habilidades

EM13MAT409 Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos, como o histograma, o de caixa (box-plot), o de ramos e folhas, reconhecendo os mais eficientes para sua análise.

Para o professor

Objetivos específicos

OE1 Comparar distribuições de uma mesma variável para grupos distintos a partir dos histogramas.

OE2 Perceber a necessidade de usar a mesma escala nos eixos do histograma, para tornar os mesmos comparáveis.

Observações e recomendações

Esta atividade introduz os elementos necessários para a comparação de dois histogramas, a saber: mesmas escalas nos eixos e colunas de frequências relativas.

Os histogramas são apresentados com uma série de perguntas de discussão que podem motivar a formulação do conceito de dispersão de forma intuitiva, que será trabalhado na seguinte seção. Além de mostrar como apenas as medidas de posição não dizem suficiente sobre uma distribuição. As perguntas não têm respostas fechadas, têm o intuito de gerar uma discussão sobre os assuntos já colocados.

Atividade

Observe os histogramas da [figura 1](#) referentes aos tempos de chegada das categorias “cadeira de rodas” e “triciclo de mão” na Maratona de Nova Iorque em 2017. Nesse caso, os tempos de chegada foram convertidos para horas e referem-se ao total de participantes que completaram a prova: 51 na categoria cadeira de rodas e 69 na categoria triciclo de mão.

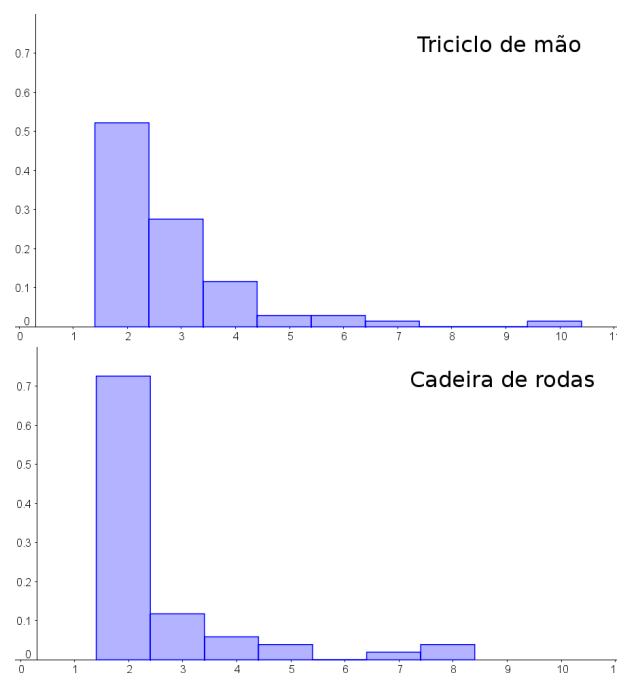


Figura 1: Histogramas comparativos das quatro modalidades da maratona de Nova Iorque 2017

- Compare as escalas utilizadas na construção destes histogramas, tanto no eixo horizontal, como no eixo vertical. O que você observou?
- Em qual categoria se encontra o atleta que completou a maratona no maior tempo?
- Você consegue estimar o tempo médio destas categorias observando os histogramas? Você acha que os tempos médios dessas duas categorias serão muito diferentes dos tempos médios das categorias de homens e de mulheres (atividade Categoria Homens na Maratona)?
- Observe o quadro a seguir e marque as médias (em horas) nos histogramas. Comente sobre a posição da média em cada caso e sobre a simetria ou assimetria de cada distribuição de dados.

Tabela 1: Média das quatro categorias da maratona de Nova Iorque 2017

Categoria	Cadeira de rodas	Triciclo de mão
Média	2,59	2,73

- Observe que as médias não são muito diferentes. Se você conhecesse apenas a média, seria capaz de perceber a forma destes histogramas? Por quê?
- Comparando os dois histogramas, qual distribuição apresenta maior dispersão? Por quê?
- A partir dos esquemas de cinco números para os tempos de chegada nas categorias cadeira de rodas e triciclo de mão, informados a seguir, construa os respectivos boxplots sem sinalização de valores discrepantes, usando a mesma escala de valores para as duas categorias.

Categoria	Mínimo	Q_1	Q_2	Q_3	Máximo
Cadeira de rodas	1,62	1,8	2,09	2,68	7,81
Triciclo de mão	1,48	1,78	2,38	3,09	9,52

Solução:

- a) As escalas horizontal e vertical são coincidentes em ambos os histogramas para permitir a comparação. Isto foi discutido no capítulo **A Natureza da Estatística**.
- b) Observando-se os histogramas o maior tempo está na categoria de Triciclo de mão.
- c) A simples visualização pode remeter a estimativa de 4 horas para estas categorias, valor bem maior do que as médias para homens e mulheres (ativ-maratona-categoria-homens). No entanto, este valor excede muito as médias reais, apresentadas. Isto mostra que estimar o centro de equilíbrio de uma distribuição a partir do histograma não é trivial.
- d) Ambos os histogramas apresentam assimetria à direita: grande concentração de dados à esquerda e forma alongada para a direita.
- e) Não é possível, a informação apenas da média é insuficiente para caracterizar a forma do histograma. Outras informações são necessárias para isto.
- f) O triciclo de mão apresenta maior dispersão, considerando a amplitude.
- g) Na figura a seguir veja uma construção dos boxplots sem sinalização dos valores discrepantes.

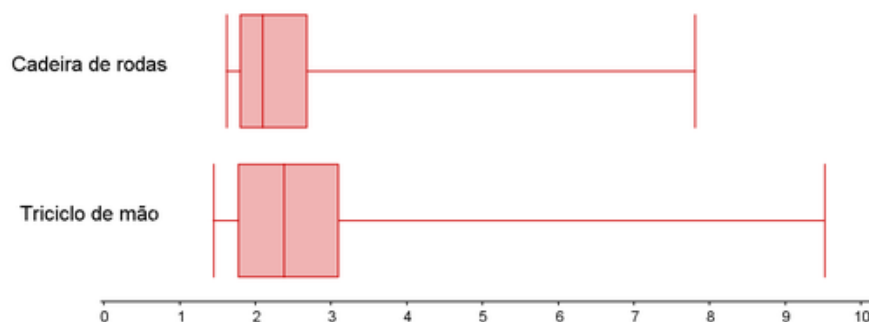


Figura 2: Boxplots sem sinalização de valores discrepantes para os tempos de chegada em horas das categorias “cadeira de rodas” e “triciclo de mão”