

Por que controlar?

Dada uma medida (presença de pessoas, temperatura, umidade, nível de luminosidade etc.) como sabemos se precisamos nos preocupar? Se precisamos nos preocupar, qual o nível dessa preocupação (criticidade)?

Ou ainda nem para preocupação, mas apenas para gerenciar algo, mudar estratégia de negócio, enxergar novas possibilidades.

“Se você não pode medir, não pode gerenciar”

Peter Drucker

Antes de tudo: a Mediana

Vamos supor que precisamos estudar uma nova campanha de marketing para atrair novos alunos à SPTEch. Se tivermos uma amostra da população com a seguinte idade:

21, 22, 20, 19, 19, 18, 19

Podemos calcular a média de aproximadamente 20 anos (19,71 anos para ser mais exato, mas quem fala que tem 0,71 anos?) e com isso podemos adotar uma estratégia de marketing que seja direcionada para um público mais jovem, com uma linguagem apropriada, imagens, músicas e talvez até seres famosos para esses seres.

Entretanto como sabemos que a média está adequada para nossa análise? E se tivermos um valor discrepante? Se por exemplo um aluno que esteja transitando de carreira faça parte da nossa amostra?

21, 22, 20, 19, 19, 18, 19, **60**

Neste caso nossa média salta de 20 para 25 anos (ou 24,75 para ser mais Caxias¹), o que implicaria numa estratégia um tanto quanto diferente (5 anos hoje em dia altera bastante o comportamento!). Chegamos que neste conjunto a média não é de muita valia, entretanto percebemos facilmente o quão deslocado o 60 é nessa lista.

Mas e numa amostra de 20 medidas? 50 medidas? Só de 3 turmas podemos ter mais de 210 medidas (isso do 1º semestre), 1000 medidas? Cada vez mais o olho humano

¹ Em homenagem ao Duque de Caxias. Termo relacionado a uma pessoa certinha pois essa deve ter o mesmo temperamento do duque de Caxias, o marechal Luís Alves de Lima e Silva (1803-1880), que era um chefe severo e minucioso

não consegue distinguir a validade ou **qualidade** da média calculada. Para podermos averiguar a qualidade da média, calculamos a **mediana** que é uma medida geográfica da nossa amostra, em termos gerais ela divide nossa amostra em partes iguais, ou seja, a quantidade de valores antes da mediana é igual ou aproximadamente igual à quantidade de valores depois da mediana.

- Para o cálculo devemos primeiramente ordenar o conjunto de dados em ordem crescente.

Conjunto 1:

18, 19, 19, 19, 20, 21, 22

Conjunto 2:

18, 19, 19, 19, 20, 21, 22, 60

- Se o número de elementos for par, então a mediana é a média dos dois valores centrais.
- Se o número de elementos for ímpar, então a mediana é o valor central.

Conjunto 1, $n = 7$ (número de elementos), logo a mediana (MED) = 19 (valor central). Já para o conjunto 2, $n = 8$, logo a MED = $(19 + 20) / 2 = 19,5$.

Comparando os valores das medianas aos calculados da média obtemos que a média do conjunto 1 tem uma maior qualidade maior. O segundo conjunto temos uma média deslocada da mediana, causada pelo 60, percebemos então quais números são **outliers** (ou fora de série, discrepantes, longe da população etc.)

De forma fina e elegante a mediana corresponde uma ordenada que separa a curva de densidade em duas partes, cada uma delas com área igual a $\frac{1}{2}$ no caso de uma distribuição contínua. [1]

De uma outra forma, mais saborosa, separa a pizza (ou o bolo) em duas partes iguais em quantidade de pepperoni (ou brigadeiro na cobertura).

Nem só de mediana vive o controle

Vimos como separar uma amostra em duas partes iguais, para por aí? Obviamente que não, no mundo encantado da matemática [2] podemos utilizar outras medidas para melhor gerenciar nosso sistema.

Os percentis são uma dessas medidas que tornam conveniente o controle. Afinal quem gostaria de pegar um extintor com o prédio pegando fogo e ele estar nessa marcação (totalmente descarregado):

Figura 1: Indicador Gauge para extintor de incêndio



Fonte 1: [3]

Carregado demais também pode ser um problema (mas isso é uma discussão para uma aula sobre vasos de pressão). A imagem acima demonstra níveis de aviso dada uma pressão presente no interior do extintor.

No caso do controle de alguma medida se torna interessante separar as medições por meio de ordenadas que separem a amostra em determinadas porcentagens. Sendo os valor mínimos e máximos os extremos que concentram 100% da nossa amostra.

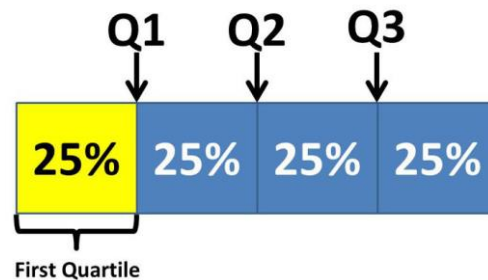
Trabalhando com percentis, portanto, podemos definir diversas divisões de nossa amostra. Uma divisão útil pode ser em **quartis**. Mas o que são quartis? Bem, um quartil é um quarto, ou $\frac{1}{4}$. Se dividirmos 1 por 4 chegamos em 0,25. Logo, os quartis dividem nossa amostra em partes separadas de 25% em 25%.

Em termos mais simples:

- Valor mínimo: o menor valor medido ou quartil 0.
- 1º quartil: até esse ponto existem 25% dos nossos dados coletados

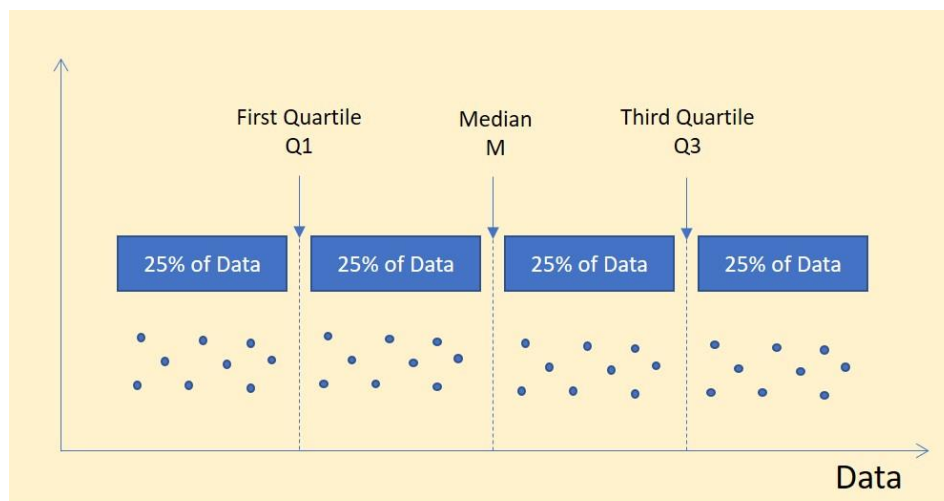
- 2º quartil: até esse ponto existem 50% dos nossos dados coletados (opa! Isso tem cara de mediana! E tem mesmo, é a mediana) pois temos 25% + 25%
- 3º quartil: até esse ponto existem 75% dos nossos dados coletados
- Valor máximo: o maior valor medido ou 4º quartil, 100% dos valores coletados estão sob esse número

Figura 2: Distribuição de quartis



Fonte 2: [4]

Figura 3: Distribuição de quartis de uma amostra



Fonte 3: [5]

Para calcular a posição do quartil podemos utilizar o seguinte conceito (sendo n o número de dados da amostra e a amostra ordenada de forma crescente:

- 1º quartil: arredondar $0,25 * (n + 1)$
- 3º quartil: arredondar $0,75 * (n + 1)$

Para nossos conjuntos:

Conjunto 1:

$$N = 7$$

> 1º quartil: posição 2 = 19

> 3º quartil: posição 6 = 21

Conjunto 2:

$N = 8$

> 1º quartil: posição $2,25 \sim 2 = 19$

> 3º quartil: posição $6,75 \sim 7 = 22$

Melhorando a precisão

Como nossas medidas são contínuas, podem ser mostradas de forma decimal (diferente da idade, a menos que você seja um infante com 8 meses e 3 semanas de vida) e numa aplicação real teremos uma enorme quantidade de dados, podemos nos dar ao luxo de utilizar ferramentas computacionais para o cálculo. O excel é uma das boas!

Podemos então calcular com as seguintes fórmulas:

- **MED(<matriz>):** calcula a mediana de uma determinada amostra [6]
- **QUARTIL.EXC(<matriz>;<quarto>):** calcula o quartil especificado de uma determinada amostra excluindo os valores mínimo e máximo (de certa forma faixas de valores mais conservadores são calculados assim) [7]
- **QUARTIL.INC(<matriz>;<quarto>):** calcula o quartil especificado de uma determinada amostra incluindo os valores mínimo e máximo. [8]
- **MÉDIA(<matriz>):** Calcula a média de uma determinada amostra [9]

Onde uso?

Com os valores coletados pelas medições é formada uma população de dados. Com essa população crio faixas de alertas baseados nos quartis, tendo por exemplo níveis de alertas como no do medidor do extintor. Esse valores então entram na lógica em JS para que o frontend reflita os alertas a partir das medidas provenientes do backend.

Um exemplo

Dado um projeto, é discutido que as faixas ideais de trabalho são entre 20°C e 29°C. Entretanto, sabemos que o gestor **não morre de más notícias, mas sim de sustos**, então não podemos avisar que a temperatura está acima de 29°C quando de fato estiver acima dos 29°C, precisamos ir avisando aos poucos, para que dê tempo de tomar alguma atitude antes mesmo de passar as temperaturas críticas.

40 medidas foram tomadas resultando no conjunto de dados abaixo.

Im35temp: Dados brutos






27,00	27,00	25,00	20,00
20,00	25,00	27,00	25,00
23,00	24,00	21,00	26,00
22,00	28,00	25,00	22,00
27,00	24,00	26,00	25,00
23,00	22,00	21,00	20,00
20,00	21,00	27,00	26,00
26,00	28,00	24,00	25,00
22,00	25,00	25,00	25,00
29,00	23,00	22,00	29,00

Em seguida foi feito o ordenamento dos dados (no excel não necessariamente é necessário realizar)

Im35temp: Dados ordenados

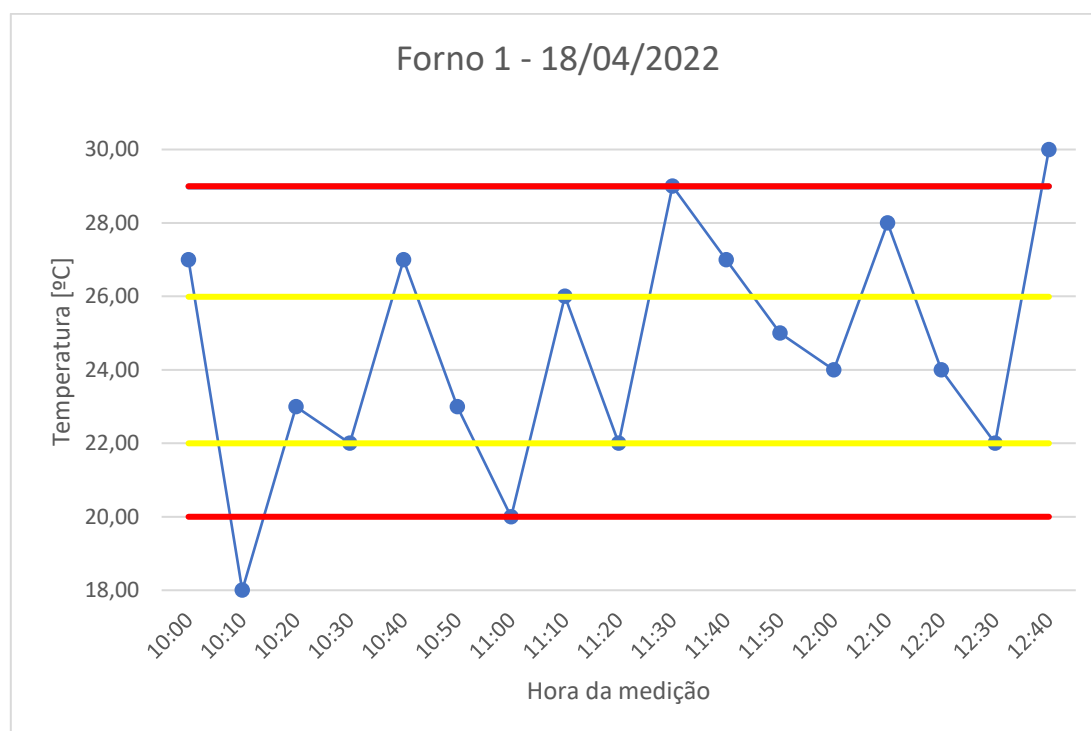
20,00	22,00	25,00	26,00
20,00	22,00	25,00	27,00
20,00	23,00	25,00	27,00
20,00	23,00	25,00	27,00
21,00	23,00	25,00	27,00
21,00	24,00	25,00	27,00
21,00	24,00	25,00	28,00
22,00	24,00	26,00	28,00
22,00	25,00	26,00	29,00
22,00	25,00	26,00	29,00

E com esses valores foram calculadas as posições de mínimo, máximo, 1º quartil, mediana e 3º quartil. Com isso podemos criar nossa faixa de alertas.

	Posição	Desc. Alerta	Faixa alerta	Cor alerta
Mínimo	20,00	Crítico frio	$\leq 20,00$	
1º quartil	22,00	Alerta frio	20,01 - 22,00	
Mediana	25,00	Ideal	22,01 - 25,99	
3º quartil	26,00	Alerta quente	26,00 - 28,99	
Máximo	29,00	Crítico quente	$\geq 29,00$	

Com essas faixas foram então feitas novas medidas em intervalos de 10min e adequadas os valores com relação às faixas dos alertas e montada a seguinte tabela para construção do gráfico.

	Medida	Crítico frio	Alerta frio	Ideal	Alerta quente	Crítico quente
10:00	27,00	20,00	22,00	25,99	28,99	29,00
10:10	18,00	20,00	22,00	25,99	28,99	29,00
10:20	23,00	20,00	22,00	25,99	28,99	29,00
10:30	22,00	20,00	22,00	25,99	28,99	29,00
10:40	27,00	20,00	22,00	25,99	28,99	29,00
10:50	23,00	20,00	22,00	25,99	28,99	29,00
11:00	20,00	20,00	22,00	25,99	28,99	29,00
11:10	26,00	20,00	22,00	25,99	28,99	29,00
11:20	22,00	20,00	22,00	25,99	28,99	29,00
11:30	29,00	20,00	22,00	25,99	28,99	29,00
11:40	27,00	20,00	22,00	25,99	28,99	29,00
11:50	25,00	20,00	22,00	25,99	28,99	29,00
12:00	24,00	20,00	22,00	25,99	28,99	29,00
12:10	28,00	20,00	22,00	25,99	28,99	29,00
12:20	24,00	20,00	22,00	25,99	28,99	29,00
12:30	22,00	20,00	22,00	25,99	28,99	29,00
12:40	30,00	20,00	22,00	25,99	28,99	29,00



Observe que as medidas entre as faixas amarelas (22°C e 26°C) se encontram na temperatura ideal. As que estão entre as faixas amarela e vermelha estão com algum alerta (quente ou frio) e as medidas abaixo ou acima das linhas vermelhas estão em estado crítico.

Tarefa para a aula

Cada grupo vai buscar os valores de referência do determinado projeto de PI para que em sala possamos calcular os valores dos quartis para confecção dos alertas!

Referências

[1] SPIEGEL, Murray R. Probabilidade e Estatística. São Paulo, McGraw-Hill do Brasil, p. 519. 1978

[2] https://www.youtube.com/watch?v=g8oqgrVhA_8&ab_channel=Azup Donald no País da Matemática e o Número de Ouro. Acessado em 18 de abril de 2022

[3] https://pt.made-in-china.com/co_rayscx/product_High-Quality-Fire-Extinguisher-Gauge_ryssyiong.html acessado em 18 de abril de 2022

[4] <https://www.sciencenewsforstudents.org/article/scientists-say-quartile> acessado em 18 de abril de 2022

[5] <https://www.sigmamagic.com/blogs/what-are-quartiles/> acessado em 18 de abril de 2022

[6] <https://support.microsoft.com/pt-br/office/fun%C3%A7%C3%A3o-med-d0916313-4753-414c-8537-ce85bdd967d2> acessado em 18 de abril de 2022

[7] <https://support.microsoft.com/pt-br/office/quartil-exc-fun%C3%A7%C3%A3o-quartil-exc-5a355b7a-840b-4a01-b0f1-f538c2864cad> acessado em 18 de abril de 2022

[8] <https://support.microsoft.com/pt-br/office/quartil-inc-fun%C3%A7%C3%A3o-quartil-inc-1bbacc80-5075-42f1-aed6-47d735c4819d> acessado em 18 de abril de 2022

[9] <https://support.microsoft.com/pt-br/office/m%C3%A9dia-fun%C3%A7%C3%A3o-m%C3%A9dia-047bac88-d466-426c-a32b-8f33eb960cf6> acessado em 18 de abril de 2022