

Lista 1 - Estatística

Gustavo Lopes Rodrigues

2021

Fórmulas usadas

Media aritmética

$$M_a = \frac{x_1 + x_2 + x_3 + \dots x_n}{n}$$

$$M_a = \frac{\sum F * X_i}{n}$$

Variância

$$V = \frac{\sum_{i=1}^n (x_i - M_a)^2}{n - 1}$$

Desvio Padrão

$$D_p = \sqrt{V}$$

$$D_v = \sqrt{\frac{\sum Fx^2 - n(M_a)^2}{n - 1}}$$

Coeficiente de Variação

$$C_v = \frac{D_p}{M_a} * 100$$

Mediana

$$M_e = li + \frac{(\frac{n}{2} - F_a A)}{f_i} * h$$

Respostas

1)

Media aritmética = **77.48**, Desvio Padrão = **8.17**, Coeficiente de variação **10.55**

Analisando os dados, percebe-se que a média aritmética é de 77.48, uma margem de erro de 8.17% além de um coeficiente de variação igual a 10.55%. Tais observações sugere que as cotações diárias das ações dessa empresa neste intervalo são homogêneas, pois não existe uma diferenciação muito grande entre os dados.

2)

Media aritmética = **56.6**, Mediana **57.5**, Desvio Padrão = **5.66**

Examinando a pontuação de cada trabalhador, percebe-se que a media aritmética das notas é igual a 56.6, a mediana é 57.5, e existe uma margem de erro entre os dados de 5.56. Os dados são homogêneos, já que a variação entre os números não é grande(demonstrado pelo desvio padrão). Além disso, percebe-se que quatro trabalhadores estão com notas abaixo da média, enquanto que o restante mantém acima, com destaque a duas notas. que superaram o desvio padrão.

3) Para estimar a nova média, podemos pegar a média(186), multiplicar pelo número de dados (30) e achar a soma total dos valores(5580). Depois disso, é só tirar 200 e dividir por 29(o número de dados menos o valor que foi retirado).

$$186 * 30 = 5580$$

$$M_a = \frac{(5580 - 200)}{29} = 185.51$$

Agora para calcular o novo desvio padrão, primeiro é preciso pegar o número atual(22), elevar ao quadrado para pegar a variância(484). E então multiplicar pelo número de elementos(30), então

$$484 * 30 = 14520$$

Por fim, agora é só pegar os dados atuais e fazer o cálculo do desvio padrão novamente, com os dados estipulados da média anterior.

$$14520 - (200 - 185.5172)^2 = 14310.2485$$

$$\frac{14310.2825}{29} = 493.4568$$

$$D_p = \sqrt{493.4568} = 22.2138$$

4)

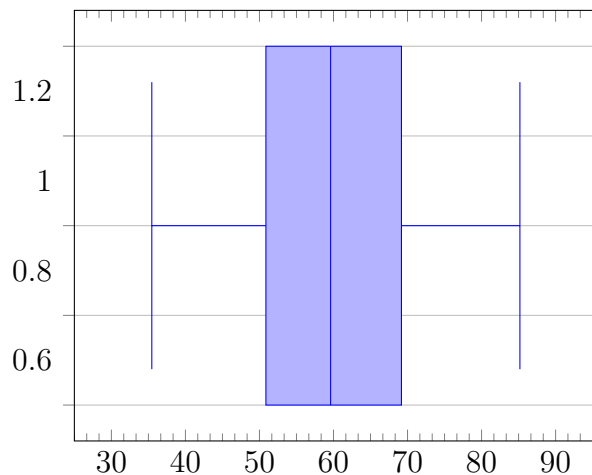
Observando os resultados: o objetivo da indústria foi não foi alcançada, pois eles conseguiram aumentar a média de 59.6% para 75.2%, um aumento de 26%, porém houve um aumento no desvio padrão, 9% indicando perca na homogeneidade

5)

- a) Média aritmética = **816.6**, Mediana **800**, Desvio Padrão = **212.48**
b) O mês com o valor mais alto está com um pouco menos de 2 desvios padrões.

6)

- a) Mediana = **59.61**
b) Q3 = **69.18**
c) Box Plot:



7)

- a) Por possuir um desvio padrão baixo(14.5%), e não ter nenhum valor muito alto, qualquer uma das medidas podem representar os dados

Ponto médio	60
Média aritmética	54.95
Mediana	56
Moda	40, 54, 60, 62 multimodal
Desvio Padrão	14.09%
Primeiro quartil	45.0
Terceiro quartil	64.28

- Um pouco mais da metade (55%) está com a nota acima da média
- Tem pelo menos 5 alunos que estão com notas preocupantes, pois estão muito abaixo da média
- Tem 2 alunos que se destacaram, um que teve um destaque ainda maior.

b) A melhor medida pode ser o quartil, ou a mediana, já que a moda é igual a 0, e sem contar que todos os outros valores são muito dispersos quando comparado ao maior número de faltas (45)

Ponto Médio	23
Media aritmética	4.25
Mediana	1.5
Moda	0
Desvio Padrão	9.86%
Primeiro quartil	0.25
Terceiro quartil	4.71

Análise: Percebe-se que em grande parte o número de faltas entre os trabalhadores são bem próximos. A única exceção é o maior número, que indica a presença de um "outlier", um número muito fora do padrão.

c) Neste caso, a melhor medida central pode ser a mediana ou a moda, pois muitas pessoas não leem o jornal.

Ponto Médio	6
Media aritmética	3.1
Mediana	0
Moda	0
Desvio Padrão	5.06%
Primeiro quartil	0
Terceiro quartil	7.14

Examinando o número de exemplares, percebe-se que este jornal não é bem vendido, já que a mediana, e o valor com maior frequência nos dados é 0.

d) Por se tratar de analisar pessoas com condição médica, o uso da Moda parece ser apropriado, assim também como a média, por conta dos dados terem uma dispersão baixa

Ponto Médio	2.01
Media aritmética	1.66
Mediana	1.62
Moda	1.6, 1.65 bimodal
Desvio Padrão	0.2%
Primeiro quartil	1.56
Terceiro quartil	1.68

Ao fazer o estudo, percebe-se que esse grupo de dados é bimodal(possui duas modas), mais específico, as alturas: 1.6,1.65, foram as que mais apareceram. A média encontrada foi de 1.65

Ainda como observação a mais, a altura 2.50 foi registrada, o que pode significar ser um possível "outlier", um possível número falso, pois está muito acima da dispersão.

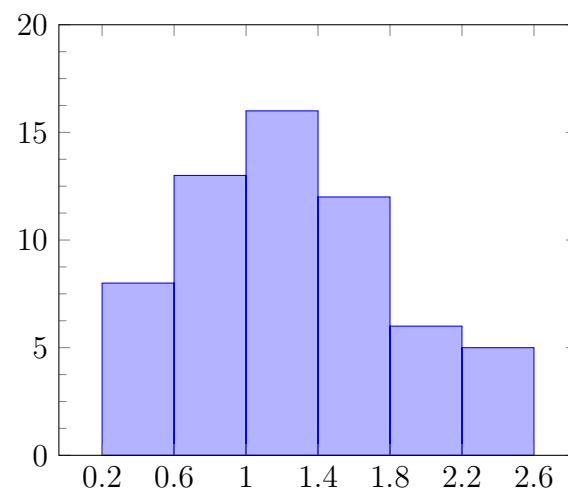
8)

Valor faturado em milhões de Reais	x_i	f	f%
0,2 —0,6	0,4	8	13,33%
0,6 —1,0	0,8	13	21,67%
1,0 —1,4	1,2	16	26,67%
1,4 —1,8	1,6	12	20,00%
1,8 —2,2	1,8	6	10,00%
2,2 —2,6	2,2	5	8,33%
Σ	8	60	100%

Media aritmética = 1.23

Desvio Padrão = 0.52

Histograma



9)

Media aritmética = **82.75**, Mediana **82**, Desvio Padrão = **11.84**

10)

Classe R\$	Frequência	F acumulada	Media	MxF	M^2xF
3 —4	3	3	3.5	10.5	36.75
4 —5	18	21	4.5	81	364.5
5 —6	13	34	5.5	71.5	393.25
6 —7	11	45	6.5	71.5	464.75
7 —8	15	60	7.5	112.5	843.75

a)

Mediana = 5.69

Media aritmética = 5.78

b)

Desvio Padrão = 1.28%

Variância = 1.63%

Coefficiente de variação = 22.07%

Análise: Analisando estes dados, percebe-se que são dados bem homogêneos, com uma margem de erro de 1.28 e um coeficiente de variação de 22.07%

11)

a) $Q1 = 2$, $Q2 = 4$, $Q3 = 5$

b) 15, o maior valor é um "outlier", pois para ser considerado um é preciso encontrar a faixa interquartil(FIQ), que é a diferença entre o terceiro e primeiro quartil. A partir dela, podemos observar o intervalo entre:

$$Q1 - 1.5FIQ \rightarrow 2 - 1.5(3) = -2.5 \quad (1)$$

$$Q3 + 1.5FIQ \rightarrow 5 + 1.5(3) = 9.5 \quad (2)$$

O número 15 como pode ser observado, é maior do que este intervalo.