

Métodos Quantitativos para a Computação

Estatística Descritiva
Média e Desvio Padrão

Medidas de tendência central

Média amostral

- Medida de tendência central é um resumo do valor central de um atributo em uma população
- A média amostral é a medida de tendência central mais importante
- Também chamada de média aritmética

Média amostral

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

- Na fórmula da média amostral, o somatório não indica explicitamente quais os valores que estão sendo somados, ficando entendido que se refere à soma dos valores de um atributo de todos os indivíduos (X_i) que fazem parte da questão

Exemplo 1

- Um gerente de supermercado deseja estudar a movimentação de pessoas em seu estabelecimento, e constata que

295 1002 941 768 1283

pessoas entraram no estabelecimento nos últimos 5 dias. Qual é o número médio de pessoas que entraram diariamente nestes 5 dias?

$$(295 + 1002 + 941 + 768 + 1283) / 5 = 857,8$$

Média de dados agrupados

- Para o cálculo da média obtém-se uma boa aproximação atribuindo a cada elemento de uma classe o valor do ponto médio correspondente
- Se uma distribuição tem k classes com pontos médios x_1, x_2, \dots, x_k , e frequências f_1, f_2, \dots, f_k , a média dos dados agrupados pode ser calculada pela seguinte expressão:

$$\bar{X} = \frac{\sum_{i=1}^k x_i \cdot f_i}{\sum_{i=1}^k f_i}$$

Exemplo 2

- Seja a distribuição de frequência para as seguintes quantidades (em toneladas) de óxidos de enxofre emitidas por uma indústria em 70 dias

$$\bar{X} = \frac{1310}{70} = 18,71$$

- A média usando-se os dados originais é 18,61

Toneladas de óxidos de enxofre	Ponto médio (x)	Frequência	$x \times f$
-5,0 – 10,0	7,5	5	37,5
-10,0 – 15,0	12,5	14	175
-15,0 – 20,0	17,5	21	367,5
-20,0 – 25,0	22,5	20	450
-25,0 – 30,0	27,5	9	247,5
-30 – 35,0	32,5	1	32,5
	Total	70	1310

Medidas de dispersão

Variabilidade

- O resumo de um conjunto de dados por uma única medida representativa de posição central esconde a informação sobre a variabilidade do conjunto de observações

Variância

- Um critério frequentemente usado para sumarizar a dispersão de um conjunto de dados é a sua variabilidade em torno da média
- A variância e o desvio padrão são calculados segundo este princípio

Desvio Padrão

- Como a variância é uma medida de dimensão igual ao quadrado da dimensão dos dados, define-se outra medida de dispersão denominada desvio padrão

Desvio padrão

- Desvio padrão amostral

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

Exemplo 3

- Em seis domingos consecutivos, um serviço de guincho recebeu 9, 7, 11, 10, 13, 7 chamadas
- Calcule a variância e o desvio padrão

$$\bar{X} = \frac{9+7+11+10+13+7}{6} = \frac{57}{6} = 9,5$$

$$S^2 = \frac{27,5}{6-1} = 5,5$$

$$S = \sqrt{5,5} = 2,3$$

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
9	-0,5	0,25
7	-2,5	6,25
11	1,5	2,25
10	0,5	0,25
13	3,5	12,25
7	-2,5	6,25
Total	0	27,5

Exemplo 4

- Calcular a variância e o desvio padrão para o seguinte conjunto de dados

$$\bar{X} = 65,1$$

$$s^2 = \frac{1289,81}{21-1} = 64,49$$

$$s = \sqrt{64,49} = 8,03$$

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
69	3,90	15,25
58	-7,10	50,34
70	4,90	24,06
66	0,90	0,82
46	-19,10	364,63
61	-4,10	16,77
65	-0,10	0,01
74	8,90	79,29
61	-4,10	16,77
75	9,90	98,10
80	14,90	222,15
55	-10,10	101,91
58	-7,10	50,34
67	1,90	3,63
56	-9,10	82,72
70	4,90	24,06
72	6,90	47,68
58	-7,10	50,34
70	4,90	24,06
68	2,90	8,44
68	2,90	8,44
Total	0,00	1289,81

Calculo da variância sem o cálculo prévio da média

- No procedimento anterior é necessário calcular a média antes de calcular a variância ou o desvio padrão
- A seguinte fórmula pode ser usada para se calcular a variância sem a necessidade de calcular a média

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2 / n}{n-1}$$

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2 / n}{n}$$

Exemplo 5

- Em seis domingos consecutivos, um serviço de guincho recebeu 9, 7, 11, 10, 13, 7 chamadas
- Calcule a variância e o desvio padrão
Calcule a variância e o desvio padrão sem calcular a média

$$S^2 = \frac{569 - (57)^2 / 6}{6 - 1} = 5,5$$

$$S = \sqrt{5,5} = 2,3$$

X_i	X_i^2
9	81
7	49
11	121
10	100
13	169
7	49
57	569

Exemplo 6

- Calcule a variância e o desvio padrão dos dados do exemplo 8 sem calcular a média

$$S^2 = \frac{90275 - (1367)^2 / 21}{21 - 1}$$

$$S^2 = \frac{90275 - 88985,19}{20} = \frac{1289,81}{20} = 64,49$$

$$S = \sqrt{64,49} = 8,03$$

X_i	$(X_i)^2$
69	4761,00
58	3364,00
70	4900,00
66	4356,00
46	2116,00
61	3721,00
65	4225,00
74	5476,00
61	3721,00
75	5625,00
80	6400,00
55	3025,00
58	3364,00
67	4489,00
56	3136,00
70	4900,00
72	5184,00
58	3364,00
70	4900,00
68	4624,00
68	4624,00
1367,00	90275,00

Variância de dados agrupados

- Para determinar a variância e o desvio padrão de uma distribuição, determina-se os pontos médios (x_1, x_2, \dots, x_n) e as frequências (f_1, f_2, \dots, f_n) das classes correspondentes

$$S^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i - \left(\sum_{i=1}^n x_i \cdot f_i \right)^2 / n}{n-1}$$

Exemplo 7

- Ache a média, a variância e o desvio padrão da seguinte distribuição que dá os tempos dedicados ao lazer de uma amostra de 80 estudantes em uma semana escolar

Horas	Frequência
-10 - 15	8
-15 - 20	28
-20 - 25	27
-25 - 30	12
-30 - 35	4
-35 - 40	1

Exemplo 7

Ponto médio (x)	x^2	f	$x \cdot f$	$x^2 \cdot f$
12,5	156,25	8	100	1250
17,5	306,25	28	490	8575
22,5	506,25	27	607,5	13668,75
27,5	756,25	12	330	9075
32,5	1056,25	4	130	4225
37,5	1406,25	1	37,5	1406,25
Total	4187,50	80	1695	38200

$$\bar{X} = \frac{1695}{80} = 21,19 \quad S^2 = \frac{38200 - (1695^2 / 80)}{80 - 1} = \frac{2287,2}{79} = 28,95$$

$$S = \sqrt{28,95} = 5,38$$

Exemplo 8

- Calcular a o desvio padrão da seguinte distribuição de freqüência para as seguintes quantidades (em toneladas) de óxidos de enxofre emitidas por uma indústria em 70 dias

Toneladas de óxidos de enxofre	Freqüência
-5,0 – 10,0	5
-10,0 – 15,0	14
-15,0 – 20,0	21
-20,0 – 25,0	20
-25,0 – 30,0	9
-30,0 – 35,0	1
	70

Exemplo 8

$$\bar{X} = \frac{1310}{70} = 18,71$$

$$S^2 = \frac{26887,75 - (1310^2 / 70)}{70 - 1}$$

$$S^2 = \frac{2371,79}{69} = 34,37$$

$$S = \sqrt{34,37} = 5,86$$

Toneladas de óxidos de enxofre	Ponto médio (x)	x^2	Freq	$x \cdot f$	$x^2 \cdot f$
-5,0 – 10,0	7,5	56,25	5	37,5	281,25
-10,0 – 15,0	12,5	156,25	14	175	2187,5
-15,0 – 20,0	17,5	306,25	21	367,5	6431,25
-20,0 – 25,0	22,5	506,25	20	450	10125,0
-25,0 – 30,0	27,5	756,25	9	247,5	6806,25
-30,0 – 35,0	32,5	1056,25	1	32,5	1056,25
		Total	70	1310	26887,5

Exemplo 8

- Os dados relativos ao exemplo de emissão de óxido de carbono

15,8	26,4	17,3	11,2	23,9	24,8	18,7	13,9	9	13,2
22,7	9,8	6,2	14,7	17,5	26,1	12,8	28,6	17,6	23,7
26,8	22,7	18	20,5	11	20,9	15,5	19,4	16,7	10,7
19,1	15,2	22,9	26,6	20,4	21,4	19,2	21,6	16,9	19
18,5	23	24,6	20,1	16,2	18	7,7	13,5	23,5	14,5
8,3	21,9	12,3	22,3	13,3	11,8	19,3	20	25,7	31,8
25,9	10,5	15,9	27,5	18,1	17,9	9,4	24,1	20,1	28,5

$$S^2 = 33,39 \quad S = \sqrt{33,39} = 5,78$$

Métodos Quantitativos para a Computação

Estatística Descritiva
Média e Desvio Padrão