

Distribuição de Frequências

Uma distribuição de frequências é um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas. Para construir uma distribuição de frequências, temos de dividir a faixa de dados em intervalos, que são geralmente chamados de intervalos de classe ou células. Se possível, os intervalos devem ser de iguais larguras de modo a aumentar a informação visual na distribuição de frequências. Algum julgamento tem de ser usado na seleção do número de intervalos de classes, de modo que uma apresentação razoável possa ser desenvolvida. O número de intervalos depende do número de observações e da quantidade de espalhamento ou dispersão dos dados. Uma distribuição de frequências não será informativa se usar um número muito baixo ou muito alto de intervalos de classe. Geralmente, achamos que 5 a 20 intervalos são satisfatórios na maioria dos casos e que o número de intervalos deve crescer com n . Na prática, trabalha-se bem se o número de intervalos de classe for aproximadamente igual à raiz quadrada do número de observações. Sendo assim, considere um banco de dados com valores x_1, x_2, \dots, x_n e os seguintes postulados para construir uma tabela de frequências:

1. Determine o número de termos (n), o menor valor dos dados $x_{(1)}$ e o maior valor dos dados $x_{(n)}$.
2. Determine a amplitude total (AT) dada por:

$$AT = x_{(n)} - x_{(1)}$$

3. Determine o número de classes n_c que pode ser calculada como:

$$n_c \approx \sqrt{n} \quad \text{ou} \quad n_c \approx 1 + 3,322 \log n$$

Obs: \approx representa o inteiro mais próximo.

4. Determine a amplitude das classes (c) que é definida como:

$$c = \frac{AT}{(n_c - 1)}$$

5. Defina $Li_1 = x_{(1)} - \frac{c}{2}$, $Li_k = Li_{k-1} + c$, $1 \leq k \leq n_c$ e $Li_k = Li_{(k-1)}$, $2 \leq k \leq n_c$. Em que Li_j representa o limite inferior da classe j e Li_j representa o limite superior da classe j .
6. A frequência absoluta (f_a) de uma classe j é encontrada contabilizando os valores x_1, x_2, \dots, x_n que pertencem ao intervalo $[Li_j, Li_j)$.

Exemplo: Considere o banco de dados seguinte e elabore uma tabela de distribuição de frequências seguindo os passos acima.

16	35	26	37	50	38	27	38	20	23	30	12	50	46	40	17
17	10	10	29	36	32	26	37	31	24	19	29	24	30	40	35
18	11	14	26	36	30	36	14	39	10	35	17	10	43	43	34
20	12	16	14	38	15	18	14	44	39	34	30	40	22	39	15
22	13	19	40	14	50	13	17	15	11	40	47	13	15	36	20
26	13	23	33	42	25	43	26	42	29	25	45	28	31	28	25
28	15	24	34	38	34	16	48	14	34	26	26	41	39	12	11
29	20	24	17	12	30	40	24	42	25	25	41	33	23	43	48
29	20	28	17	11	40	46	31	35	43	44	22	13	38	44	49
30	26	30	40	42	50	16	28	43	21	29	23	29	20	14	11
30	32	31	22	27	20	23	45	19	23	17	10	10	30	14	32
32	41	37	30	21	25	47	38	22	49	32	48	47	35	37	29
35	43	40	38	40	25	43	18	32	12	36	21	11	19	24	21
36	44	44	41	33	26	37	34	46	47	39	27	32	50	40	32
46	49	47	41	45	44	26	44	13	44	23	28	29	33	16	41

Seguindo os postulados temos: $n = 240$, $x_{(1)} = 10$, $x_{(n)} = 50$, $AT = 40$, $n_c = \sqrt{240} \approx 15$, $c = 2,857$, $Li_1 = 8,572$ e $Li_{15} = 11,4285$. Assim, temos a seguinte distribuição de frequências:

Tabela 1: Distribuição de Frequências

Classes	f_a	f_r	f_{acm}	fr_{acm}
[8,572; 11,4285)	12	0,05	12	0,05
[11,4285;14,2855)	19	0,079167	31	0,129167
[14,2855;17,1425)	17	0,070833	48	0,2
[17,1425;19,9995)	7	0,029167	55	0,229167
[19,9995;22,8565)	16	0,066667	71	0,295833
[22,8565;25,7135)	20	0,083333	91	0,379167
[25,7135;28,5705)	19	0,079167	110	0,458333
[28,5705;31,4275)	23	0,095833	133	0,554167
[31,4275;34,2845)	18	0,075	151	0,629167
[34,2845;37,1415)	17	0,070833	168	0,7
[37,1415;39,9985)	12	0,05	180	0,75
[39,9985;42,8555)	21	0,0875	201	0,8375
[42,8555;45,7125)	19	0,079167	220	0,916667
[45,7125;48,5695)	12	0,05	232	0,966667
[48,5695;51,4265)	8	0,033333	240	1
Total	240	1	-	-

f_a é a frequência absoluta. f_r é a frequência relativa. f_{acm} representa a frequência absoluta acumulada e fr_{acm} representa a frequência relativa acumulada.

Uma distribuição de frequência é um método de se agrupar dados em classes de modo a fornecer a quantidade (e/ou a percentagem) de dados em cada classe. O método considerado neste texto será o padrão nas aulas de MAF 105 e deve ser seguido na resolução de problemas da disciplina.