

Estatística Descritiva

Julienne Borges

Distribuição de frequências

Um dos primeiros passos para analisar um arquivo de dados, especialmente quando o número de observações for grande, é a distribuição de frequências de cada variável.

- Uma distribuição de frequências é uma tabela que mostra categorias, valores ou intervalos de valores de acordo com as ocorrências.

Elementos de uma tabela

The diagram illustrates the components of a table. Labels with arrows point to specific parts: 'Título' points to the main title 'Tabela 11 - Evolução da renda dos Empreendedores Brasileiros'; 'Cabeçalho' points to the header row containing the categories and percentages; 'Casa ou Célula' points to a specific data cell containing the value '43'; 'Corpo' points to the main body of the table containing the data rows; 'Coluna Indicadora' points to the first column which lists the income ranges; and 'Fonte' points to the source information 'Fonte: Pesquisas GEM 2000 - 2003'.

Tabela 11 – Evolução da renda dos Empreendedores Brasileiros				
Faixa de Renda	2000(%)	2000(%)	2000(%)	2000(%)
Menos de 3 SM	30	39	43	53
De 3 a 6 SM	30	31	34	22
Mais de 6 a 9 SM	14	12	11	8
Mais de 9 a 15 SM	12	12	9	6
Mais de 15 SM	1	3	3	1
Não sabe Recusou	4	6	2	2
Total	100	100	100	100
Fonte: Pesquisas GEM 2000 – 2003				

Distribuição de frequências

Tabela 1 – Distribuição de frequências do provedor usado pelo visitante do site.

Provedor	Frequência	Porcentagem
A	10	25,0
B	17	42,5
C	7	17,5
D	6	15,0
Total	40	100,0

Tabela 2 – Distribuição de frequências do número de defeitos encontrados em escadas no final da linha de produção.

Nº de defeitos encontrados	Frequência	Porcentagem	Frequência acumulada	Porcentagem acumulada
0	13	27,1	13	27,1
1	15	31,3	28	58,4
2	10	20,8	38	79,2
3	7	14,6	45	93,8
4	2	4,2	47	98
7	1	2,1	48	100
Total	48	100,0		

Exemplo 1

Os dados, a seguir, representam o tempo (em segundos) que operadores gastam para montar um equipamento na linha de produção de uma empresa:

4,7	4,9	5,1	5,4	5,7	6	6,3	6,8	7,3	8,9
4,8	4,9	5,2	5,5	5,7	6,2	6,4	6,9	8,2	9,1
4,8	5	5,3	5,6	5,7	6,2	6,5	7	8,2	9,9
4,9	5	5,4	5,6	5,9	6,2	6,7	7,1	8,3	14,1
4,9	5	5,4	5,7	6	6,3	6,8	7,3	8,4	15,2

Distribuição de frequências por intervalos

A Regra de Sturges é uma das regras mais utilizadas na Estatística para construção de uma tabela de frequências por intervalos. Isso porque a fórmula de Sturges nos fornece uma quantidade adequada de classes para os mais variados tamanhos de amostras.

Distribuição de frequências por intervalos

A regra de Sturges envolve os seguintes passos:

Cálculo da amplitude: $A = \text{máx} - \text{mín}$

Cálculo do número de *intervalos* ou classes da tabela:

$$k = 1 + 3,322 \cdot \log n$$

Cálculo da amplitude dos *intervalos*: $A_k = \frac{A}{k}$

Exemplo 1

Os dados, a seguir, representam o tempo (em segundos) que operadores gastam para montar um equipamento na linha de produção de uma empresa:

4,7	4,9	5,1	5,4	5,7	6	6,3	6,8	7,3	8,9
4,8	4,9	5,2	5,5	5,7	6,2	6,4	6,9	8,2	9,1
4,8	5	5,3	5,6	5,7	6,2	6,5	7	8,2	9,9
4,9	5	5,4	5,6	5,9	6,2	6,7	7,1	8,3	14,1
4,9	5	5,4	5,7	6	6,3	6,8	7,3	8,4	15,2

Exemplo 1

Utilizando os passos da regra de Sturges, temos:

$$A = 15,2 - 4,7 = 10,5$$

$$K = 1 + 3,322 \cdot \log 50 = 1 + 3,322 \cdot 1,699 = 6,644 \text{ (pode-se escolher entre 6 ou 7 classes)}$$

$A_k = 10,5 / 6 = 1,75$ (nesse caso, escolhendo 6 classes, cada classe deverá ter uma amplitude de 1,8 segundos, seguimos o mesmo número de casas decimais dos dados da amostra)

Tabela 3 - Distribuição de frequências do tempo gasto (em segundos) para a montagem de um equipamento na linha de produção.

Tempo	Frequência	Porcentagem	Frequência acumulada	Porcentagem acumulada
4,7 ┤ 6,5	32	64,0	32	64,0
6,5 ┤ 8,3	11	22,0	43	86,0
8,3 ┤ 10,1	5	10,0	48	96,0
10,1 ┤ 11,9	0	0,0	48	96,0
11,9 ┤ 13,7	0	0,0	48	96,0
13,7 ┤ 15,5	2	4,0	50	100,0
Total	50	100,0	-	-

Simbologia utilizada

Observe que o símbolo $|$ — é utilizado para representar intervalo aberto/fechado, ou seja, o valor que se encontra do lado direito do símbolo não está incluído no intervalo (intervalo aberto) e o valor que se encontra do lado esquerdo do intervalo está incluído no intervalo (intervalo fechado).

Tabelas de contingência

As tabelas de contingência são a forma usual de apresentar uma distribuição de frequência conjunta de duas ou mais variáveis.

Exemplo 2

Tabela 4 – Distribuição de frequências do número de defeitos de acordo com a posição e o lado do pneu em veículos utilitários.

Lado	Posição		Total	
	Dianteira	Traseira		
Direito	32	28	60	Totais das linhas →
Esquerdo	35	57	92	
Totais das colunas ← Total	67	85	152	Total geral →

Exemplo 2

Tabela 5 – Distribuição de defeitos encontrados em veículos utilitários de acordo com a posição e o lado do pneu.

Lado	Posição		Total
	Dianteira	Traseira	
Direito	53%	47%	100%
Esquerdo	38%	62%	100%
Total	44%	56%	100%

Exemplo 2

Tabela 6 – Distribuição de defeitos encontrados em veículos utilitários de acordo com a posição e o lado do pneu.

Lado	Posição		Total
	Dianteira	Traseira	
Direito	48%	33%	39%
Esquerdo	52%	67%	61%
Total	100%	100%	100%

Exemplo 2

Tabela 7 – Distribuição de defeitos encontrados em veículos utilitários de acordo com a posição e o lado do pneu.

Lado	Posição		Total
	Dianteira	Traseira	
Direito	21%	18%	39%
Esquerdo	23%	38%	61%
Total	44%	56%	100%

Exemplo 3

Quantidade de acidentes de trabalho 2007-2009,
segundo a Classificação Nacional de Atividades Econômicas (CNAE)

	Com CAT			Sem CAT			Total		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
Bancos comerciais	737	785	671	136	205	211	873	990	882
Bancos múltiplos	3.783	3.875	3.319	1.395	2.239	2.484	5.178	6.114	5.803
Caixas econômicas	889	693	620	207	357	412	1.096	1.050	1.032

Fonte: Ministério da Previdência Social, com dados do INSS

*CAT: Comunicação de acidente de trabalho

Exemplo 4

Estatística	Matemática			Totais
	0 ┤ 4	4 ┤ 7	7 ┤ 10	
0 ┤ 4	32	25	5	62
4 ┤ 7	20	183	82	285
7 ┤ 10	7	27	19	53
Totais	59	235	106	400



PUC Minas
Virtual