

Distribuição de Frequência

ML205 - Estatística I

João Luis Gomes Moreira

DMAT/CCT/UFRR

4 de março de 2021

Conteúdo

- 1 Distribuição de Frequência
 - Dos Dados às Tabelas
 - Terminologia
 - Gráficos

Situação Problema

“A direção de uma empresa está estudando a possibilidade de fazer um seguro saúde para seus funcionários e respectivos familiares. Para isso, ela faz um levantamento junto a seus 500 funcionários, obtendo informações sobre sexo, estado civil, idade, número de dependentes e salário.”

Tomemos inicialmente a *variável qualitativa* sexo.

Resumindo os Dados

Como são 500 funcionários, temos que achar uma forma de resumir os dados. Uma forma é através de uma tabela e outra é através de um gráfico.

Contudo, plotar um gráfico neste momento não seja muito fácil. Assim, trabalhamos os dados em uma tabela e por fim plotamos o gráfico.

Além disso, o que nos interessa não é o sexo de cada funcionário, mas quantos funcionários são do sexo masculino e quantos do sexo feminino. Ou seja, queremos saber a frequência de funcionários masculinos e femininos. Neste exemplo, masculino e feminino são as classes, ou categorias, da variável sexo.

Frequência Absoluta

A tabela abaixo é conhecida como Distribuição de Frequência:

Sexo	Número de Funcionários
Masculino	270
Feminino	230
Total	500

A tabela nos informa que dos 500 funcionários o sexo Masculino aparece 270 vezes, ou com uma frequência de 270 vezes.

Esta frequência também é dita simples ou absoluta.

Frequência Relativa

É interessante também expressar os dados das frequências em relação ao total, ao que chamamos de frequência relativa. A tabela abaixo mostra esta representação.

Sexo	Frequência Simples	
	Absoluta (n)	Relativa (%)
Masculino	270	0,54 ou 54%
Feminino	230	0,46 ou 46%
Total	500	1,00 ou 100%

Note que a soma das frequências relativas é sempre 1, enquanto a soma das frequências absolutas deve ser igual ao número total de elementos sendo investigados.

Note também que $0,54 \times 100 = 54\%$, ou seja, a notação da frequência relativa pode ser visualizada facilmente como uma porcentagem.

Exemplo de Variável Discreta

Também nomeada como Distribuição de Frequência do Tipo A.

X = número de erros por página observados em um livro

Número de Erros (X_i)	Número de Páginas (F_i)
0	35
1	20
2	13
3	6
4	4
5	2

Onde:

X_i = identifica as classes, ou categorias, em que o fato se subdivide

F_i = informa a frequência absoluta de cada classe

N = informa a soma dos elementos observados na população

n = informa a soma dos elementos observados na amostra

Obs: Perceba que para este caso $n = N = 80$.

Exemplo de Variável Contínua.

Também nomeada como Distribuição de Frequência do Tipo B.

X = notas finais de 50 estudantes

Sejam as notas dos estudantes :

22	46	9	40	57	22	22	13	50	42
35	2	15	41	34	52	32	75	69	44
26	42	60	56	30	3	17	79	45	37
0	12	62	50	45	41	59	11	66	39
43	33	70	50	47	20	36	40	67	29

Exemplo de Variável Contínua

Com as notas geramos a seguinte distribuição de frequência :

Notas (X_i)	Número de Estudantes (F_i)
0 ┤ 10	4
10 ┤ 20	5
20 ┤ 30	6
30 ┤ 40	8
40 ┤ 50	12
50 ┤ 60	7
60 ┤ 70	5
70 ┤ 80	3

Terminologia.....

Dados Brutos: são os dados originais, que ainda não se encontram prontos para análise, por não estarem numericamente organizados.

Também são conhecidos como Tabela Primitiva.

Por exemplo, as 50 notas dos alunos no exemplo anterior de distribuição tipo B.

22	46	9	40	57	22	22	13	50	42
35	2	15	41	34	52	32	75	69	44
26	42	60	56	30	3	17	79	45	37
0	12	62	50	45	41	59	11	66	39
43	33	70	50	47	20	36	40	67	29

Terminologia.....

Rol: é o arranjo dos dados brutos em ordem crescente ou decrescente de grandeza.

O Rol em ordem crescente dos dados brutos do exemplo anterior é :

0	2	3	9	11	12	13	15	17	20
22	22	22	26	29	30	32	33	34	35
36	37	39	40	40	41	41	42	42	43
44	45	45	46	47	50	50	50	52	56
57	59	60	62	66	67	69	70	75	79

Terminologia....

Dados discretos: dados de uma variável discreta, que assume valores em pontos da reta real.

Exemplos:

número de erros em um livro: $0, 1, 2, 3, 4, \dots$

número de filhos de vários casais: $0, 1, 2, 3, 4, \dots$

número de acidentes em uma rodovia: $4, 10, 12, 15, \dots$

Dados contínuos: dados de uma variável contínua, que assume qualquer valor em certo intervalo da reta real.

Exemplos:

peso de alunos: $55, 5\text{kg}, 61, 0\text{kg}, 63, 4\text{kg}, 68, 1\text{kg}, \dots$

distância entre cidades: $35, 5\text{km}, 48, 6\text{km}, 100, 10\text{km}, \dots$

Terminologia...

Dados Tabelados não agrupados em classes: os valores da variável aparecem individualmente. São representados por uma distribuição de frequência tipo A.

Dados Tabelados agrupados em classes: os valores da variável não aparecem individualmente, mas agrupados em classes. São representados por uma distribuição de frequência tipo B.

Limites de Classe: são os extremos de cada classe, existindo um limite inferior e um superior.

Amplitude Total ou Range: É a diferença entre o maior e o menor dado da distribuição ($79 - 0 = 79$).

Terminologia..

Ponto Médio das Classes (X_i): é a média aritmética entre o limite superior e o limite inferior da classe.

Por exemplo, se a classe for $0 \vdash 10$, teremos:

$$\text{Ponto Médio da Classe} = \frac{0 + 10}{2} = 5$$

Número de Classes (K): Número de classes da distribuição de frequência. Não existe uma regra fixa para se determinar o número de classes, sendo o número de classes calculado em função do número n de valores da variável ou de itens da amostra. São geralmente utilizados os procedimentos:

- 1 A Regra de Sturges: $K = 1 + 3,3 \cdot \log(n)$
- 2 $K = 5$ para $n \leq 25$ e $K \cong \sqrt{n}$, para $n > 25$

Terminologia.

Exemplo: Seja ainda o exemplo com os 50 alunos ($n = 50$):

Pela formula de Sturges:

$$K = 1 + 3,3.\log(50) = 6,61 \rightarrow K = 7$$

Adotando $K \approx \sqrt{n}$, temos

$$K = \sqrt{50} = 7,07 \rightarrow K = 7$$

Observe-se que na geração da distribuição de frequência tipo B desses dados, foram utilizadas 8 classes com amplitude de classe de 10.

Ou seja, o cálculo do número de classes de uma distribuição de frequência não define obrigatoriamente o número de suas classes.

Terminologia

Amplitude ou Intervalo de Classes (h):

$$h = \text{Amplitude Total} / \text{Número de Classes } (K)$$

Notações: $0 - 10$, $0 \vdash 10$, $0 \dashv 10$ e $0 \vdash\vdash 10$
(sendo $h = 10 - 0 = 10$)

Frequência Relativa da Classe: corresponde ao quociente entre a frequência absoluta da classe e o total de elementos

Frequência Acumulada Crescente: corresponde à soma da frequência de uma classe com as frequências das classes anteriores

Frequência Acumulada Decrescente: corresponde à soma da frequência de uma classe com as frequências das classes posteriores

Gráficos

Além dos gráficos já vistos, podem ser citados como gráficos representativos:

- Histograma: que associa as classes com suas frequências.
- Polígono de Frequência: também associa as classes com suas frequências, sendo traçado ligando-se os pontos médios de cada barra de um histograma.
- Polígono de Frequência Acumulada: que associa os limites das classes com suas frequências acumuladas.
- Função Distribuição Acumulada - FDA: associa dados (suas classes e seus limites) com a percentagem total dos dados.

That's all, Folks !!!