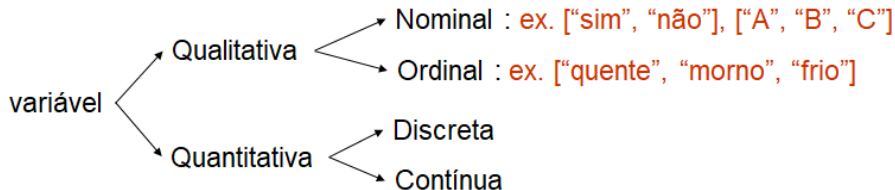


# Estatística e Probabilidade

Carolina Silva Pena

# Tipos de variáveis



# Exemplo Base de Dados

Id	Turma	Sexo	Idade	Horas_estudo	ENEM_MAT	Notas_EST
1	B	M	21	Até 1h	346,82	2,79
2	B	F	22	Até 1h	634,38	4,30
3	B	M	21	Mais de 1h e até 3h	561,03	4,46
4	B	F	22	Mais de 3h e até 5h	602,91	7,64
5	A	F	21	Mais de 3h e até 5h	751,11	7,70
6	B	F	22	Até 1h	441,68	2,09
7	A	M	24	Até 1h	571,72	4,94
8	A	M	23	Mais de 1h e até 3h	606,97	5,78
9	A	M	21	Mais de 5h	731,62	8,33
10	B	F	18	Mais de 3h e até 5h	709,59	7,45
11	B	M	21	Até 1h	502,09	5,28
12	B	F	20	Mais de 5h	779,17	10,00
13	B	M	23	Mais de 5h	640,57	7,80
14	A	F	17	Mais de 1h e até 3h	564,35	5,56
15	B	F	20	Até 1h	504,76	4,15

# Distribuição de frequência

Sexo	Frequência ( $n_i$ )	Proporção ( $f_i$ )	Percentual ( $100 \times f_i$ )
F	8	0,5333	53,33%
M	7	0,4667	46,67%
Total	15	1	100%

em que:

- $n$  representa o número total de observações;
- $n_i$  representa a frequência absoluta de cada classe;
- $f_i = n_i/n$  indica a proporção (ou frequência relativa) de cada classe.

# Construção de tabelas de frequência para variáveis contínuas

- A solução empregada é agrupar os dados por faixa.

Nota ENEM	Frequência ( $n_i$ )	Proporção ( $f_i$ )	Percentual ( $100 \times f_i$ )
300 $\vdash$ 400	1	0,0667	6,67%
400 $\vdash$ 500	1	0,0667	6,67%
500 $\vdash$ 600	5	0,3333	33,33%
600 $\vdash$ 700	4	0,2667	26,67%
700 $\vdash$ 800	4	0,2667	26,67%
Total	15	1	100%

- em que a notação  $a \vdash b$  equivale à notação  $[a; b)$  e indica que o intervalo contém o extremo  $a$ , mas não contém o extremo  $b$ .

# Gráficos de barras (Frequência)

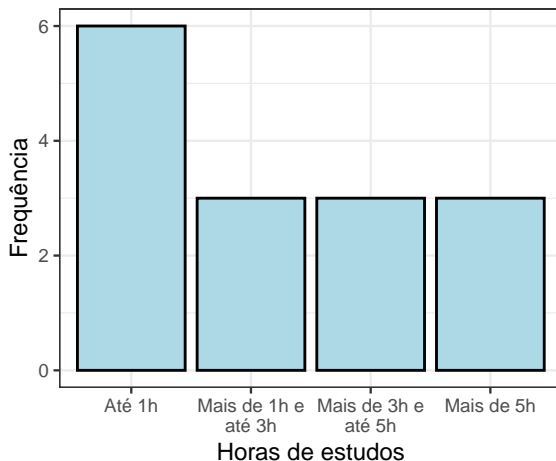


Figura 1: Gráfico de barras da variável Horas de Estudo

# Gráfico de barras (Percentual)

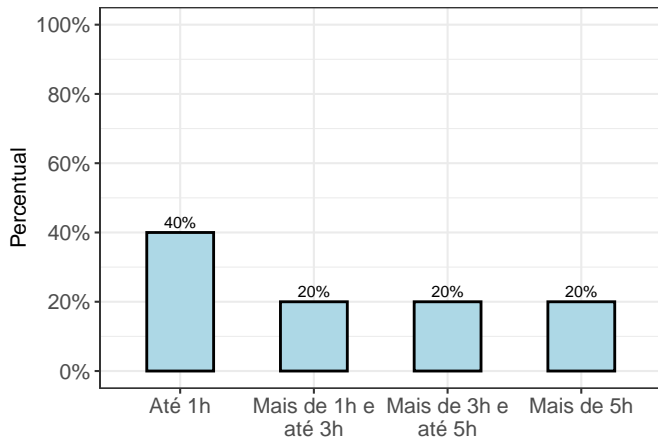


Figura 2: Gráfico de barras da variável Horas de Estudo (versão 2)

# Gráfico de Pizza

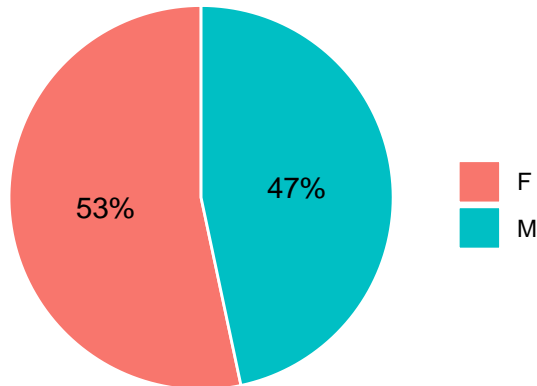


Figura 3: Gráfico de pizza da variável Sexo



# Gráfico de Barras (variáveis discretas)

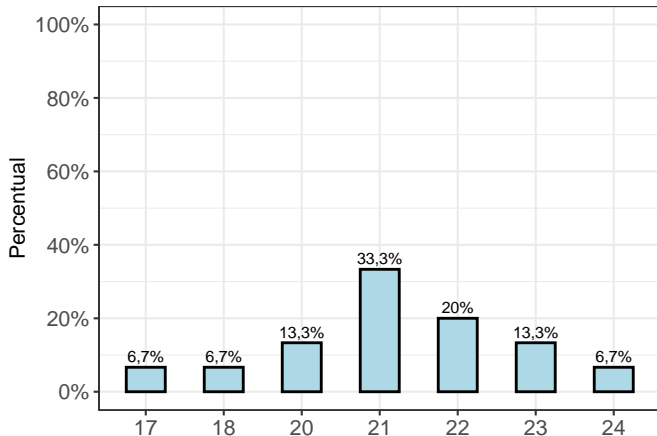


Figura 4: Gráfico de barras da variável Idade

# Histograma

O histograma é um gráfico de barras contínuas.

- O eixo X representa o intervalo de classe;
- O eixo Y pode representar a frequência ou a Densidade.

$$Densidade = \frac{Frequência\ relativa}{Largura\ do\ intervalo}$$

Quando o eixo Y é representado pela Densidade:

- a área total do histograma será igual a 1;
- a área de cada intervalo será igual a sua frequência relativa.

# Histogramas

