

# **Machine Learning com Python**

Prof. Luciano Galdino

# Distribuições de probabilidades

Variáveis aleatórias: são variáveis quantitativas cujos valores estão associados a fatores aleatórios.

1) Variável aleatória discreta: número contável de possíveis resultados ou atributos (Ex.: Número de pessoas em uma sala, número de gols em uma partida, número de filhos...).

2) Variável aleatória contínua: número incontável de possíveis resultados ou atributos (Ex.: intervalo de tempo, temperatura, altura...).

# Distribuições de probabilidades discretas

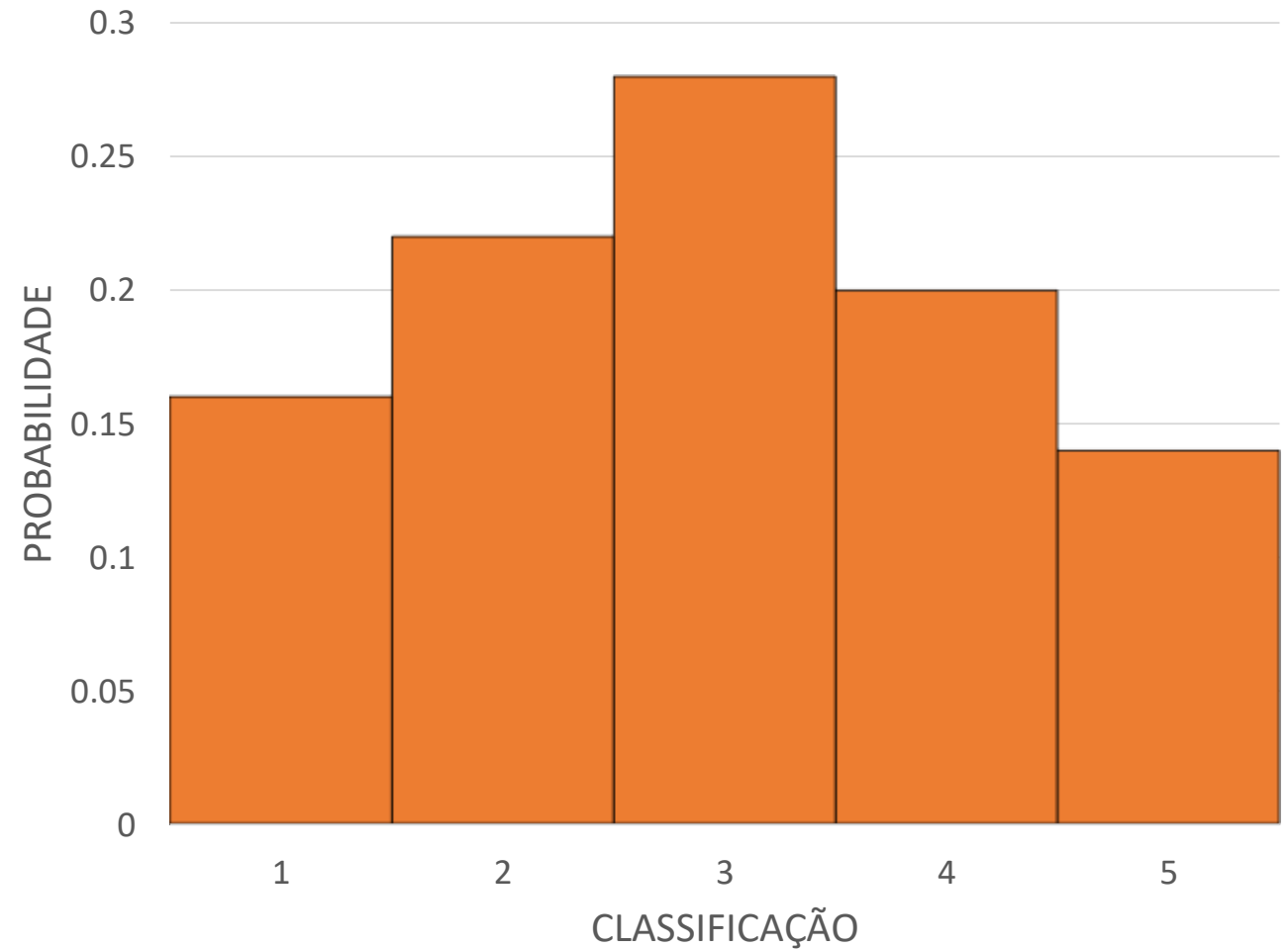
Lista cada valor possível que a variável aleatória pode assumir, assim como a sua probabilidade.

Condições:

- 1) Probabilidade cada valor é entre 0 e 1:  $0 \leq P(x) \leq 1$ .
- 2) A soma das probabilidades deve ser igual a 1:  $\sum P(x) = 1$

## Representação das distribuições de probabilidades discretas

Classificação (x)	Frequência (f)	P(x)
1	24	0,16
2	33	0,22
3	42	0,28
4	30	0,20
5	21	0,14
Soma	150	1



# Média de variáveis discretas

Representa a média teórica de um experimento de probabilidade.

$$\mu = \sum x \cdot P(x)$$

Classificação (x)	Frequência (f)	P(x)	x.P(x)
1	24	0,16	1.0,16
2	33	0,22	2.0,22
3	42	0,28	3.0,28
4	30	0,20	4.0,20
5	21	0,14	5.0,14
Soma	150	1	<b>μ=2,94</b>

Valor esperado de uma variável aleatória é igual a média e pode ser negativa.

$$E(x) = \mu$$

**Variância** de variáveis discretas:  $\sigma^2 = \sum (x - \mu)^2 \cdot P(x)$

**Desvio padrão** de variáveis discretas:  $\sigma = \sqrt{\sigma^2}$

Classificação (x)	Frequência (f)	P(x)	x.P(x)	(x - $\mu$ ) <sup>2</sup>	P(x). (x - $\mu$ ) <sup>2</sup>
1	24	0,16	1.0,16	3,764	0,602
2	33	0,22	2.0,22	0,884	0,194
3	42	0,28	3.0,28	0,004	0,001
4	30	0,20	4.0,20	1,124	0,225
5	21	0,14	5.0,14	4,244	0,594
Soma	150	1	$\mu=2,94$		$\sigma^2 = 1,616$

$$\sigma = \sqrt{1,616} = 1,27$$

# Distribuição Binomial

Possibilidade de apenas dois resultados.

Tentativas idênticas e independentes.

$$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Fatorial:

$$4! = 4.3.2.1 = 24$$

$$5! = 5.4.3.2.1 = 120$$

n = número de tentativas ou amostras.

p = probabilidade de sucesso em uma tentativa.

q = probabilidade de fracasso em uma tentativa.

x = variável aleatória que representa a contagem do número de sucessos.

# Distribuição Geométrica

Probabilidade de repetidas tentativas até que o sucesso ocorra.

Tentativas independentes.

A probabilidade de sucesso é constante para todas as tentativas.

$$P(x) = p(q)^{x-1}$$

p = probabilidade de sucesso em uma tentativa.

q = probabilidade de fracasso em uma tentativa.

x = variável aleatória que representa o número de falhas até o sucesso.



# Distribuição de Poisson

Número de vezes que um evento ocorre num determinado intervalo (tempo, área ou volume).

Probabilidade do evento ocorrer é a mesma para cada intervalo.

Número de ocorrências nos intervalos são independentes.

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

$\mu$  = número médio de ocorrências (frequência).

$x$  = número de ocorrência num dado intervalo.