

Arquivo 'BostonHousing.csv' carregado inicialmente com encoding 'utf-8' e separador ','.

# Análise do Dataset: BostonHousing.csv

## Visualização das primeiras linhas do dataset

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat
0	0.0063	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.9
1	0.0273	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.1
2	0.0273	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.0
3	0.0324	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.9
4	0.0691	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.9	5.3

## Visualização dos tipos de dados das colunas

	0
nox	float64
rm	float64
age	float64
dis	float64
rad	int64
tax	int64
ptratio	float64
b	float64
lstat	float64
medv	float64

# Verificar e remover dados duplicados

Número de linhas duplicadas: 0

Número de linhas duplicadas após remoção: 0

## Estatísticas descritivas

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio
count	506	506	506	506	506	506	506	506	506	506	506
mean	3.6135	11.3636	11.1368	0.0692	0.5547	6.2846	68.5749	3.795	9.5494	408.2372	18.4555
std	8.6015	23.3225	6.8604	0.254	0.1159	0.7026	28.1489	2.1057	8.7073	168.5371	2.1649
min	0.0063	0	0.46	0	0.385	3.561	2.9	1.1296	1	187	12.6
25%	0.082	0	5.19	0	0.449	5.8855	45.025	2.1002	4	279	17.4
50%	0.2565	0	9.69	0	0.538	6.2085	77.5	3.2074	5	330	19.05
75%	3.6771	12.5	18.1	0	0.624	6.6235	94.075	5.1884	24	666	20.2
max	88.9762	100	27.74	1	0.871	8.78	100	12.1265	24	711	22

## Verificar se há valores nulos

	0
crim	0
zn	0
indus	0
chas	0
nox	0
rm	0
age	0
dis	0
rad	0
tax	0

# As variaveis que escolhi para analise sao:

## crim e medv

- crim: Taxa de criminalidade per capita por cidade
- medv: Valor mediano das casas ocupadas pelos proprietários (em milhares de dólares)

## Estatísticas para crim (x) e medv (y)

### Estatísticas para crim:

	Média	Variância	Desvio Padrão	Mediana
crim	3.6135	73.9866	8.6015	0.2565

### Estatísticas para medv:

	Média	Variância	Desvio Padrão	Mediana
medv	22.5328	84.5867	9.1971	21.2

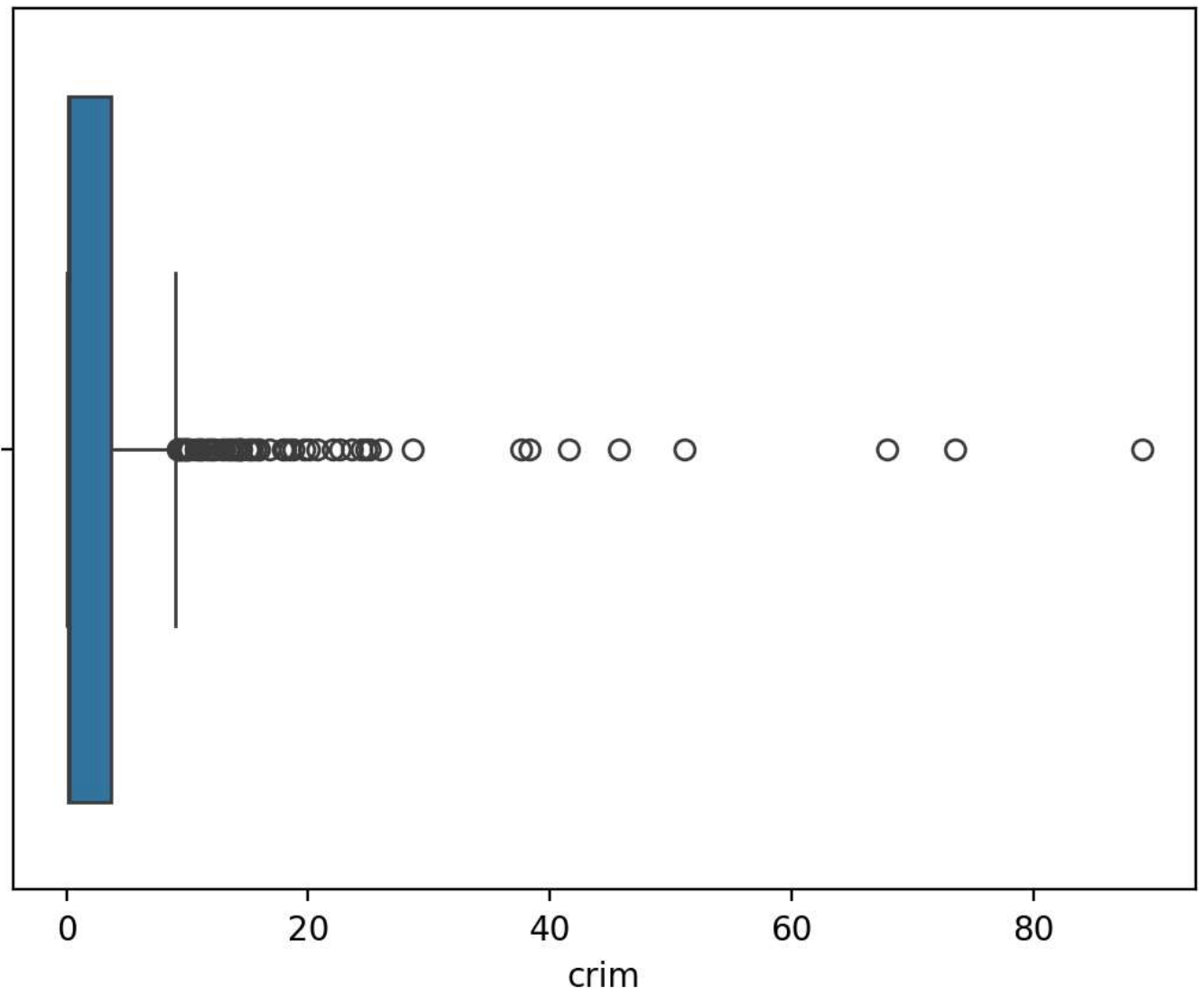
## Coeficiente de Correlação entre crim e medv

-0.3883

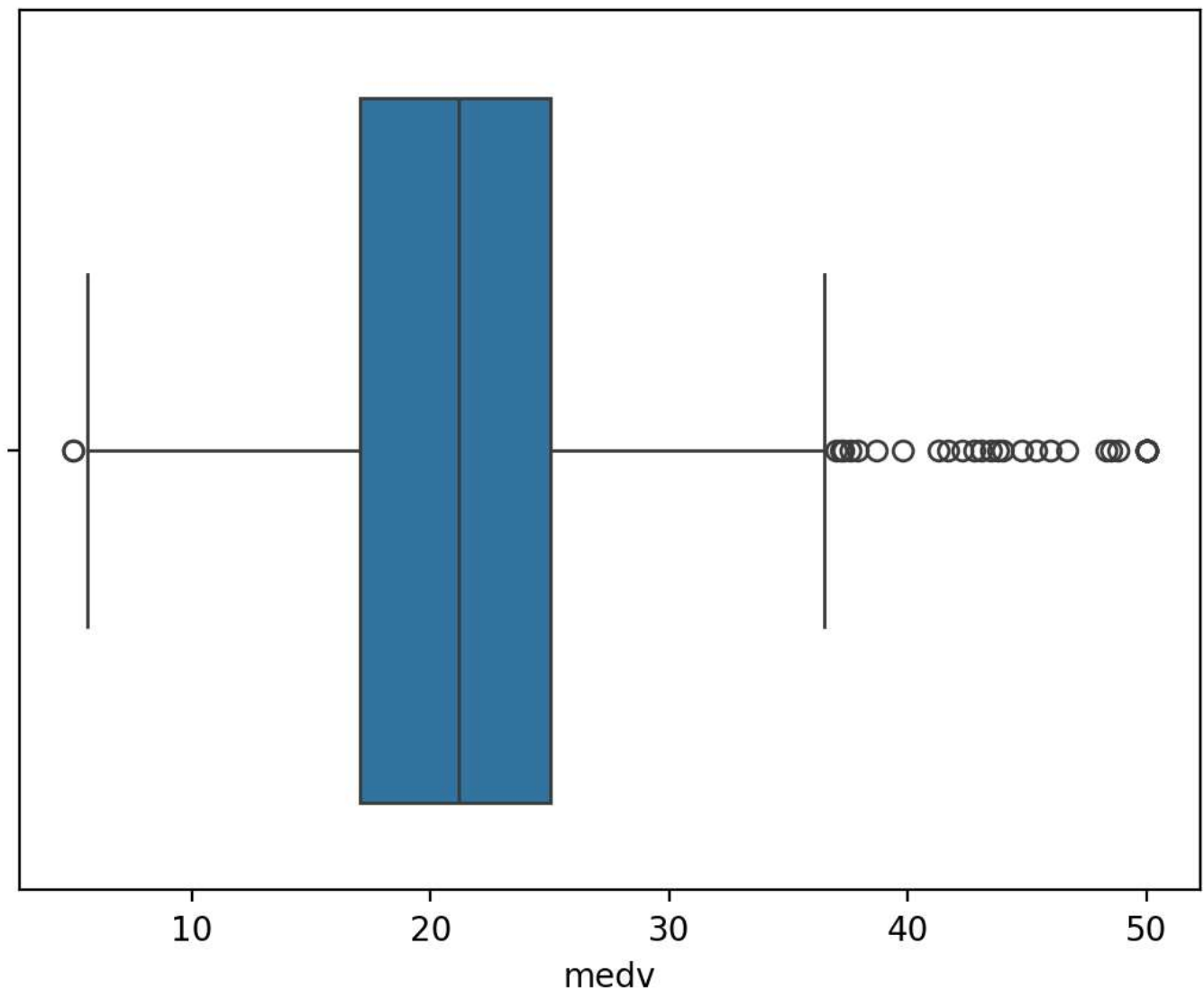
Correlação pelo numpy: -0.3883

## Boxplots para x e y

### Boxplot para crim



Boxplot para medv



## Teste de Normalidade (Shapiro-Wilk)

Teste de Shapiro-Wilk para crim:

Estatística=0.4500, p-valor=0.0000

crim não parece ser normalmente distribuído (rejeita-se H0)

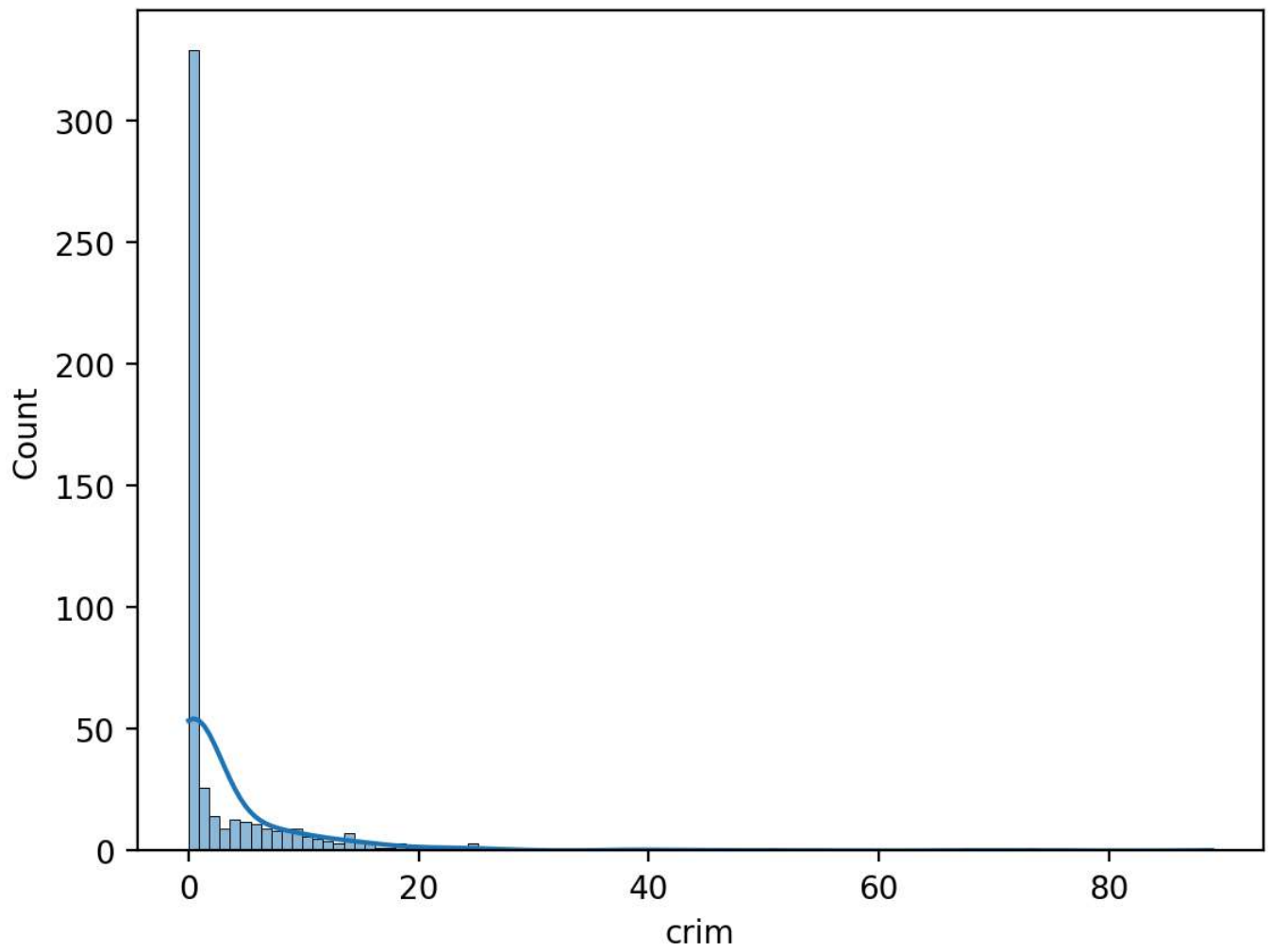
Teste de Shapiro-Wilk para medv:

Estatística=0.9172, p-valor=0.0000

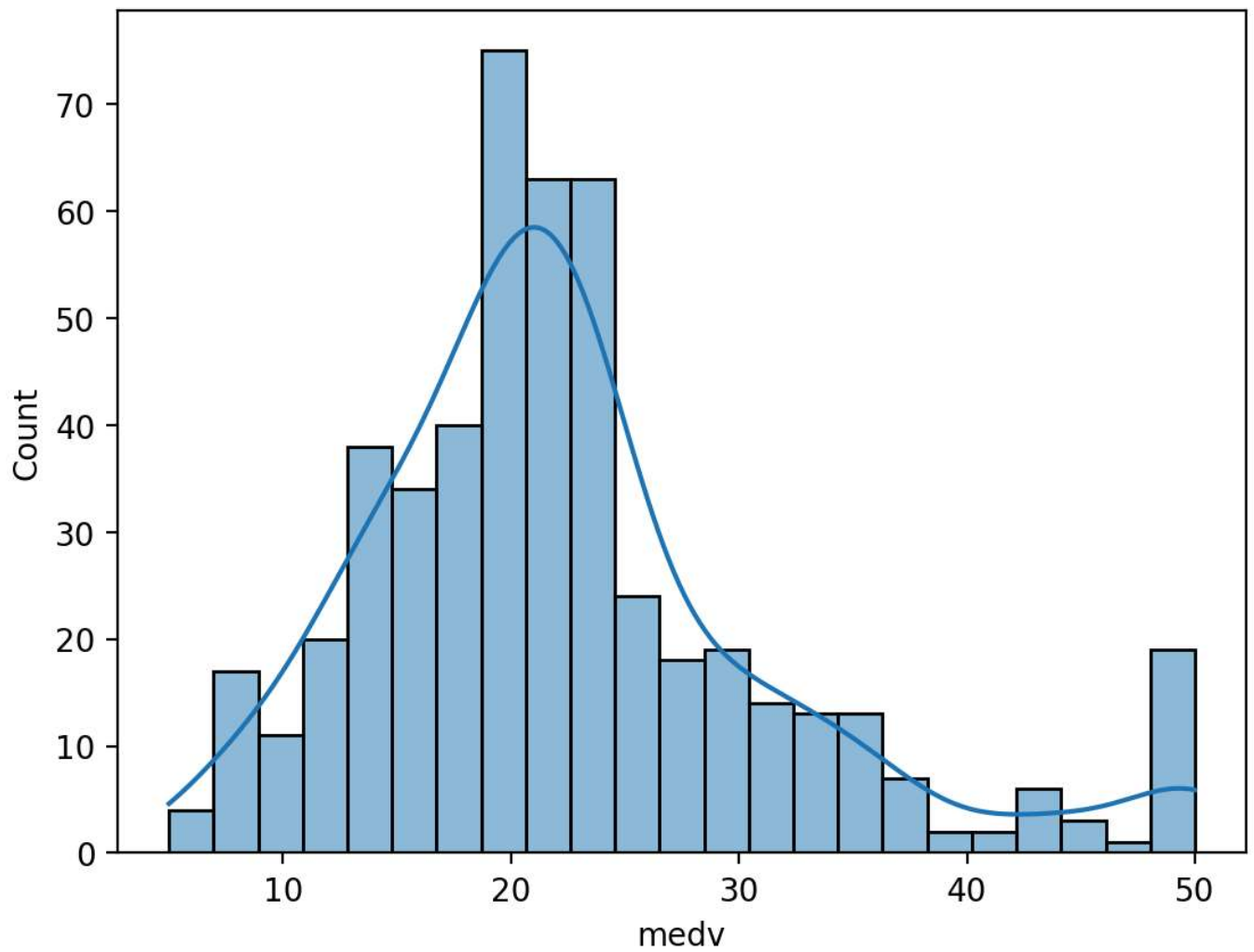
medv não parece ser normalmente distribuído (rejeita-se H0)

## Visualizar a distribuição das variáveis selecionadas

### Histograma para crim



Histograma para medv



**Gráfico de dispersão entre crim e medv**

