

Lab02 - Análise Exploratória de Dados com o R

Machine Learning com R

Thalles Quinaglia Liduares

2022-08-01

Upload dos dados

```
price = as.numeric(read.csv2('http://bit.ly/2AcWxsr', header=T,
sep=',', dec='.')$price)
```

Verifique o valor mínimo, máximo, a média, mediana, o primeiro e terceiro quartil dos dados;

A função `summary` fornece as principais estatísticas descritivas do conjunto de dados

```
summary(price)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3800   10995   13592   12962   14904   21992
```

Qual o preço mais frequente nesse conjunto de dados?

```
as.numeric(names(which(table(price)==max(table(price)))))
```

```
## [1] 12995
```

O preço mais frequente no conjunto de dados é igual a \$12995

Calcule o IQR

O IQR, que representa a dispersão dos dados em torno de uma medida de centralidade (2° e 3° quartis) é igual a \$3909.5.

```
IQR(price)
```

```
## [1] 3909.5
```

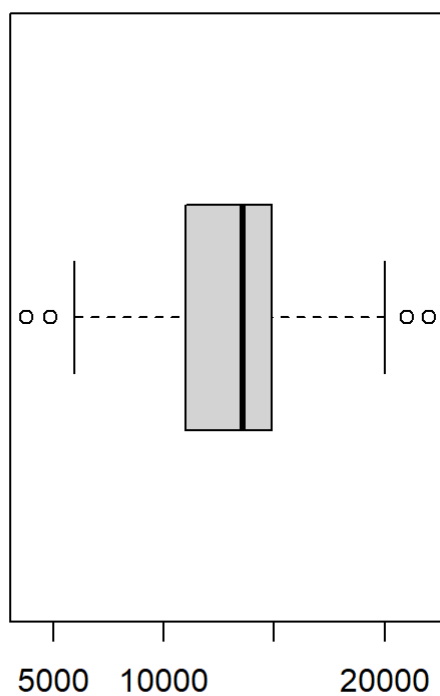
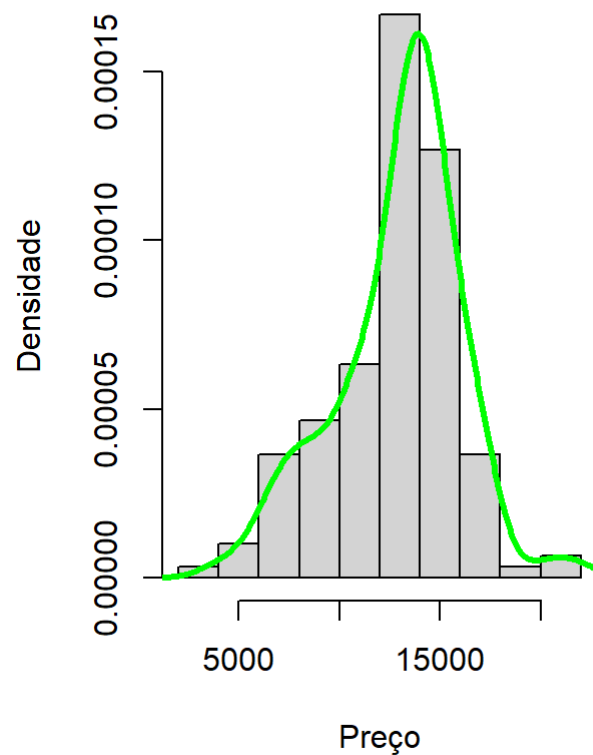
Construa o boxplot e histograma

```
par(mfrow=c(1,2))

boxplot(price,
        main='Boxplot',
        horizontal = T)

hist(price,
     main="Histograma ",
     xlab="Preço",
     ylab="Densidade",
     prob=T
    )

lines(density(price),lwd=3,col='green')
```

Boxplot**Histograma****Qual o formato da distribuição desse conjunto de dados?**

Pela forma do histograma, percebe-se que há uma assimetria a esquerda, dado que a mediana é maior do que a média.

Calcule a variância e o desvio-padrão do seu conjunto de dados

```
dados<-c(round(var(price),2), round(sd(price),2))
```

```
dados
```

```
## [1] 9749892.18    3122.48
```

Dados estes números, o conjunto de dados de preços de automóveis apresenta uma elevada variabilidade.

Qual o intervalo de preços mais provável de ser encontrado nesse conjunto de dados?

Com base nas estatísticas descritivas e gráficos analisados nos itens acima, pode-se constatar que o intervalo mais provável de preços de automóveis situa-se entre \$12500 e \$15000.