

Trabalho de Probabilidade e Estatística

Prof. Afrânio Vieira

11/04/2022

Sobre os Dados

Com a intenção de entender como variáveis associadas às características de casas na região de Ames, Iowa, USA, influenciam o preço de venda, um corretor solicitou ajuda a um cientista de dados para a análise dos dados.

A base de dados coletada, chamada **BasePE.csv**, contém 17 variáveis, todas quantitativas, associadas às 1460 casas previamente comercializadas por uma agência imobiliária. Uma breve descrição das variáveis segue abaixo:

- *Id*: identificador único;
- *LotFrontage*: área do lote à frente da casa;
- *LotArea*: Área total do lote;
- *OverallQual*: Índice de qualidade geral;
- *OverallCond*: Índice da condição de estado do imóvel;
- *YearsBult*: Idade do imóvel;
- *YearRemodAdd*: Ano em que houve reforma de expansão do imóvel;
- *TotalBsmntSF*: Área total do porão;
- *GRLivArea*: Área da sala de estar;
- *BedroomAbvGr*: Números de quarto no segundo piso;
- *TotRmsAbvGrd*: Total de cômodos no segundo piso;
- *Fireplaces*: número de lareiras na casa;
- *GarageYrBlt*: Ano que foi construída a garagem;
- *GarageCars*: Número de carros dentro da garagem;
- *GarageArea*: Área da garagem;
- *WoodDeckSF*: Área do deck de madeira;
- *SalePrice*: Preço de venda da casa (USD/10).

```
library(readr)
BasePE <- read_csv("BasePE.csv")
dim(BasePE)
```

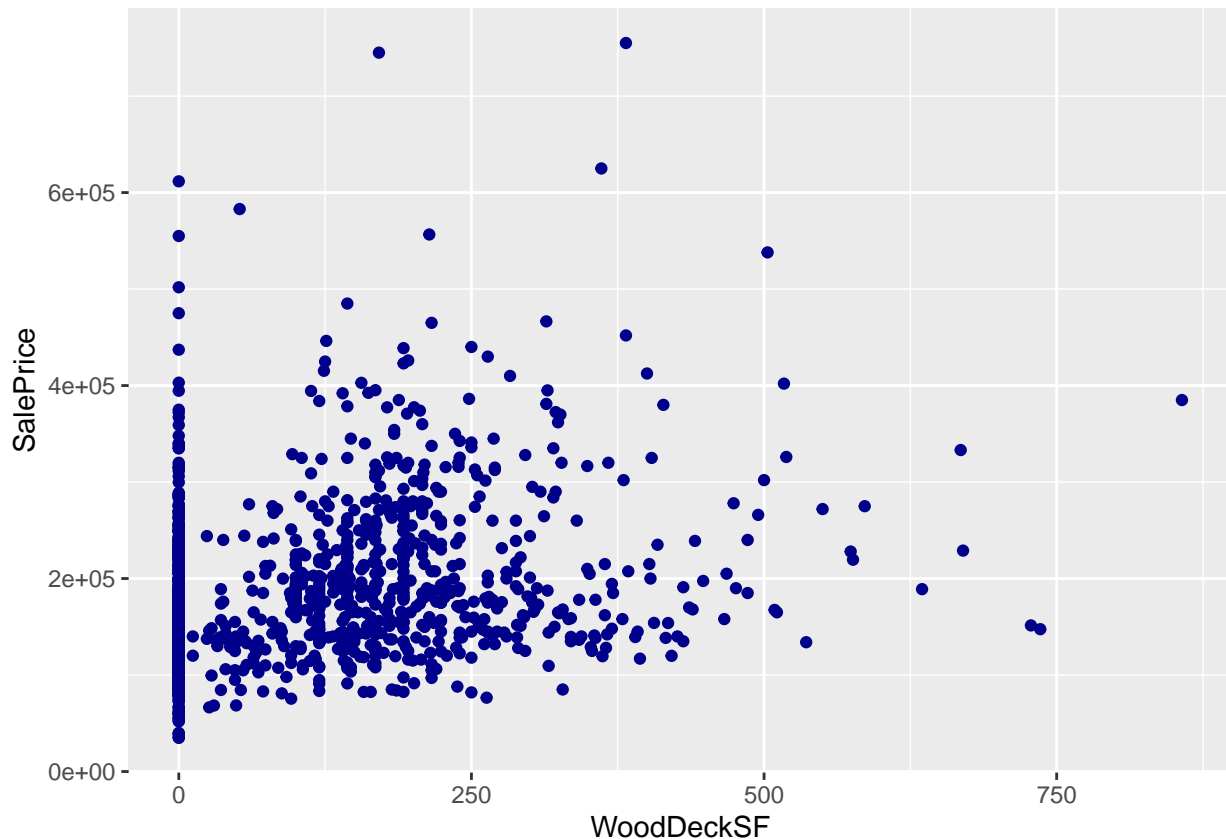
```
## [1] 1460 17
```

```
summary(BasePE)
```

```
##      Id      LotFrontage      LotArea      OverallQual
## Min.   : 1.0    Min.     : 21.00    Min.     : 1300    Min.     : 1.000
## 1st Qu.: 365.8  1st Qu.: 59.00    1st Qu.: 7554    1st Qu.: 5.000
## Median : 730.5  Median : 69.00    Median : 9478    Median : 6.000
```

```
## Mean : 730.5 Mean : 70.05 Mean : 10517 Mean : 6.099
## 3rd Qu.:1095.2 3rd Qu.: 80.00 3rd Qu.: 11602 3rd Qu.: 7.000
## Max. :1460.0 Max. :313.00 Max. :215245 Max. :10.000
## NA's :259
## OverallCond YearBuilt YearRemodAdd TotalBsmtSF GrLivArea
## Min. :1.000 Min. :1872 Min. :1950 Min. : 0.0 Min. : 334
## 1st Qu.:5.000 1st Qu.:1954 1st Qu.:1967 1st Qu.: 795.8 1st Qu.:1130
## Median :5.000 Median :1973 Median :1994 Median : 991.5 Median :1464
## Mean :5.575 Mean :1971 Mean :1985 Mean :1057.4 Mean :1515
## 3rd Qu.:6.000 3rd Qu.:2000 3rd Qu.:2004 3rd Qu.:1298.2 3rd Qu.:1777
## Max. :9.000 Max. :2010 Max. :2010 Max. :6110.0 Max. :5642
##
## BedroomAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt
## Min. :0.000 Min. : 2.000 Min. :0.000 Min. :1900
## 1st Qu.:2.000 1st Qu.: 5.000 1st Qu.:0.000 1st Qu.:1961
## Median :3.000 Median : 6.000 Median :1.000 Median :1980
## Mean :2.866 Mean : 6.518 Mean :0.613 Mean :1979
## 3rd Qu.:3.000 3rd Qu.: 7.000 3rd Qu.:1.000 3rd Qu.:2002
## Max. :8.000 Max. :14.000 Max. :3.000 Max. :2010
## NA's :81
## GarageCars GarageArea WoodDeckSF SalePrice
## Min. :0.000 Min. : 0.0 Min. : 0.00 Min. : 34900
## 1st Qu.:1.000 1st Qu.: 334.5 1st Qu.: 0.00 1st Qu.:129975
## Median :2.000 Median : 480.0 Median : 0.00 Median :163000
## Mean :1.767 Mean : 473.0 Mean : 94.24 Mean :180921
## 3rd Qu.:2.000 3rd Qu.: 576.0 3rd Qu.:168.00 3rd Qu.:214000
## Max. :4.000 Max. :1418.0 Max. :857.00 Max. :755000
##
```

```
library(ggplot2)
ggplot(BasePE, aes(x=WoodDeckSF, y=SalePrice) ) +
  geom_point(colour = "darkblue")
```



Tarefas

1. Definir os membros da **DUPLA**;
2. escolher 4 variáveis + Preço de venda da casa;
3. Fazer a análise descritiva univariada e bivariada, resumindo os dados numericamente e graficamente. Na análise bivariada, calcule correlações, construa diagramas de dispersão. Comente, **para cada variável, sobre os gráficos e estatísticas**. Lembre-se que o projeto gráfico é como você quer comunicar visualmente. Visite (<https://r-graph-gallery.com/index.html>) para liberar a criatividade.
4. Ajuste, para cada uma das 4 variáveis escolhidas, o modelo de regressão linear simples. Faça a análise de resíduos para verificar se os pressupostos de normalidade, homogeneidade da variância e independência dos resíduos.
5. Estando os modelos bem ajustados, calcule intervalos de confiança para os coeficientes angulares das retas e interprete-os.
6. Conclua o trabalho, apresente as bibliografias utilizadas.

O trabalho deve ser feito em RMarkdown. Maiores dúvidas, estou a disposição para esclarecer durante a aula de quarta feira.