



5 DE FEVEREIRO DE 2022


# DWDM – TABALHO PRATICO

OE – OFICINA DE ESTATÍSTICA

ANDRÉ PINTO 8200613 / SERGIO FELIX 8200615

ESTG – ESCOLA SUPERIOR DE TECNOLÓGIA E GESTÃO

[Endereço da empresa]



## Conteúdo

Introdução.....	2
População e Amostra .....	3
Caracterização das Variáveis.....	3
Meio .....	3
Antiguidade .....	3
Preço .....	4
Eficiência .....	4
Lucro .....	4
Comparação de Variáveis .....	5
Meio .....	5
Eficiência Energética .....	6
Regressão Linear .....	7
Preço vs Antiguidade .....	7
Lucro vs Antiguidade.....	8
Preço vs Lucro .....	8
Estatística Indutiva.....	9
Conclusão.....	12

## Introdução

No âmbito da disciplina de Oficina de Estatística do 2º ano, 1º semestre do Curso Técnico Superior Profissional em Desenvolvimento para a Web e Dispositivos Móveis, foi requisitado a análise dos dados recolhidos relativos a 25 apartamentos de uma agência imobiliária, ou seja, uma amostra, com a finalidade de inferir sobre uma população a partir de uma parte dela.

Os integrantes do grupo inicializando uma análise exploratória de dados conseguirão um entendimento básico dos seus dados e das relações existentes entre as variáveis analisadas. Após isso, o próximo passo será a análise descritiva, esta etapa é fundamental, pois uma análise descritiva detalhada vai permitir ao grupo familiarizar-se com os dados, organizá-los e sintetizá-los de forma a obter as informações necessárias do conjunto de dados para responder as questões que estão sendo estudadas.

Por fim, através da estatística indutiva será necessário procurar retirar conclusões para a população em geral, com base na análise dos resultados obtidos na amostra. Como tal, técnicas como a estimação para determinar o valor dos parâmetros desconhecidos, como por exemplo os intervalos de confiança ou então os testes de hipóteses para testar suposições acerca das características da população.

## População e Amostra

O presente estudo foi realizado em 25 imóveis de uma agência imobiliária em que, o catálogo de imóveis há venda representa a nossa população. Com ela foram retirados os seguintes dados, o meio envolvente sendo esses compreendidos entre rural e urbano, o valor antiguidade, o preço, a sua eficiência e o suposto lucro.

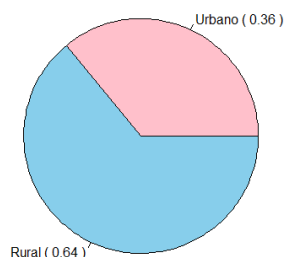
## Caracterização das Variáveis

### Meio

Esta variável, identifica o meio onde o imóvel está inserido. Os valores possíveis que esta variável pode assumir são: U (Urbano) e R (Rural). Para efeitos de codificação e manipulação de dados no R, esta variável foi codificada para inteira, em que Urbano corresponde 0 e Rural corresponde 1.

Sendo uma variável qualitativa nominal, não faz sentido o cálculo de médias. Assim apenas foram calculadas as frequências e criado o respetivo gráfico de distribuição.

Distribuição dos Imóveis por Meio



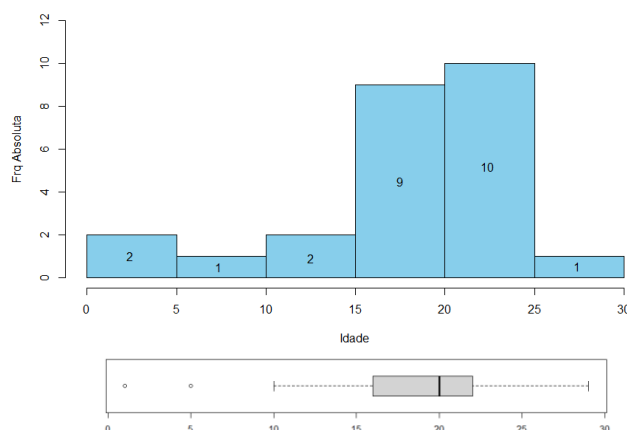
Realizamos um gráfico circular, para nos ajudar a observar a distribuição. Concluimos que, na nossa amostra, 64% dos imóveis estão inseridos em um meio rural e os restantes 36% são provenientes de um meio urbano.

### Antiguidade

Esta variável, refere-se à idade do imóvel. Os valores possíveis que esta variável pode assumir, variam entre 0 anos e mais infinito.

Sendo uma variável inteira numérica podemos classificar como uma variável quantitativa discreta. No entanto, como observamos uma grande disparidade de valores, deverá ser tratada para efeitos de análise como uma variável quantitativa contínua.

Distribuição dos Imóveis por Antiguidade



Os valores amostrais variam entre 1 até 29 anos de idade. Em média os imóveis tem aproximadamente 19 anos e mediana 20 anos. Assim identificamos dois "outliers" nomeadamente 1 e 5 anos de idade.

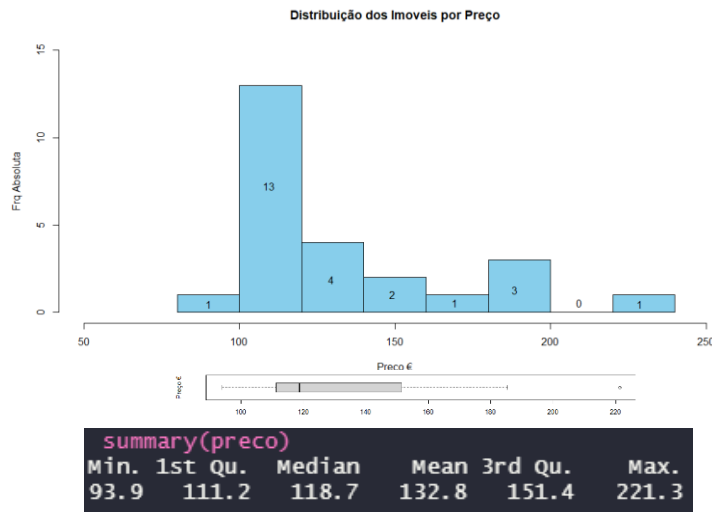
Concluimos ainda que, 50% na nossa amostra está compreendida entre 16 e 22 anos de idade. Aparentemente, será bastante significativo.

```
summary(antiguidade)
Min. 1st Qu. Median Mean 3rd Qu. Max.
1.0 16.0 20.0 18.6 22.0 29.0
```

## Preço

Esta variável, refere-se ao preço do imóvel. Os valores possíveis que esta variável pode assumir, variam entre 0 euros e mais infinito.

Sendo uma variável decimal deverá ser tratada para efeitos de análise como uma variável quantitativa continua.

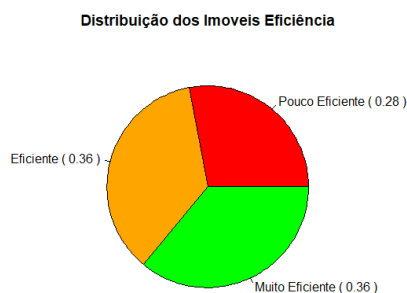


Os valores amostrais variam entre 93.9 mil e 221.3 mil euros. Em média os imoveis tem aproximadamente 132.8 mil euros. Assim identificamos 1 “outlier” nomeadamente o valor máximo de 221.3 mil euros.

De forma aparentemente significativa, podemos dizer que 50% dos imoveis custam até 118.2 mil euros e que a maior concentração está no segundo quartil de 11.2 mil a 118.2mil euros. Corresponde com o histograma que nos mostra que a classe modal é aproximadamente 100 mil a 120mil euros.

## Eficiência

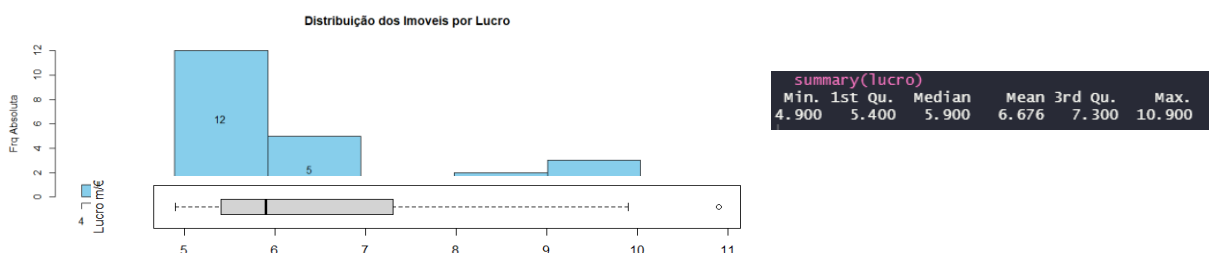
Esta variável, refere-se à classe de eficiência energética do imóvel. Estamos perante uma variável qualitativa ordinal. Pode assumir valores entre 1 para pouco eficiente e 3 muito eficiente. As variáveis já estando em valores inteiros não foi necessário realizar uma codificação.



Realizamos um gráfico circular, para nos ajudar a observar a distribuição. Concluimos que, aparentemente na nossa amostra, as distribuições são uniformes. Apenas os imoveis pouco eficientes apresentam uma ligeira discrepância de 8% com os restantes que possuem 36%.

## Lucro

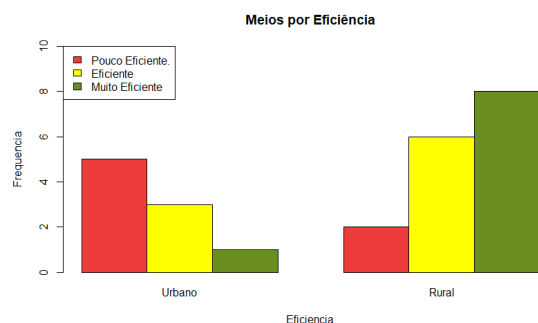
Esta variável, refere-se ao lucro estimado na venda do imóvel. Os valores possíveis que esta variável pode assumir, variam entre 0 euros e mais infinito. Sendo uma variável decimal deverá ser tratada para efeitos de análise como uma variável quantitativa continua.



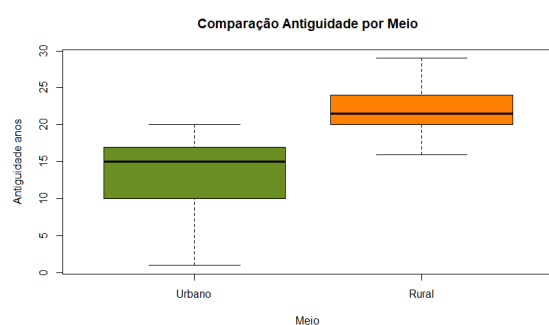
Os valores amostrais variam entre 4.9 e 10.9 milhares de euros. Em média o lucro estimado da amostra é aproximadamente 6.7 milhares de euros. Assim identificamos 1 “outlier” nomeadamente o valor máximo de 10.9 milhares de euros, pois é um valor com bastante discrepância e único. De forma aparentemente significativa, podemos dizer que 50% dos lucros variam entre 4.9 a 5,9 mil euros.

## Comparação de Variáveis

### Meio

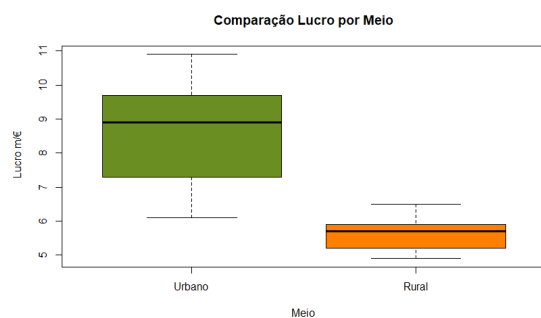


Nesta comparação entre grupos, meio por eficiência, utilizamos um gráfico de barras. Nele podemos verificar que na nossa amostra, aparentemente, os imóveis urbanos tendem a ter uma eficiência menor comparado com os imóveis rurais que tem uma eficiência energética mais elevada.



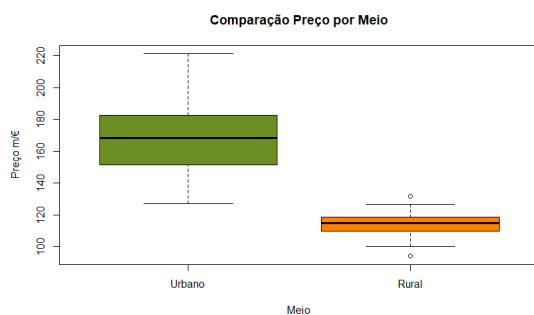
```
> tapply(antiguidade, meio, summary)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1     10     15     13     17     20
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
16.00  20.00  21.50  21.75  24.00  29.00
```

Nesta comparação entre Meio e Antiguidade, utilizamos um gráfico de extremos e quartis com 2 entradas. Nele aparentemente verificamos, uma enorme discrepância dos gráficos. No meio urbano, 50% dos imóveis estão compreendidos nas idades de 10 a 17 anos enquanto no meio rural 50% da amostra está compreendido entre 20 e 24 anos. Ou seja, aparentemente, os imóveis do meio urbano são mais recentes que os imóveis do meio rural.



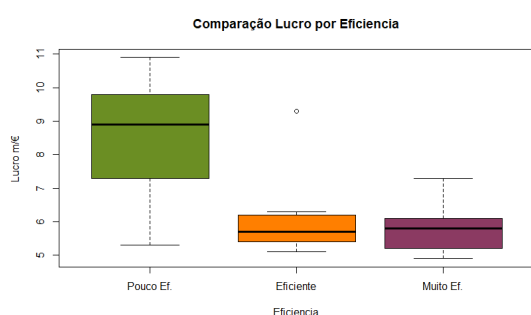
```
> tapply(lucro, meio, summary)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 6.100  7.300  8.900  8.567  9.700 10.900
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 4.900  5.200  5.700  5.612  5.900  6.500
```

Nesta comparação entre Meio e Lucro, utilizamos um gráfico de extremos e quartis com 2 entradas. Nele aparentemente verificamos, uma enorme discrepância dos gráficos. No meio rural os valores estão bastante concentrados em que 50% na amostra está concentrada entre 5.2 e 5.9 m/€ e o seu máximo apenas chega aos 6.5 m/€. Comparando com o meio urbano, este apresenta um lucro mínimo praticamente igual ao lucro máximo do meio rural. A pesar de terem uma maior amplitude, podemos dizer que aparentemente o meio urbano é o meio mais lucrativo.



Nesta comparação entre Meio e Preço, utilizamos um gráfico de extremos e quartis com 2 entradas. Nele aparentemente verificamos, uma enorme discrepância dos gráficos. No meio rural os valores estão bastante concentrados em que 50% na amostra está concentrada entre 109.7 e 118.2 m/€ e o seu máximo apenas chega aos 131.7 m/€. Comparando com o meio urbano, este apresenta um preço mínimo praticamente igual ao preço máximo do meio rural. A pesar de terem uma maior amplitude, podemos dizer que aparentemente o meio urbano o preço dos imóveis são mais elevados comparado com o meio rural. O meio urbano é o mais caro.

## Eficiência Energética



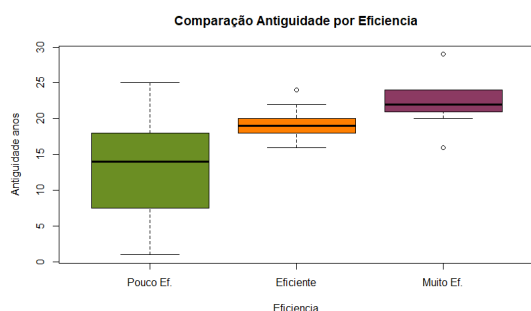
```
> tapply(Lucro, eficiencia, summary)
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 5.300  7.300   8.900   8.471  9.800  10.900

$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 5.100  5.400   5.700   6.111  6.200   9.300

$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 4.900  5.200   5.800   5.844  6.100   7.300
```

Nesta comparação entre Eficiência e Lucro, utilizamos um gráfico de extremos e quartis com 3 entradas. Nele aparentemente verificamos uma enorme discrepância no primeiro gráfico. Na eficiência energética muito eficiente e eficiente é onde há uma maior aproximação onde ambas as amostras de 50% estão concentradas entre os 5.2 e os 6.2 m/€, aparentemente tendo uma maior significância a eficiência energética eficiente no seu máximo através do outlier chegando aos 9.3 m/€ em relação aos 7.3 m/€ máximo da eficiência energética muito eficiente. Já na eficiência energética pouco eficiente existe uma grande diferença, 50% da amostra está compreendida entre 7.3 e 9.8 m/€, tendo como seu máximo os 10.9 m/€.

Com isto, aparentemente, os imóveis com eficiência energética pouco eficiente dão um retorno de lucro maior em relação aos imóveis de eficiência energética eficiente e muito eficiente.



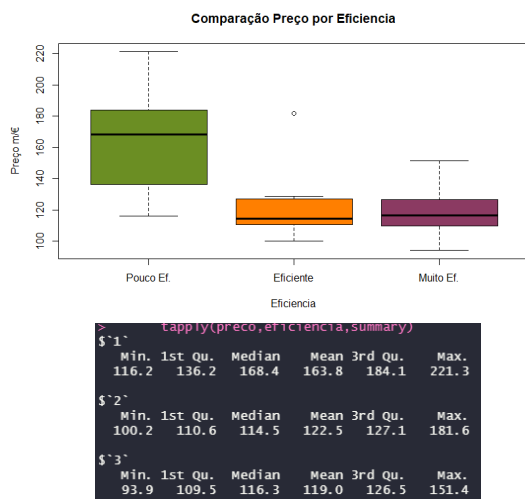
```
> tapply(antiguidade, eficiencia, summary)
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.0   7.5   14.0   13.0   18.0   25.0

$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 16.00  18.00  19.00  19.44  20.00  24.00

$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 16.00  21.00  22.00  22.11  24.00  29.00
```

Nesta comparação entre Eficiência e Antiguidade, utilizamos um gráfico de extremos e quartis com 3 entradas. Nele aparentemente verificamos uma enorme discrepância dos gráficos. Na eficiência energética pouco eficiente, 50% dos imóveis estão compreendidos nas idades de 7.5 a 18 anos enquanto na eficiência energética eficiente 50% da amostra está compreendida entre os 18 e 20 anos, já na eficiência energética muito eficiente os 50% da amostra estão compreendidos entre os 21 e 24 anos.

Ou seja, aparentemente, os imóveis com eficiência energética pouco eficiente são mais recentes que os imóveis de eficiência energética muito eficiente.

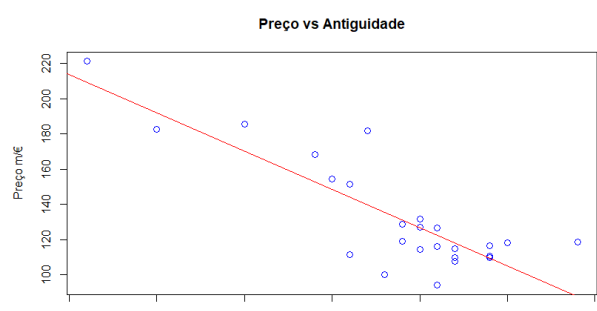


Nesta comparação entre Eficiência e Preço, utilizamos um gráfico de extremos e quartis com 3 entradas. Nele aparentemente verificamos uma enorme discrepância apenas no primeiro gráfico. Ambas as eficiências energéticas eficiente e muito eficiente possuem valores muito próximos, não demonstrando assim muita significância, por outro lado, na eficiência energética pouco eficiente é onde há uma maior significância, possuindo 50% da amostra compreendido entre 136.2 e 184.1 m/€ com o seu máximo de 221.3 m/€.

Em conclusão, aparentemente os imóveis com eficiência energética pouco eficiente têm o seu preço mais elevado em relação aos imóveis de eficiência energética eficiente e muito eficiente.

## Regressão Linear

### Preço vs Antiguidade



```

cor(antiguidade, preco)
-0.8361285
cor(antiguidade, preco)^2
0.6991108

Call:
lm(formula = preco ~ antiguidade)

Residuals:
    Min       1q   Median       3q      Max
-35.203  -9.188   0.387   7.307  41.852

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  213.6132    11.6248   18.38 3.04e-15 ***
antiguidade   -4.3450     0.5944   -7.31 1.94e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.97 on 23 degrees of freedom
Multiple R-squared:  0.6991, Adjusted R-squared:  0.686
F-statistic: 53.44 on 1 and 23 DF, p-value: 1.943e-07

```

De uma forma visual, esta nuvem de pontos é possível verificar que os dados tendem para uma relação negativa moderada, possivelmente influenciada pelos outliers. No entanto quanto maior a idade menor será o lucro.

- *Coef. de CoRelação* = - 0.84
- *Coef. de Regressão* = - 4.35
- *Coef. de Determinação* = 0.70

Segundo o coeficiente de correlação - 0.84, verificamos então que a correlação é negativa e forte pois apresenta um coeficiente de correlação maior que -0.80.

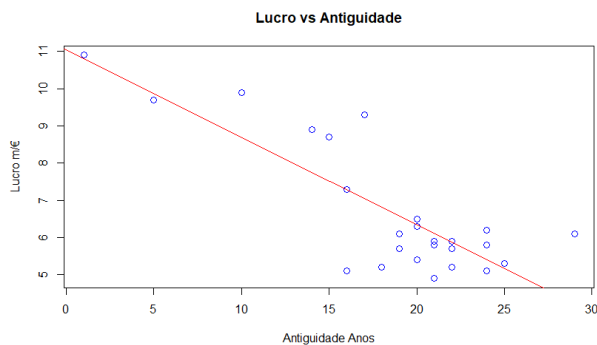
Segundo o coeficiente de determinação mostra-nos aproximadamente 70% da variância do preço é explicada pela variância da antiguidade.

Em ambos, o valor p-value é menor que o valor de significância (0.05), então rejeitamos a hipótese nula,  $\beta_0$  igual a 0, logo ficamos com hipótese alternativa,  $\beta_0$  diferente 0, e o modelo é válido. Sendo válido, é representação pela seguinte reta.  $\text{preço}^{\wedge} = 213.6 + (-4.35 * \text{antiguidade})$

Se a antiguidade for 0, o valor esperado para o preço será o valor do Intercept 213.6 m/€. Por outro lado, por cada ano de antiguidade, prevê-se que diminua -4.35 m/€ no preço do imóvel.



## Lucro vs Antiguidade



```
cor(antiguidade, lucro)
-0.8172094
cor(antiguidade, lucro)^2
0.6678313
```

```
Call:
lm(formula = lucro ~ antiguidade)

Residuals:
    Min       1Q   Median       3Q      Max
-2.18629 -0.48211 -0.04738  0.39152  2.24844

Coefficients:
(Intercept) 11.04191  0.67510  16.36 3.68e-14 ***
antiguidade -0.23473  0.03452  -6.80 6.19e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.044 on 23 degrees of freedom
Multiple R-squared:  0.6678,    Adjusted R-squared:  0.6534 
F-statistic: 46.24 on 1 and 23 DF,  p-value: 6.185e-07
```

De uma forma visual, esta nuvem de pontos é possível verificar que os dados tendem para uma relação negativa fraca, pois quanto maior a idade menor o lucro, no entanto os valores aparentemente estão muito dispersos.

- *Coef. de CoRelação* = - 0.82
- *Coef. de Regressão* = - 0.23
- *Coef. de Determinação* = 0.67

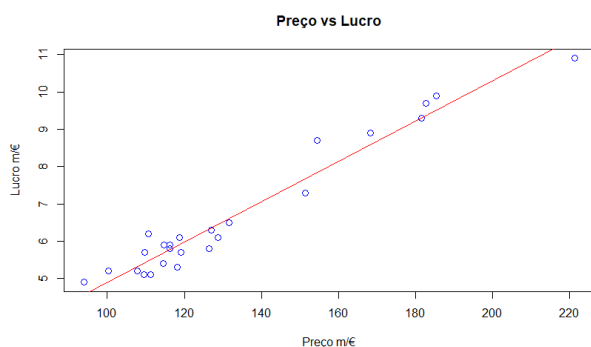
Segundo o coeficiente de correlação - 0.82, verificamos então que a correlação é negativa e forte pois apresenta um coeficiente de correlação maior que -0.80.

Segundo o coeficiente de determinação mostra-nos aproximadamente 67% da variância do lucro é explicada pela variância da antiguidade.

Em ambos, o valor p-value é menor que o valor de significância (0.05), então rejeitamos a hipótese nula,  $\beta_0$  igual a 0, logo ficamos com hipótese alternativa,  $\beta_0$  diferente 0, e o modelo é válido. Sendo válido, é representação pela seguinte reta.  $lucro^{\wedge} = 11.04 + (-0.23 * antiguidade)$

Se a antiguidade for 0, o valor esperado para o preço será o valor do Intercept 11.04 m/€. Por outro lado, por cada ano de antiguidade, prevê-se que diminua - 0.23 m/€ no preço do imóvel.

## Preço vs Lucro



```
cor(preco, lucro)
0.9758705
cor(preco, lucro)^2
0.9523232
```

```
Call:
lm(formula = lucro ~ preco)

Residuals:
    Min       1Q   Median       3Q      Max
-0.58331 -0.31944 -0.00844  0.28220  0.85870

Coefficients:
(Intercept) -0.486885  0.343411 -1.418  0.17
preco       0.053939  0.002517 21.434 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3954 on 23 degrees of freedom
Multiple R-squared:  0.9523,    Adjusted R-squared:  0.9503 
F-statistic: 459.4 on 1 and 23 DF,  p-value: < 2.2e-16
```

De uma forma visual, esta nuvem de pontos é possível verificar que os dados tendem para uma relação forte e positiva, pois quanto maior o preço, maior é lucro.

- *Coef. de CoRelação* = 0.98
- *Coef. de Regressão* = 0.054
- *Coef. de Determinação* = 0.95

Segundo o coeficiente de correlação 0.98, bastante próximo de 1, verificamos então que a correlação é positiva forte. Segundo o coeficiente de determinação mostra-nos que 95% da variância do lucro é explicada pela variância do preço.

Em ambos, o valor p-value é menor que o valor de significância (0.05), então rejeitamos a hipótese nula,  $\beta_0$  igual a 0, logo ficamos com hipótese alternativa,  $\beta_0$  diferente 0, e o modelo é válido. Sendo válido, é representação pela seguinte reta.  $\text{lucro}^{\wedge} = -0.49 + (0.054 * \text{preço})$

Se o preço for 0, o valor esperado para o preço será o valor do Intercept – 0.49 m/€. Por outro lado, por cada unidade que varie no preço, prevê-se que o lucro varie 0.054 m/€.

## Estatística Indutiva

3.1.1 Após a análise dos resultados, visto que a % obtida é muito baixa podemos considerar que a representatividade da amostra face ao meio envolvente dos apartamentos pode não ser representativa ou seja a amostra pode ser enviesada.

```
> dbinom(9,25,0.72) # P(U = 9 em 25, 0.72%) = 0.0001516296 %  
[1] 0.0001516296
```

3.1.2 Para realizar os testes a normalidade das variáveis foi realizado um teste de Shapiro-Wilk, após a análise dos resultados foi concluído que as variáveis por apresentarem um p-value inferior ao nível de significância, ou seja, que rejeitaram a hipótese nula (Ser normal), ficando assim com a hipótese alternativa ou seja não seguem uma distribuição normal.

```
data: preco  
W = 0.84014, p-value = 0.001151
```

```
data: lucro  
W = 0.80953, p-value = 0.0003292
```

```
data: antiguidade  
W = 0.89583, p-value = 0.0149
```

3.1.3 Após realizar o teste a média do preço dos imóveis considerada pelo gestor e obtendo um p-value inferior ao nível de significância (95%), ou seja, existindo evidências estatísticas de que se deva rejeitar a hipótese nula ( $h_0: \mu=150$ ), ficando assim com a hipótese alternativa ( $h_1: \mu \neq 150$ ), conclui-se que o gestor estava errado em relação a média do preço dos imóveis.

```
data: preco  
t = -2.6823, df = 24, p-value = 0.01302  
alternative hypothesis: true mean is not equal to 150  
95 percent confidence interval:  
 119.5586 146.0334  
sample estimates:  
mean of x  
 132.796
```

3.1.4 Após realizar testes a média do variável lucro ( $h_0: \mu = 6.5$ ) e antiguidade ( $h_0: \mu = 18$ ) e obtendo um p-value superior ao nível de significância (90%), ou seja, não existindo evidências estatísticas de que se deva rejeitar a hipótese nula ( $h_0: \mu=6.5$ ) e ( $h_0: \mu=18$ ), conclui-se que as médias são aceites.

```
data: lucro  
t = 0.49646, df = 24, p-value = 0.6241  
alternative hypothesis: true mean is not equal to 6.5  
90 percent confidence interval:  
 6.069478 7.282522  
sample estimates:  
mean of x  
 6.676  
data: antiguidade  
t = 0.48613, df = 24, p-value = 0.6313  
alternative hypothesis: true mean is not equal to 18  
90 percent confidence interval:  
 16.48837 20.71163  
sample estimates:  
mean of x  
 18.6
```

3.1.5 Após realizar testes as médias entre os grupos meio através do t-test com as variáveis preço, lucro e antiguidade, obtendo um p-value inferior ao nível de significância (95%), ou seja, existindo evidências estatísticas de que se deve rejeitar a hipótese nula ( $H_0: \mu_1 = \mu_2$ ), ficando assim com a hipótese alternativa ( $H_0: \mu_1 \neq \mu_2$ ).

```
data: preco by meio
t = 5.175, df = 8.8303, p-value = 0.0006197
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 29.81836 76.37470
sample estimates:
mean in group 0 mean in group 1
 166.7778      113.6813

data: lucro by meio
t = 5.2335, df = 8.6907, p-value = 0.0006043
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 1.670282 4.238052
sample estimates:
mean in group 0 mean in group 1
  8.566667      5.612500

data: antiguidade by meio
t = -3.8376, df = 10.081, p-value = 0.00323
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
-13.824723 -3.675277
sample estimates:
mean in group 0 mean in group 1
 13.00      21.75
```

Ao realizar testes as médias entre os grupos eficiências foi necessário desta vez usar o teste anova de forma a ter mais do que 3 grupos, apesar de muitas incertezas achamos que seria o mais adequado a usar.

```
Response: preco
      Df Sum Sq Mean Sq F value    Pr(>F)
eficiencia 1  7362.2   7362.2    9.7766 0.004737 **
Residuals 23 17319.9    753.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Response: antiguidade
      Df Sum Sq Mean Sq F value    Pr(>F)
eficiencia 1  316.45   316.45   12.181 0.001975 **
Residuals 23  597.55    25.98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Response: lucro
      Df Sum Sq Mean Sq F value    Pr(>F)
eficiencia 1  25.384   25.384   11.672 0.002363 **
Residuals 23   50.022    2.1749
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.2.1 Após os testes aos coeficientes dos modelos de regressão anteriormente abordados concluímos que o p-value é inferior a 0,05, para todos os testes de hipóteses, rejeitando-se assim a hipótese nula, com 95% de confiança. O que significa que, tanto o  $b_0$  como o  $b_1$ , têm um efeito significativo no modelo estimado.

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 213.6132    11.6248   18.38 3.04e-15 ***
antiguidade  -4.3450     0.5944   -7.31 1.94e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.97 on 23 degrees of freedom
Multiple R-squared:  0.6991,    Adjusted R-squared:  0.686
F-statistic: 53.44 on 1 and 23 DF,  p-value: 1.943e-07
```

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.04191     0.67510   16.36 3.68e-14 ***
antiguidade -0.23473     0.03452   -6.80 6.19e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.044 on 23 degrees of freedom
Multiple R-squared:  0.6678,    Adjusted R-squared:  0.6534
F-statistic: 46.24 on 1 and 23 DF,  p-value: 6.185e-07
```

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.486885     0.343411   -1.418    0.17
preco        0.053939     0.002517   21.434 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3954 on 23 degrees of freedom
Multiple R-squared:  0.9523,    Adjusted R-squared:  0.9503
F-statistic: 459.4 on 1 and 23 DF,  p-value: < 2.2e-16
```

3.2.2 Em todos os testes efetuados a normalidade dos erros o valor de p-value foi superior com 95% de confiança, ou seja, a hipótese nula não foi rejeitada e a normalidade dos erros segue uma distribuição normal.

Referente aos testes efetuados a média dos erros o valor de p-value foi comum em todos testes o que nos leva a não rejeitar a hipótese nula, mostrando que a média dos erros é 0, ou seja, existem poucos ou nenhuns erros.

```
data: model$residuals
W = 0.97353, p-value = 0.735
```

```
data: model$residuals
t = 3.3669e-17, df = 24, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4216906 0.4216906
sample estimates:
mean of x
6.879263e-18
```

```
data: model$residuals
W = 0.97353, p-value = 0.735
```

```
data: model$residuals
t = 3.3669e-17, df = 24, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4216906 0.4216906
sample estimates:
mean of x
6.879263e-18
```

```
data: model$residuals
W = 0.97353, p-value = 0.735
```

```
data: model$residuals
t = 3.3669e-17, df = 24, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4216906 0.4216906
sample estimates:
mean of x
6.879263e-18
```

3.2.3 Com base na reta ajustada do modelo Preço vs. Antiguidade proveu-se que o valor do preço de um determinado imóvel com 7 anos é de aproximadamente 183.15m€.

```
# preco ^ = 213.6 + (-4.35 * antiguidade)
preco = 213.6 + (-4.35 * 7) = 183.15 # aproximadamente 183.15m$
```

3.2.4 Com base na reta ajustada do modelo Preço vs. Antiguidade proveu-se que a idade máxima que um imóvel deve ter para ser vendido, no mínimo a 150m€ é de aproximadamente 15 anos.

```
# preco ^ = 213.6 + (-4.35 * antiguidade)
150 = 213.6 + (-4.35 * x) = 14.620 # aproximadamente anos = 14.620/y
```

3.2.5 Utilizando o modelo Preço vs. Lucro com uma previsão de valores “in the box” de um lucro de 10m€, o resultado do preço esperado é de aproximadamente 197.84m/€.

```
# lucro ^ = -0.49 + (0.054 * preço)
10 = -0.486 + (0.053 * x) = 197.849 # aproximadamente preco = 197.849m/$
```

Utilizando o mesmo modelo da previsão anterior, mas com uma previsão de valores “out of box” de um preço de 250m/€, o resultado do lucro esperado é de aproximadamente 13.01m€.

```
# lucro ^ = -0.49 + (0.054 * preço)
x = -0.49 + (0.054 * 250) = 13.01 # aproximadamente lucro = 13.01m$
```

Utilizando o modelo Lucro vs. Antiguidade com uma previsão de valores “in the box” de um lucro de 9m€, o resultado de anos esperado é de aproximadamente 9 anos.

```
# lucro ^ = 11.04 + (-0.23 * antiguidade)
9 = 11.04 + (-0.23 * x) = 8.869 # aproximadamente antiguidade = 8.869/y
```

Utilizando o mesmo modelo da previsão anterior, mas com uma previsão de valores “out of box” de 33 anos, o resultado do lucro esperado é de aproximadamente 3.319m€.

```
# lucro ^ = 11.04 + (-0.23 * antiguidade)
x = 11.041 + (-0.234 * 33) = 3.319 # aproximadamente lucro = 3.319m$
```

## Conclusão

Em suma, todos os objetivos propostos pela docente foram realizados com sucesso.

Durante o desenvolvimento do trabalho deparamo-nos com várias dificuldades, no entanto, estas foram facilmente ultrapassadas, consultando o material disponibilizado nas aulas, por outro lado, em decisão unânime dos elementos do grupo a parte mais complicada do desenvolvimento do trabalho foi a fase dos testes na estatística indutiva.

Graças á dedicação do grupo, foi possível cumprir todos os objetivos idealizados para a realização deste trabalho e solidificar assim ainda mais os nossos conhecimentos acerca da análise exploratória de dados amostrais, a estimação de parâmetros e a previsão com base no modelo de regressão linear simples.

Conforme o exposto, conclui-se que foi um trabalho bem conseguido, que do qual estamos orgulhosos de todos processo de aprendizagem e do resultado final.