

Preparado para:



# REFORM/SC2022/126

## DELIVERABLE 4

### MÓDULO 5

### O MODELO

DESIGNING A NEW VALUATION MODEL  
FOR RURAL PROPERTIES IN PORTUGAL

## Parte II

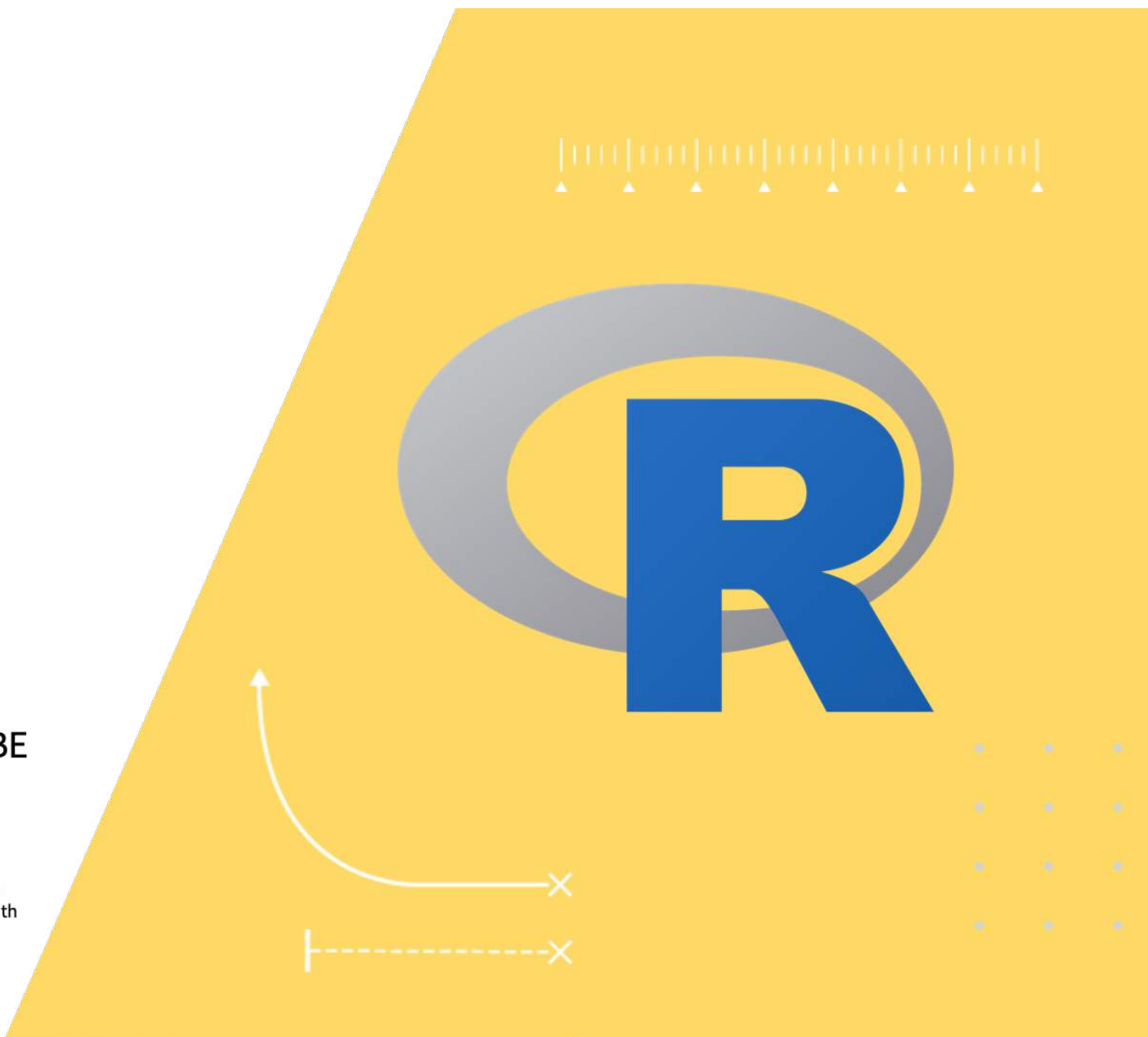
Formador: Luís Teles Morais | Nova SBE  
*Lisboa, 27 julho 2023*



This project is carried out with funding by the European Union via the Structural Reform Support Programme and in cooperation with the Directorate General for Structural Reform Support of the European Commission



LOBO VASQUES



# Programa

MÓDULOS	DURAÇÃO
<b>Módulo 1 – Introdução ao R:</b> <ul style="list-style-type: none"><li>- O que é o R?</li><li>- Como instalar e configurar o R.</li><li>- Sintaxe básica e comandos.</li><li>- Tipos de dados, objetos e classes.</li></ul>	<b>4 Horas</b>
<b>Módulo 2 – Gestão e tratamento de dados em R:</b> <ul style="list-style-type: none"><li>- Carregar dados no R.</li><li>- Perceber as estruturas de dados e <i>subsetting</i>.</li><li>- Limpeza de dados: <i>missing values</i>, <i>outliers</i> e transformações</li><li>- Juntar bases de dados</li></ul>	<b>8 Horas</b>
<b>Módulo 3 – Estatística básica em R:</b> <ul style="list-style-type: none"><li>- Estatísticas descritivas: medidas de dispersão central e variação.</li><li>- Distribuições probabilísticas: variáveis discretas e contínuas.</li><li>- Testes de hipóteses.</li></ul>	<b>8 Horas</b>

MÓDULOS	DURAÇÃO
<b>Módulo 4 – Regressão Linear:</b> <ul style="list-style-type: none"><li>- O modelo classico linear.</li><li>- Estimação de parametros segundo o MMQ.</li><li>- Testes de hipóteses: significância estatística e ajuste do modelo.</li><li>- Modelo de regressão múltipla.</li><li>- Testar as premissas: multicolinearidade, heteroscedasticidade e normalidade dos resíduos.</li><li>- Critérios de seleção dos modelos.</li></ul>	<b>12 Horas</b>
<b>Módulo 5 – O modelo:</b> <ul style="list-style-type: none"><li>- Estrutura do modelo e premissas – Perceber o modelo (4 Hours).</li></ul>	
<ul style="list-style-type: none"><li>- Uso e tratamento dos dados (4 Hours).</li><li>- Descrição do modelo (4 Hours).</li></ul>	<b>32 Horas</b>
<ul style="list-style-type: none"><li>- Aplicação do modelo a cada piloto (12 Hours).</li><li>- Aplicação autónoma do modelo a uma região (8 Hours).</li></ul>	

# across()

Operações em várias colunas ao mesmo tempo, permite repetir **mutate** e **summarize** de forma eficiente:

Coluna(s) a  
alterar

Função a aplicar:  
nome de função **f** ou **~ f(.)**

```
mutate(across(.cols, .fns, .names = "{.col}"))
```

chamar **across** dentro  
de **mutate/summarize**

```
pp_rect %>% summarize(  
  across(c("mass", "height"), mean)  
)
```

## Exercício II

1. Transforme os dados das classes para considerar apenas as classes de aptidão 1-6, seguindo a fórmula:
  - Valor de classe novo = Valor de classe antigo / (1 - A0)
2. Corra o modelo com estas novas variáveis e compare os resultados com o anterior.
3. Corra o modelo em versão log-linear e compare os resultados com o anterior.

# Flowchart

# A regressão base

- Estimar, para a amostra composta pelas freguesias  $i$  no último ano disponível:

Rendibilidade média (€/ha) - **var. dep.**

Majorantes e minorantes

$$Profitability_i = \alpha + \beta_1 A1B + \beta_2 A2B + \beta_3 A3B + \beta_4 F1B + \beta_5 F2B + \beta_6 Majs_i + \beta_7 Mins_i + \varepsilon_i$$

---

## Classes de aptidão

- Para cada freguesia, a % do território atribuída a cada uma de entre 6 categorias/classes de aptidão para a atividade agrícola
- Em princípio,  $A1B + A2B + A3B + F1B + F2B + F3B = 100\%$

# Modelo proposto

Rendibilidade média (€/ha) - **var. dep.**

**Classes de aptidão**

$$Profitability_i = \alpha + \beta_1 A1B + \beta_2 A2B + \beta_3 A3B + \beta_4 F1B + \beta_5 F2B + \beta_6 DIST\_CONC\_KM + \beta_7 Majs_i + \beta_7 Mins_i + \varepsilon_i$$

**Distância ao centro do município**

- Impacto na rendibilidade do terreno face a custos de distribuição etc.

# Seleccção de variáveis

## Critérios

- Aumentar R2 do modelo
- Coeficiente estatisticamente significativo

VARIABLE CODE	VARIABLE NAME	STATISTICAL SIGNIFICANCE	ADJUSTED $R^2$
CAALL	All conservation areas	SIGNIFICANT AT 1%	↑
TI_M	<u>Thermicity index</u>	SIGNIFICANT AT 1%	↑
OI_M	Annual <u>ombrothermic index</u>	SIGNIFICANT AT 1%	↑
ELE_M	Elevation	SIGNIFICANT AT 1%	↑
TRI5_M	Terrain ruggedness index (5-cell window)	SIGNIFICANT AT 1%	↑
SRR25_M	Surface relief ratio (25-cell window)	NOT SIGNIFICANT	↓
TPI5_M	Topographic position index (5-cell window)	NOT SIGNIFICANT	↓
TPI15_M	Topographic position index (15-cell window)	NOT SIGNIFICANT	↓
TPI25_M	Topographic position index (25-cell window)	NOT SIGNIFICANT	↓
GPIN_M	Grassland Productivity Index	SIGNIFICANT AT 10%	↓
CPIN_M	Cropland Productivity Index	SIGNIFICANT AT 10%	↑
N2ALL	All Natura 2000	NOT SIGNIFICANT	=
NAR	National Agricultural Reserve	SIGNIFICANT AT 1%	↑
PNF	Primary Network of <u>Fuelbreaks</u>	SIGNIFICANT AT 1%	↑
HAA	Hydro-Agricultural Areas	NOT SIGNIFICANT	=



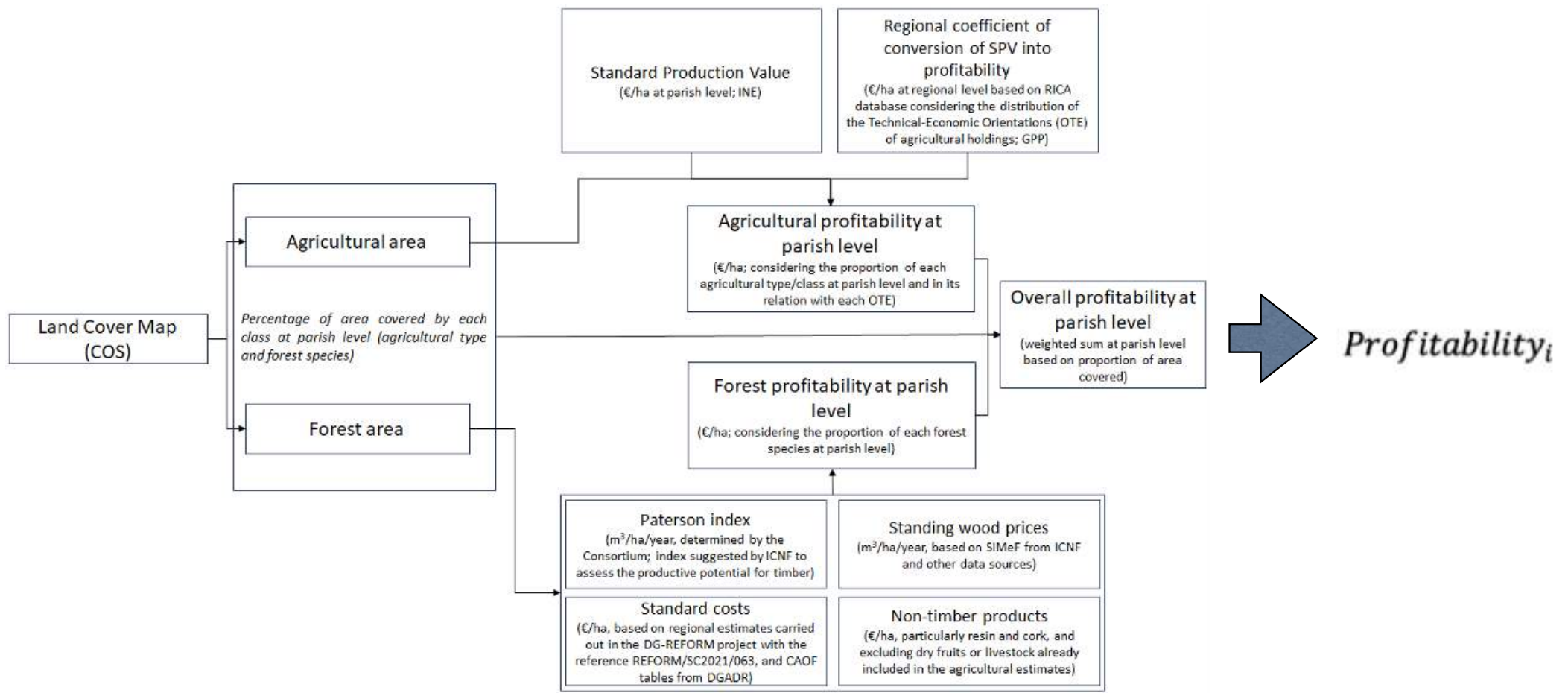
# Modelo proposto

## Especificação

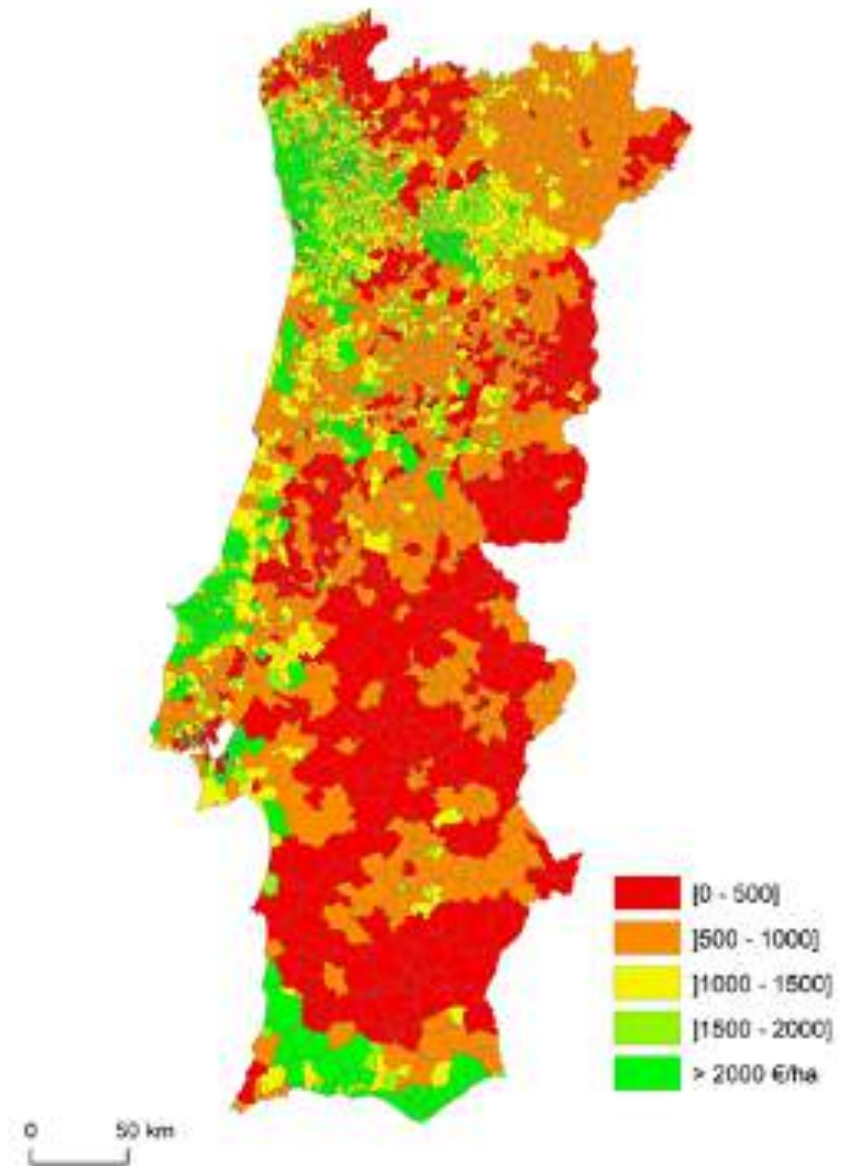
$$\begin{aligned} Profitability_i = & \alpha + \beta_1 A1B + \beta_2 A2B + \beta_3 A3B + \beta_4 F1B + \beta_5 F2B + \beta_6 DIST\_CONC_{KM} \\ & + \beta_7 CAAll + \beta_8 Ti\_M + \beta_9 Oi\_M + \beta_{10} ELE\_M + \beta_{11} TRI5\_M + \beta_{12} CPIn\_M + \beta_{13} NAR \\ & + \beta_{14} PNF + \varepsilon_i \end{aligned}$$

VARIABLE CODE	VARIABLE NAME
CAALL	All conservation areas
TI_M	<u>Thermicity Index</u>
OI_M	Annual <u>ombrothermic index</u>
ELE_M	Elevation
TRI5_M	Terrain ruggedness index (5-cell window)
CPIN_MN	Cropland Productivity Index
NAR	National Agricultural Reserve
PNF	Primary Network of <u>Fuelbreaks</u>

# Rendibilidade



# Rendibilidade



# add\_predictions()

Acrescenta uma coluna com valores previstos com base num modelo já estimado, utilizando os dados contidos numa tabela. **Pode ser dados novos, não utilizados na estimação.**

```
add_predictions(data, model, var = "pred")
```

Utilizar este  
modelo

Para gerar valores  
previstos/  
estimados, em  
função destes  
dados

Nome da nova  
coluna

```
wages %>% add_predictions(mod_h)
```

	height <dbl>	weight <int>	age <int>	marital <fctr>	sex <fctr>	education <int>	afqt <dbl>	pred <dbl>
	60	155	53	married	female	13	6.841	10.102158
	70	156	51	married	female	10	49.444	10.621947
	65	195	52	married	male	16	99.393	10.362053
	63	197	54	married	female	14	44.022	10.258095
	66	190	49	married	male	14	59.683	10.414032
	68	200	49	divorced	female	18	98.798	10.517989
	64	160	54	divorced	female	12	50.283	10.310074
	69	162	55	divorced	male	12	89.669	10.569968
	69	194	54	divorced	male	13	95.977	10.569968
	64	145	53	married	female	16	67.021	10.310074

1–10 of 5,266 rows | 2–9 of 9 columns

Previous 1 2 3 4 5 6 ... 100 Next

# spread\_predictions()

Semelhante a `add_predictions` mas para vários modelos. Acrescenta valores previstos para vários modelos, cada um na sua própria coluna.

```
spread_residuals(data, m1, m2, m3, ...)
```

**Acrescenta  
valores previstos  
com bases nestes  
modelos**

**A esta tabela,  
calculados com base  
nos valores da  
mesma**

```
wages %>%
```

```
spread_residuals(mod_h, mod_eh, mod_ehs)
```

income <int>	height <dbl>	sex <fctr>	education <int>	mod_h <dbl>	mod_eh <dbl>	mod_ehs <dbl>
19000	60	female	13	-0.2499641042	-0.197967581	-0.2638576582
35000	70	female	10	-0.1588437767	0.345993609	0.7237344726
105000	65	male	16	1.1996628894	0.853872025	0.5063344524
40000	63	female	14	0.3385397443	0.262834114	0.3124199119
75000	66	male	14	0.8112117773	0.746516842	0.4591017275
102000	68	female	18	1.0147387260	0.402532860	0.6229479494
70000	64	female	12	0.8461766568	1.051566955	1.1612752115
60000	69	male	12	0.4321315995	0.655873056	0.5117444599
150000	69	male	13	1.3484223314	1.433450940	1.2800521289
115000	64	female	16	1.3426135431	0.993152447	1.0657798463

1-10 of 5,266 rows

Previous [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) ... [100](#) Next



# Exercício III

1. Carregue os dados `fake_parcelas.csv` (processando corretamente quaisquer NAs).
2. Com base no modelo **m2** já estimado, calcule os valores previstos para a rendibilidade de cada parcela utilizando “`add_predictions`”.
  - Acrescente também uma coluna com o LBV estimado para cada parcela.
  - Componha um gráfico que relacione a distância ao centro urbano com o LBV.
  - Repita os passos para os dados originais (`dados_modelo`). O que se conclui?



**Obrigado  
e até à próxima!**

luis.morais@novasbe.pt