

Preparado para:



REFORM/SC2022/126

DELIVERABLE 4

MÓDULO 5

O MODELO

DESIGNING A NEW VALUATION MODEL
FOR RURAL PROPERTIES IN PORTUGAL

Parte I

Formador: Luís Teles Morais | Nova SBE
Lisboa, 4 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

AARC

NOVA
NOVA SCHOOL OF
BUSINESS & ECONOMICS

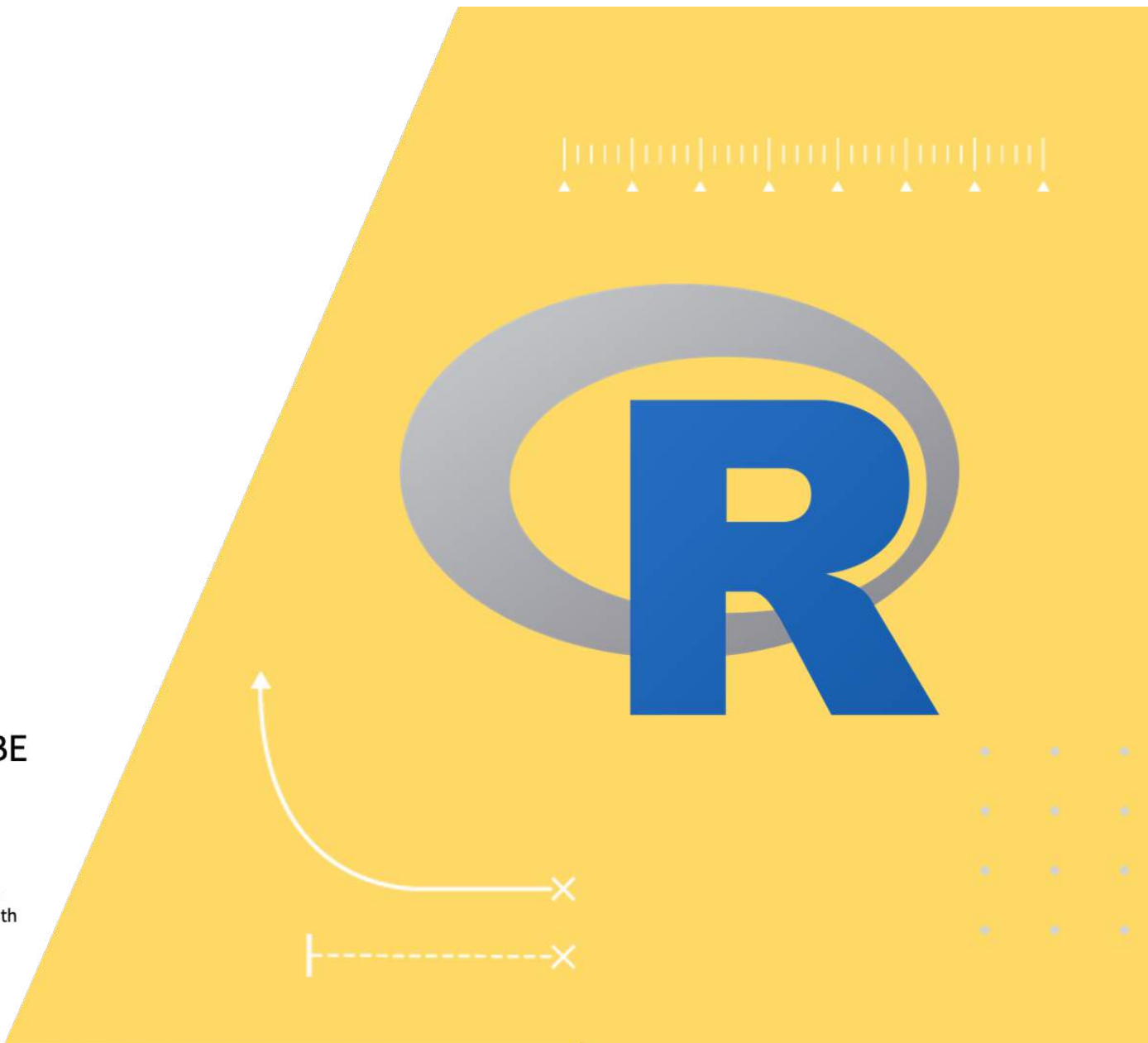
LOBO VASQUES



esri



INNERLANDS

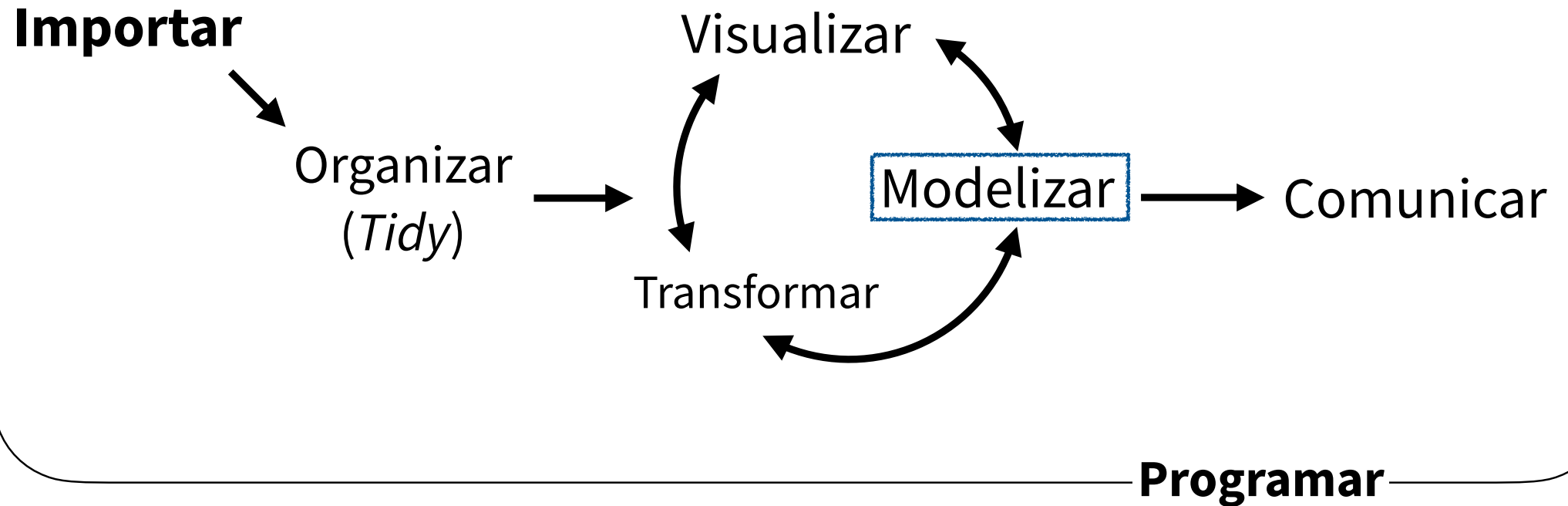


Programa

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

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

Ciência de dados



readxl::read_excel()

Importar dados a partir de Excel

```
read_excel("ficheiro.xlsx", sheet = "data", ...)
```

**Nome de caminho do
ficheiro a importar**

A partir do
working directory

getwd()

**Nome de folha
dentro do livro
Excel**

Outras opções

```
modelo_quad <- lm(price ~ Width_in, data = pp_rect)
modelo_quad
```

Call:

```
lm(formula = price ~ Width_in, data = pp_rect)
```

Coefficients:

(Intercept)	Width_in
376.77	19.47
$\hat{\alpha}$	$\hat{\beta}$

Incerteza?

tidy()

Obtém resultados essenciais do modelo

```
modelo_quad %>% tidy()
```

```
# A tibble: 2 × 5
```

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Intercept)	368.	59.3	6.21	6.15e-10
2	Width_in	19.6	2.22	8.81	1.95e-18



glance()

Diagnóstico da qualidade do ajustamento

```
modelo_quad %>% glance()
```

```
# A tibble: 1 × 12
```

	r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.0242	0.0239	1941.	77.7	1.95e-18	1	-28200.	56405.	56424.

summary()

Obtém resultados essenciais do modelo (não em tabela)

```
modelo_quad %>% summary()
```

```
Call:
lm(formula = price ~ Width_in, data = pp_rect)

Residuals:
    Min       1Q   Median       3Q      Max
-4790.8  -702.6  -516.5  -124.8  27985.8

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   376.77      60.40    6.237 5.06e-10 ***
Width_in      19.46       2.25    8.652 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1956 on 3081 degrees of freedom
Multiple R-squared:  0.02372,    Adjusted R-squared:  0.0234
F-statistic: 74.85 on 1 and 3081 DF,  p-value: < 2.2e-16
```

ao contrário de tidy() não
são tabelas reutilizáveis

*** níveis de
significância



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

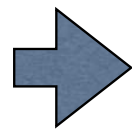
- Para cada freguesia, a % do território atribuída a cada uma de entre 6 categorias/classes de aptidão para a atividade agrícola
- Em princípio, $A1B + A2B + A3B + F1B + F2B + F3B = 100\%$

Objetivo

- Estimar os diferentes **coeficientes** β_1, β_2, \dots

$$Profitability_i = \alpha + \beta_1 A1B + \beta_2 A2B + \beta_3 A3B + \beta_4 F1B + \beta_5 F2B + \beta_6 Maj s_i + \beta_7 Min s_i + \varepsilon_i$$

- Utilizá-los para calcular **valores previstos** para parcelas ao nível local, com diferentes combinações de valores de aptidão (e majorantes/minorantes)



$\widehat{Profitability}_i$

`%<>%`

Passa objeto à função e atribui resultado ao mesmo objeto (a um objeto novo com o mesmo nome)

```
# library(magrittr)
starwars %<>% mutate(palavra = "abc")
```

**Objeto a
alterar**

qualquer verbo dplyr

Sinónimo de

```
starwars <- starwars %>% ...
```

Exercício I

```
library(tidyverse)  
library(magrittr)  
library(broom)
```

1. Carregue os dados da matriz de teste e junte com os dados da variável dependente, combinando (**join**) os dados por código de freguesia [basta completar os espaços no código].
2. Corra o modelo já preparado no código, na secção Exercício I. Qual é o problema deste modelo?

**Obrigado
e até à próxima!**

luis.morais@novasbe.pt