



Universidade de Brasília  
Departamento de Estatística  
EST0077 - 2023/1

## **SOLUÇÃO DOS EXERCÍCIOS PARA ENTREGA 5**

Igor de Oliveira Barros Faluhelyi  
Prof. George

Brasília  
2023

## Sumário

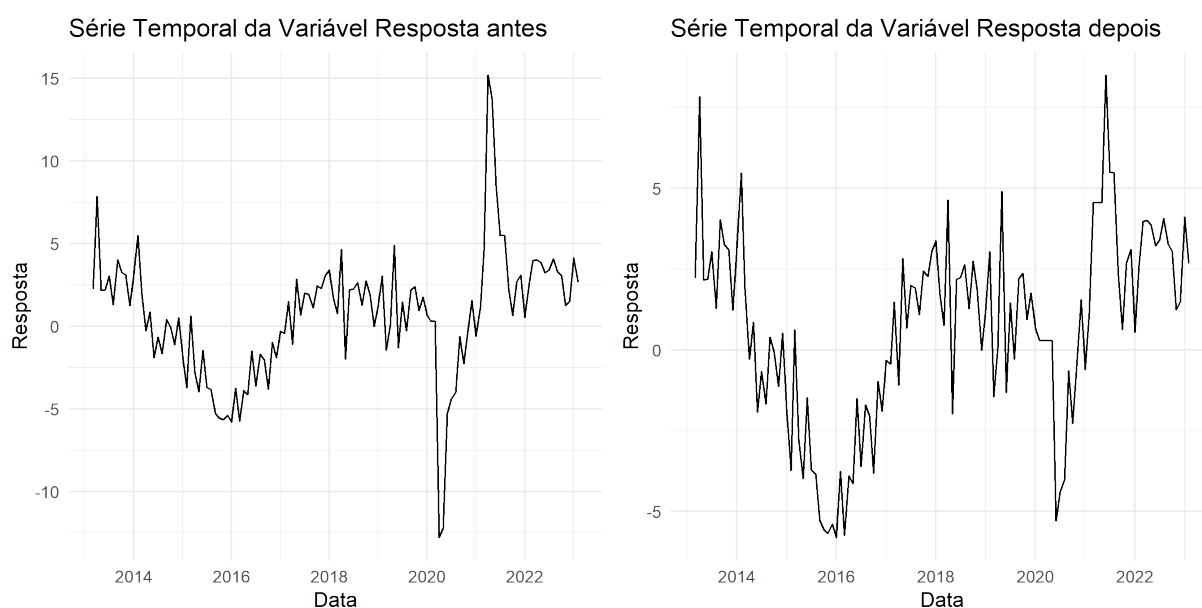
<b>1 Lista 6 - Redução PCR e PLS . . . . .</b>	<b>4</b>
1.1 Ajuste no banco de dados . . . . .	4
1.2 Resumo dos resultados. . . . .	4
1.2.1 <i>Leave-One-Out</i> . . . . .	5
1.2.2 RMSEP fora da amostra . . . . .	6

# 1 Lista 6 - Redução PCR e PLS

## 1.1 Ajuste no banco de dados

Antes de propor a solução para o exercício, foram feitas algumas alterações em períodos específicos nas séries temporais. Fixados os períodos, as mudanças seguem a mesma regra para todas as variáveis. No primeiro passo, os valores observados para todas variáveis nos meses nos anos "2020-04", "2020-05", "2021-04" e "2021-05" foram substituídos por *missing values*. No segundo momento, é feita a imputação desses "dados faltantes" pelo respectivo valor observado no mês anterior.

Para ilustrar o resultado do procedimento, segue o plot para a série temporal da variável resposta, respectivamente, antes e depois:



Veja que o comportamento geral da série se mantém, porém o efeito dos períodos, que antes geravam dados discrepantes, mencionados vai poder ser melhor captado pelo modelo linear ao decorrer do exercício e o ajuste aos dados vai ser melhor. A argumentação segue válida para todas as covariáveis do exercício.

## 1.2 Resumo dos resultados

Os códigos R foram escritos e executados em um computador com Windows 10 Pro; Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz como Processador e 8.00 GB (utilizável:

7.72 GB) de RAM instalada.

### 1.2.1 *Leave-One-Out*

Segue, na tabela, o respectivo RMSEP para as diferentes retenções de informação em  $\mathbf{X}$  considerando PCR ou PLS.

Retenção	# Componentes	RMSEP		
		PCR manual	PCR no pls do R	PLS no pls do R
70%	4	1.1386270	1.1351774	0.9153081
80%	8	0.9768882	0.9768882	0.7565003
90%	17	0.9811538	0.9811014	0.6518926
95%	29	0.9760095	0.9444652	0.7177428

O RMSEP pela implementação do pacote pls do R não bate exatamente com a implementação manual do PCR, pois, esta, considera um número de CPs variável para cada iteração do LOOCV, visto a retenção especificada. Então a coluna de Números de componentes principais da tabela anterior é exatamente válida para os RMSEPs calculados pela implementação no pacote pls e, como *Trade off*, a coluna de Retenção vale aproximadamente para eles (i.e. a % de retenção é ligeiramente maior). O exato contrário vale para a implementação manual, isto é, a coluna de Retenção vale exatamente, porém a coluna de Número de componentes seria uma média dos números de componentes utilizadas nas iteração do LOOCV para a respectiva retenção. Sendo essas médias: 3.991667, 8, 16.95 e 29.05. Por isso na retenção de 80% os RMSEPs "batem na vírgula".

O número de componentes utilizados no modelo linear em cada iteração do LOOCV feito manualmente com a retenção especificada segue na relação (ao total são 120 iterações do LOOCV):

1. Retenção de 70%: 99% das iterações usaram 4 CPs e 1% usaram 3;
2. Retenção de 80%: 100% das iterações usaram 8 CPs;
3. Retenção de 90%: 99% das iterações usaram 17 CPs e 1% usaram 16;
4. Retenção de 95%: 95% das iterações usaram 29 CPs e 5% usaram 30.

Então fica claro o motivo do maior descolamento entre a implementação manual e a do pls no R na retenção de 95% da informação e os menores nos demais, inclusive o descolamento zero na retenção de 80%.

A implementação manual do PCR e o LOOCV escrita em R rodou eu 2.737376 mins, enquanto a implementação no pls do R rodou, tanto o PCR, quanto o PLS, e seus LOOCVs em aproximadamente 10 secs cada um.

### 1.2.2 RMSEP fora da amostra

O LOOCV sugere a retenção de 80% para o PCR e 90% para o PLS. Então o quadro resumo dos resultados (predição para cada iteração) vão seguir essa retenção fixada.

#### Predições

Date	y	PCR manual	PLS
Mar-21	4.55873758	3.27802186	3.66312892
Apr-21	4.55873758	4.35099649	4.53308627
May-21	4.55873758	4.46993962	4.55056124
Jun-21	8.50793651	11.8831792	11.7487741
Jul-21	5.4958184	4.66762158	5.22993512
Aug-21	5.49253731	6.0214596	4.69262521
Sep-21	2.29007634	2.75386228	2.29631703
Oct-21	0.63805104	0.60614503	1.23659813
Nov-21	2.69005848	1.91826894	1.73973517
Dec-21	3.1030445	2.61265923	3.53011114
Jan-22	0.54777845	1.53430489	2.20005558
Feb-22	2.61437908	2.18126438	0.3021276
Mar-22	3.9686976	2.02635185	2.71137766
Apr-22	4.00929692	3.29537467	5.7326736
May-22	3.85514019	5.63583808	4.30441957
Jun-22	3.21825629	4.88287067	3.4667894
Jul-22	3.39750849	4.1641852	4.08451288
Aug-22	4.07470289	4.58135083	3.8324967
Sep-22	3.27210103	3.46802459	4.03405896
Oct-22	3.05475504	2.70474811	3.10962758
Nov-22	1.25284738	2.24309107	2.36987121
Dec-22	1.47643384	0.99468647	0.44957715
Jan-23	4.11622276	1.17673092	1.5428811
Feb-23	2.66357846	0.38483796	1.31717573

O RMSEP foi de 1.261655 para o PLS e de 1.331318 para o PCR manual. Como indicado no LOOCV, o PLS se mostrou mas adequado nesse caso.