

**Matheus Carl Beer**

**Modelos de tendência e ARIMA para  
análise e previsão de matrículas de uma  
escola de ensino de idiomas nos anos de  
2017 a 2022**

Niterói - RJ, Brasil

31 de maio de 2023

**Matheus Carl Beer**

**Modelos de tendência e ARIMA para  
análise e previsão de matrículas de  
uma escola de ensino de idiomas nos  
anos de 2017 a 2022**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof<sup>ª</sup>. Dr<sup>ª</sup>. Márcia Marques de Carvalho

Niterói - RJ, Brasil

31 de maio de 2023

**Matheus Carl Beer**

**Modelos de tendência e ARIMA para análise  
e previsão de matrículas de uma escola de  
ensino de idiomas nos anos de 2017 a 2022**

Monografia de Projeto Final de Graduação sob o título  
*“Modelos de tendência e ARIMA para análise e previsão de  
matrículas de uma escola de ensino de idiomas nos anos de  
2017 a 2022”*, defendida por Matheus Carl Beer e aprovada em  
31 de maio de 2023, na cidade de Niterói, no Estado do Rio de  
Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

---

**Prof. Dr. Moisés Lima de Menezes**  
Departamento de Estatística – UFF

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Ludmilla da Silva Viana Jacobson**  
Departamento de Estatística – UFF

Niterói, 31 de maio de 2023

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

B415m Beer, Matheus Carl  
Modelos de tendência e ARIMA para análise e previsão de matrículas de uma escola de ensino de idiomas nos anos de 2017 a 2022 / Matheus Carl Beer. - 2023.  
56 f.: il.

Orientador: Márcia Marques de Carvalho.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2023.

1. Covid-19. 2. Modelo de Regressão Linear. 3. Séries temporais. 4. Quebra estrutural. 5. Produção intelectual. I. Carvalho, Márcia Marques de, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

# Resumo

Por causa da pandemia de Covid-19, diversos setores trabalhistas tiveram que fazer adaptações para a modalidade virtualmente ou remota. Esse trabalho tem como objetivo principal analisar o impacto da pandemia na quantidade de alunos ingressantes e matriculados em uma escola de ensino de idiomas na cidade do Rio de Janeiro. Modelos de Regressão Linear e o teste de Chow foram estimados para verificar se houve quebra estrutural no número de alunos matriculados por mês durante a pandemia, segundo a cliente e a filial. Os resultados mostraram que houve quebra estrutural em quase todas as localidades / clientelas. Modelos ARIMA foram estimados para previsão de ingressantes no último trimestre de 2022 e para o primeiro trimestre de 2023.

Palavras chaves: Covid-19. Modelo de Regressão Linear. Séries temporais. Teste de Chow. Quebra estrutural.

# Agradecimentos

Agradeço ao apoio da minha família, especialmente minha orientadora Márcia que mesmo nesse momento de pós pandemia sempre esteve disponível em todos os momentos para me orientar da melhor forma possível. À todos amigos que me ajudaram a superar esta jornada e ao meu estágio pela disponibilidade das informações para a realização deste trabalho.

# Lista de Figuras

1	Funil Lead . . . . .	p. 11
2	Matrículas Ativas filial 1 . . . . .	p. 12
3	Matrículas Ativas filial 2 . . . . .	p. 13
4	Matrículas Ativas - Criança e adolescente nas filiais 1 e 2 . . . . .	p. 14
5	Matrículas Ativas - Adulto nas filiais 1 e 2 . . . . .	p. 14
6	Ingressante Geral . . . . .	p. 15
7	Ingressante filial 1 . . . . .	p. 15
8	Ingressante filial 2 . . . . .	p. 15
9	Modelo sem quebra estrutural . . . . .	p. 30
10	Modelo com quebra estrutural . . . . .	p. 30
11	Modelo sem quebra estrutural . . . . .	p. 32
12	Modelo com a queda estrutural . . . . .	p. 32
13	Modelo sem quebra estrutural . . . . .	p. 34
14	Modelo com quebra estrutural . . . . .	p. 34
15	Modelo sem quebra estrutural . . . . .	p. 35
16	Modelo com quebra estrutural . . . . .	p. 35
17	Modelo sem quebra estrutural . . . . .	p. 38
18	Modelo com quebra estrutural . . . . .	p. 38
19	Modelo sem quebra estrutural . . . . .	p. 40
20	Modelo com a queda estrutural . . . . .	p. 40
21	Modelo sem quebra estrutural . . . . .	p. 41
22	Modelo com quebra estrutural . . . . .	p. 41

23	Modelo sem quebra estrutural . . . . .	p. 43
24	Modelo com quebra estrutural . . . . .	p. 43
25	Correlograma e correlograma parcial - Filial 1 . . . . .	p. 45
26	ARIMA (4,0,0) . . . . .	p. 46
27	ARIMA (3,0,2) . . . . .	p. 47
28	Correlograma dos resíduos - Filial 1 . . . . .	p. 48
29	Previsão 2 trimestres - Filial 1 . . . . .	p. 49
30	Correlograma e correlograma parcial - Filial 2 . . . . .	p. 49
31	ARIMA(4,0,0) . . . . .	p. 51
32	ARIMA(3,0,0) . . . . .	p. 51
33	Correlograma dos resíduos - Filial 2 . . . . .	p. 52
34	Previsão próximos 2 trimestres - Filial 2 . . . . .	p. 53



# Lista de Tabelas

1	Matrículas Ativas por Clientelas, total de 2017 a 2022 . . . . .	p. 13
2	Variável binária $D_1$ . . . . .	p. 21
3	Padrões Teóricos de FAC e FACP . . . . .	p. 26
4	Resultados do modelo sem quebra - Filial 1 - Criança e Adolescente . .	p. 29
5	Resultados da estatística de aderência . . . . .	p. 29
6	Resultados do modelo com quebra - Filial 1 - Criança e Adolescente . .	p. 29
7	Resultados estatística de aderência . . . . .	p. 30
8	Resultados do modelo sem quebra - Filial 1 - Adulto . . . . .	p. 31
9	Resultados estatística de aderência . . . . .	p. 31
10	Resultados do modelo com quebra - Filial 1 - Adulto . . . . .	p. 31
11	Resultados estatística de aderência . . . . .	p. 32
12	Resultados do modelo sem quebra - Filial 2 - Criança e Adolescente . .	p. 32
13	Resultados estatística de aderência . . . . .	p. 33
14	Resultados do modelo com quebra - Filial 2 - Criança e Adolescente . .	p. 33
15	Resultados estatística de aderência . . . . .	p. 33
16	Resultados do modelo sem quebra - Filial 2 - Adulto . . . . .	p. 34
17	Resultados estatística de aderência . . . . .	p. 34
18	Resultados do modelo com quebra - Filial 2 - Adulto . . . . .	p. 35
19	Resultados estatística de aderência . . . . .	p. 35
20	Resumo da análise da quebra estrutural antes e durante a pandemia . .	p. 36
21	Resultados do modelo sem quebra - Filial 1 - Criança e Adolescente . .	p. 37
22	Resultados estatística de aderência . . . . .	p. 37

23	Resultados do modelo com quebra - Filial 1 - Criança e Adolescente . .	p. 37
24	Resultados estatística de aderência . . . . .	p. 37
25	Resultados do modelo sem quebra - Filial 1 - Adulto . . . . .	p. 38
26	Resultados estatística de aderência . . . . .	p. 38
27	Resultados do modelo com quebra - Filial 1 - Adulto . . . . .	p. 39
28	Resultados estatística de aderência . . . . .	p. 39
29	Resultados do modelo sem quebra - Filial 2 - Criança e Adolescente . .	p. 40
30	Resultados estatística de aderência . . . . .	p. 40
31	Resultados do modelo com quebra - Filial 2 - Criança e Adolescente . .	p. 41
32	Resultados estatística de aderência . . . . .	p. 41
33	Resultados do modelo sem quebra - Filial 2 - Adulto . . . . .	p. 42
34	Resultados estatística de aderência . . . . .	p. 42
35	Resultados do modelo com quebra - Filial 2 - Adulto . . . . .	p. 42
36	Resultados estatística de aderência . . . . .	p. 42
37	Resumo da análise da quebra estrutural durante e após a pandemia . .	p. 44
38	Resultado do Modelo ARIMA - Filial 1 . . . . .	p. 46
39	Estatística de Aderência . . . . .	p. 46
40	Estatística de Aderência - Filial 1 . . . . .	p. 47
41	Resultado do modelo ARIMA - Filial 1 . . . . .	p. 47
42	Estatística de Aderência . . . . .	p. 48
43	Previsão dos Ingressantes - Filial 1 . . . . .	p. 48
44	Resultado do Modelo ARIMA - Filial 2 . . . . .	p. 50
45	Estatística de Aderência . . . . .	p. 50
46	Estatística de Aderência - Filial 2 . . . . .	p. 51
47	Resultado do melhor modelo ARIMA - Filial 2 . . . . .	p. 52
48	Estatística de Aderência . . . . .	p. 52

49	Previsão dos Ingressantes - Filial 2 . . . . .	p.53
----	--	------

# Sumário

<b>1</b>	<b>Introdução</b>	p. 8
1.1	Objetivos . . . . .	p. 9
1.2	Organização . . . . .	p. 10
<b>2</b>	<b>Materiais e Métodos</b>	p. 11
2.1	Dados . . . . .	p. 11
2.1.1	Matrícula ativa . . . . .	p. 12
2.1.2	Ingressante . . . . .	p. 15
2.2	Métodos . . . . .	p. 16
2.2.1	Série Temporal . . . . .	p. 16
2.2.2	Modelo de Regressão Linear . . . . .	p. 18
2.2.3	Teste de Chow . . . . .	p. 21
2.2.4	Modelos ARIMA . . . . .	p. 23
2.2.5	Estatísticas de Aderência . . . . .	p. 26
2.2.6	Proposta . . . . .	p. 27
<b>3</b>	<b>Resultados</b>	p. 28
3.1	Quebra estrutural . . . . .	p. 28
3.2	Quebra estrutural, 2017 a 2021, antes e durante a pandemia . . . . .	p. 28
3.2.1	Filial 1: criança e adolescente . . . . .	p. 28
3.2.2	Filial 1, Adulto . . . . .	p. 31
3.2.3	Filial 2, Criança e adolescente . . . . .	p. 32

3.2.4	Filial 2: Adulto . . . . .	p. 34
3.3	Quebra estrutural, 2020 a 2022: durante e após a pandemia . . . . .	p. 36
3.3.1	Filial 1: Criança e adolescente . . . . .	p. 37
3.3.2	Filial 1: Adulto . . . . .	p. 38
3.3.3	Filial 2: Criança e adolescente . . . . .	p. 40
3.3.4	Filial 2: Adulto . . . . .	p. 42
3.4	Previsão, Introdução . . . . .	p. 44
3.4.1	Modelagem ARIMA para o Filial 1 . . . . .	p. 45
3.4.2	Modelagem ARIMA para a Filial 2 . . . . .	p. 49
<b>4</b>	<b>Conclusão</b>	<b>p. 54</b>
	<b>Referências</b>	<b>p. 56</b>

# 1 Introdução

No *marketing* tradicional, costuma-se utilizar a comunicação para divulgação de mensagens e promover serviços ou produtos em anúncios em rádios, televisão, outdoor ou mídias em geral. Estas abordagens são classificadas como **diretas** e estão presentes em horários comerciais ou dentro de produções, sendo encontradas em diálogos e na produção do cenário utilizando um grande volume financeiro investido para atender a essas realizações.

Segundo Moraes (2018), o *marketing* no setor digital atua em meios “como” sites de pesquisas e propagandas em redes sociais utilizados por diversos usuários por um financiamento menor em relação proporcional ao números de pessoas atingidas pelo método tradicional. Outra vantagem está na facilidade de conseguir informações pessoais dos usuários quando disponibilizam o nome, número de celular, e-mail e idade a fim de obter divulgações de serviços ou ofertas daquela marca. Além disso, temos acesso às atividades mapeadas do usuário com a utilização de *cookies*, onde toda movimentação é registrada e armazenada desde o momento da entrada do site, cliques em *links*, abertura de e-mails, tempo médio gasto nas plataformas, entre outros, podendo assim obter análises de acesso e assim associar a uma publicação de *marketing* em determinado tempo com o comportamento dos usuarios dentro das plataformas. Esse conjunto de pessoas é denominado por *Lead*.

A fim de obter um filtro mais preciso para denominar os níveis de interesse, o *Lead* é dividido em MQL (*Marketing Qualified Lead*) e SQL (*Sales Qualified Lead*). Os MQL são *Leads* com estágio avançados afetados pelo *marketing*. Estes MQL podem se tornar SQL, que são pessoas com grande potencial de se tornar clientes preenchendo dados de identificação a fim de obter contato com a empresa BANFIELD (2017). Após passar por um tratamento interno entre *Leads*, o cliente por fim adquire o produto.

Em uma escola de idiomas, a fim de obter uma análise comportamental no número de entrada e saída de alunos, foi adotado o mesmo método utilizado nas empresas de capital

aberto de divulgação trimestral SOUZA (2018). A cada 3 meses fecha-se um trimestre e começa um novo. A contagem começa em janeiro e acaba em dezembro, totalizando 4 trimestres:

- Inter I: Janeiro a Março;
- Intra II: Abril a Junho;
- Inter II: Julho a Setembro;
- Intra II: Outubro a Dezembro.

Os modelos propostos neste trabalho serão aplicados a duas filiais de uma escola de idiomas localizada no município do Rio de Janeiro e utilizam a metodologia *Lead*. Os trimestres *Inter* representam momentos de grande procura na escola de idiomas por se tratar do início das aulas, já os trimestres *Intra* são momentos de baixa procura.

Um desafio presente nesta escola de idiomas está na previsão de entrada de novos alunos e no impacto sofrido durante os anos entre 2020 a 2021, uma vez que o foco do ensino é a modalidade presencial, seguido da implementação do ensino remoto durante a pandemia de Covid-19. Este TCC tem objetivo de analisar a entrada de novos alunos por trimestre através da modelagem ARIMA.

Os modelos e as análises serão realizados pelo software livre GRETLL (*acrônimo de Gnu Regression, Econometrics and Time-series Library*) para compilar e interpretar dados econométricos.

## 1.1 **Objetivos**

Os objetivos da pesquisa são:

1- Avaliar a quebra estrutural da matrícula ativa mensal durante e após a pandemia de Covid-19.

2- Avaliar a tendência de entrada da matrícula ativa mensal para cada filial durante e após a pandemia de Covid-19.

3- Gerar previsão de entrada de ingressantes para os trimestres *Inter II* de 2022 e *Intra I* de 2023 por filial e verificar se atingiu a meta.

Para atingir esses objetivos, os dados serão fornecidos pela escola de idiomas para as filiais 1 e 2 por tipo de clientela.

## 1.2 Organização

Este trabalho está organizado em mais 3 capítulos, além deste capítulo introdutório. No Capítulo 2 encontra-se os materiais e métodos onde serão apresentados um resumo das bases de dados e a descrição da metodologia de modelagem. No Capítulo 3 estão as análises dos resultados e os modelos estimados e no Capítulo 4 a conclusão.



## 2 Materiais e Métodos

Neste capítulo, na seção 2.1 são apresentados os dados a serem utilizados neste trabalho e na seção 2.2 é responsável por apresentar os métodos.

### 2.1 Dados

Criado por Lewis (1915), o funil de vendas tem por objetivo ser um medidor de etapas para separar a amostra em subconjuntos. A Figura 1 apresenta os níveis de filtro estratégico do Lead MQL até se tornar Ingressante.



Figura 1: Funil Lead

Onde o Lead MQL está em fase de interação com propagandas e cliques, seja no site da empresa ou até mesmo nos emails recebidos. Após atingir certo nível de interação, o Lead passa a se tornar SQL, que são pessoas aptos a venda. Nesse momento a escola entra em contato com o aluno para um tratamento nivelando o nível de conhecimento e assim realizar a venda do curso.

A variável de interesse desta pesquisa é o número de alunos matriculados em uma escola privada de idiomas presencial e ensino a distância, preservando a identidade da instituição, utilizando dados desde 2017 a 2022 de alunos pagantes de matrículas ativas mensais totalizando 3.802.327 observações e dados de 2016 a 2022 de ingressantes trimestrais totalizando 111.750 observações, ambas localizadas no município do Rio de Janeiro.

### 2.1.1 Matrícula ativa

Matrícula Ativa é um conjunto de alunos ativos junto aos novos alunos ingressantes na instituição de ensino. Este conjunto é formado por um subconjunto chamado de *clientela* representado por: conversação; preparatório; avançado; adulto; criança e adolescente.

As Figuras 2 e 3 representam as matrículas ativas com as cinco clientelas agrupadas entre 2017 a 2022 nas filiais 1 e 2, respectivamente.

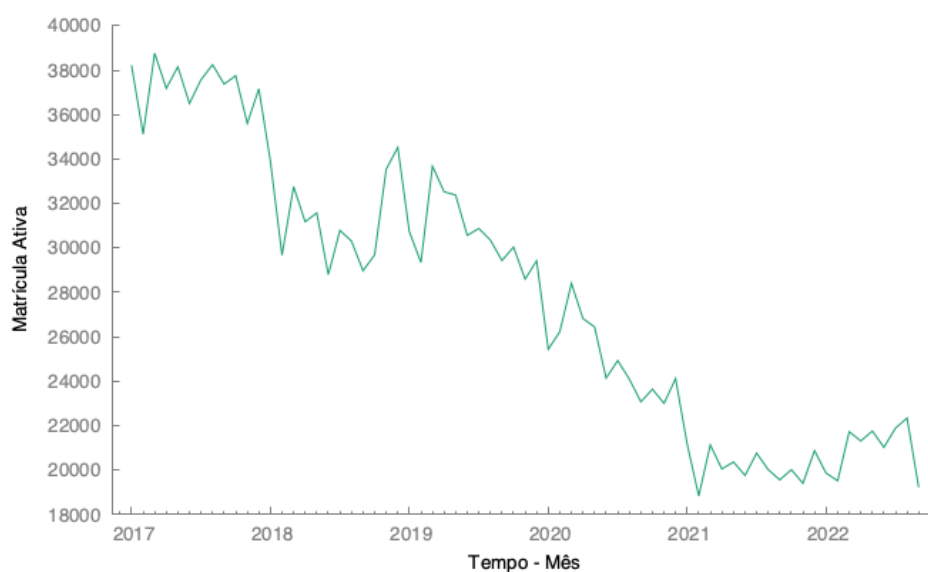


Figura 2: Matrículas Ativas filial 1

Pode-se perceber, na Figura 2, uma diminuição no número de matrículas ativas entre 2018 e 2019. Durante a pandemia de 2020 e 2021 houve menos matrículas ativas. Após 2021 não houve grandes ausências de matrículas ativas.



Figura 3: Matrículas Ativas filial 2

Conforme pode ser visto na Figura 3, as matrículas ativas tiveram um comportamento decrescente desde 2019 até 2022 de forma gradual e não houve momentos agressivos de redução.

A Tabela 1 apresenta a proporção por clientela por filial nos anos de 2017 a 2022.

Tabela 1: Matrículas Ativas por Clientelas, total de 2017 a 2022

filial	Clientela	N.º	%
1	Adulto	190.319	10,24
	Avançado	427.395	22,99
	Conversação	19.104	1,03
	Criança e adolescente	1.165.052	62,67
	Preparatório	57.094	3,07
	<b>Total</b>	<b>1.858.964</b>	<b>100</b>
2	Adulto	261.472	13,45
	Avançado	442.686	22,78
	Conversação	57.388	2,95
	Criança e adolescente	1.090.087	56,09
	Preparatório	91.730	4,72
	<b>Total</b>	<b>1.943.363</b>	<b>100</b>

Como as clientelas Adulto, Criança e Adolescente são maiores em termos de matrículas, esta pesquisa está concentrada somente nestas 2 clientelas. A clientela Avançado está ausente das análises por não ser um curso regular, cursado semestralmente.

A Figura 4 apresenta as séries temporais do número de matrículas ativas da clientela "Criança e Adolescente" nas filiais 1 e 2.

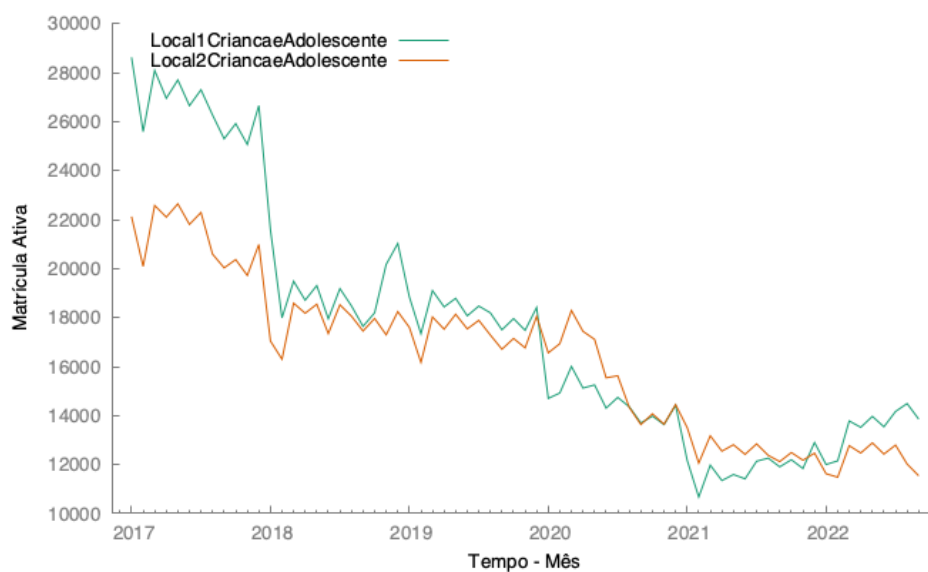


Figura 4: Matrículas Ativas - Criança e adolescente nas filiais 1 e 2

Pode-se perceber que ambas as séries apresentam comportamento similares. Contudo, a partir de 2022 a filial 1 apresenta uma tendência de crescimento maior que a filial 2. Já a Figura 5 apresenta as séries temporais do número de matrículas ativas da clientela Adulto nas filiais 1 e 2.

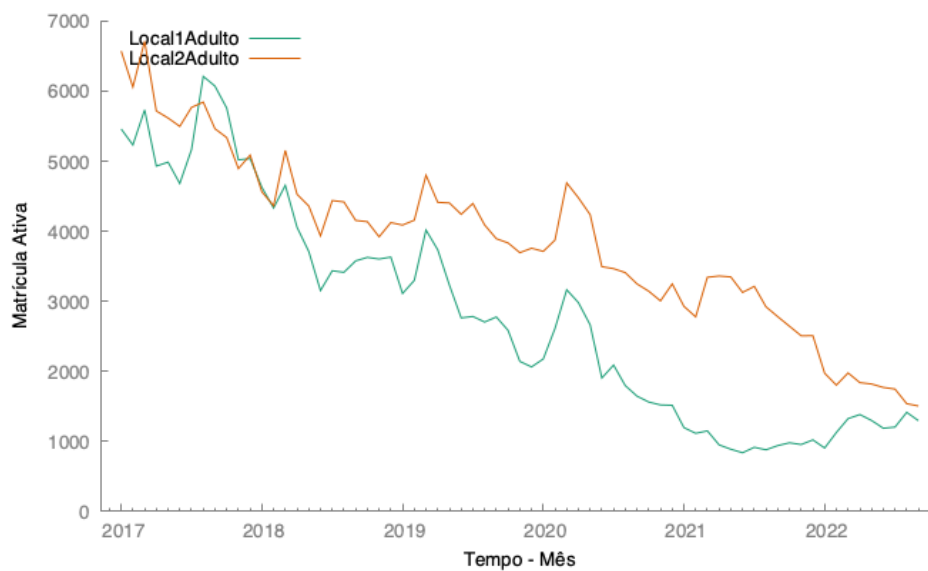


Figura 5: Matrículas Ativas - Adulto nas filiais 1 e 2

Pode-se observar uma queda mais expressiva no filial 1 no início de 2018. E ambos possuem tendência decrescentes de matrículas ativas.

## 2.1.2 Ingressante

Ingressantes é um conjunto de alunos novos na instituição de ensino após ter seu pagamento confirmado.

Foram coletados dados da entrada de novos alunos ingressantes por trimestre nos anos de 2016 a 2022. Na Figura 6, referente às unidades do Rio de Janeiro, são observados grandes entradas nas épocas do início do ano, dando inícios a novas turmas com durações semestrais, referentes aos trimestre Inter.

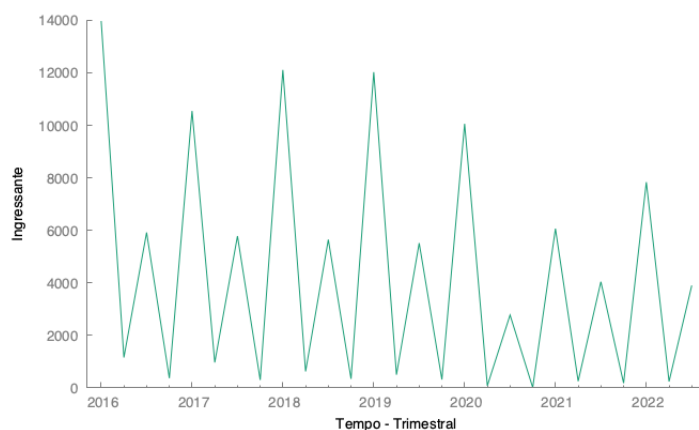


Figura 6: Ingressante Geral

Nas Figuras 7 e 8 são observadas a quantidade de alunos ingressantes em duas filiais.

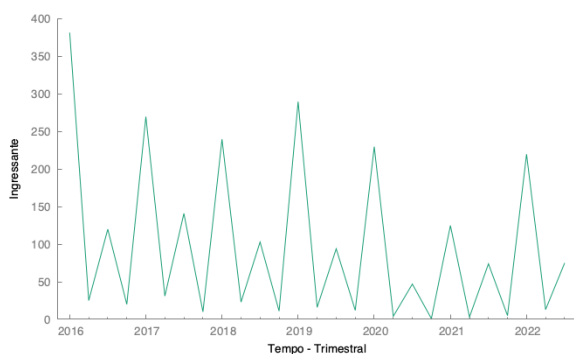


Figura 7: Ingressante filial 1

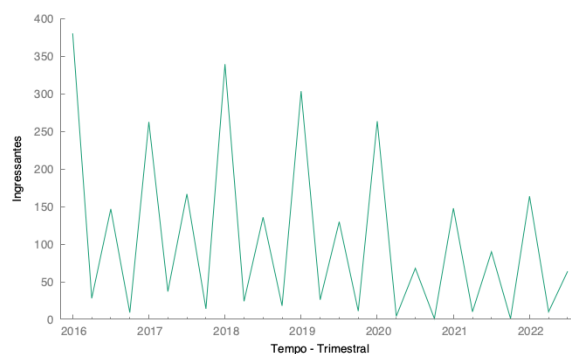


Figura 8: Ingressante filial 2

As duas filiais apresentam comportamentos sazonais similares. A filial 1 apresenta o máximo no primeiro trimestre de 2019, enquanto a filial 2 apresenta o máximo no primeiro trimestre de 2018. Com um comportamento proporcional ao ingressante geral.

## 2.2 Métodos

Esta seção é responsável por apresentar os métodos de série temporais, MQO e teste de Chow.

### 2.2.1 Série Temporal

Série temporal é uma sequência de observações no mesmo espaçamento de tempo dependentes sendo do interesse modelar sua correlatividade a fim de prever valores futuros em determinado período (BOX G. E. P.; REINSEL 1994).

A análise de uma série temporal tem como objetivo entender o seu processo gerador, a previsão dos dados futuros, além de compreender o comportamento de sua trajetória retirar e, conseqüentemente, tomar decisões apropriadas e reduzir ao máximo os erros de previsão. Para modelagem e previsão de séries temporais, é necessário definir: Processos Estocásticos; Processos Estacionários; Processo Ruído Branco; Teste de Raiz Unitária.

#### a) Processos estocásticos

Um processo é um modelo matemático para estudo de conjuntos (sequências) de variáveis aleatórias  $X = \{X(t) : t \in T\}$ , definidas em um mesmo espaço de probabilidade indexadas por um conjunto T.

Um no processo estocástico pode ser definido como estacionário e não estacionário.

#### b) Processos estacionários

Um processo estocástico é dito estacionário quando sua distribuição de probabilidade possui uma variabilidade constante ao longo de um período de tempo pré-definido em torno da média de modo que a escolha de uma origem não seja importante.

Um processo estocástico é **estritamente estacionário** se não houver mudança na sua estruturação ao longo do tempo. Dessa forma:

$$F(Z_1, \dots, Z_n; t_1 + \tau, \dots, t_n + \tau) = F(Z_1, \dots, Z_n; t_1, \dots, t_n) \quad \forall t_1, \dots, t_n \text{ e } \tau \in T \quad (2.1)$$

Um processo estocástico é **francamente estacionário** ou estacionário de segunda ordem se e somente se:

$$E(Z(t)) = \mu(t) = \mu \quad \forall t \in T \quad (2.2)$$

$$E(Z^2(t)) < \infty \quad \forall t \in T \quad (2.3)$$

$$\gamma(t_1, t_2) = Cov(Z(t_1); Z(t_2)) \rightarrow |t_1 - t_2| \quad (2.4)$$

Será considerada a estacionaridade de segunda ordem neste trabalho

### c) Ruído branco

É uma conjunto de variáveis aleatórias não correlacionadas  $\varepsilon_t$ , a cada momento  $t$ , com média 0 e variância finita  $\sigma_t^2$ . A série temporal gerada a partir dessas variáveis não correlacionadas é utilizada como modelo para ruído, denominado ruído branco.

Um ruído branco possui as seguintes características.

Média zero  $\rightarrow E(\varepsilon_t) = 0$

Homocedasticidade  $\rightarrow Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2$

Não - auto correlação  $\rightarrow Cov(\varepsilon_t, \varepsilon_j) = E(\varepsilon_t, \varepsilon_j) = 0$  para  $t \neq j$

Sendo as três propriedades independente e identicamente distribuída (i.i.d)

### d) Teste da Raiz Unitária

Para verificar se a série é estacionária, verifica-se os resultados do teste de raiz unitária. para a verificação de estacionaridade, podemos destacar o teste da raiz unitária, utiliza-se teste Dickey e Fuller aumentado (DFA). O teste consiste em estimar:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad (2.5)$$

Onde:

- $\Delta Y_t$  : número de termos de diferenças defasados a serem incluídos
- $\beta_1$  é intercepto
- $\beta_2$  é coeficiente de tendência
- $\delta$ : é o coeficiente de presença da raiz unitária
- m é o número de defasagens tomadas na série

em que t é a variável de tendência ou temporal. Em cada caso, as hipóteses são:

Hipótese nula:  $H_0$ :  $\delta = 0$  ou  $p = 1$ , há uma raiz unitária ou a série temporal é não estacionária, ou ela possui uma tendência estocástica

Hipótese alternativa:  $H_1$ :  $\delta \neq 0$  ou  $p \neq 1$ , ou ainda  $p \geq 1$  a série temporal é estacionária, possivelmente em torno de uma tendência determinística.

O Gretl apresenta o p-valor, calculado a partir da estatística do teste  $\tau$ . Se o p-valor for maior que 0,05, não rejeita-se a hipótese nula, ao nível de significância de 5%, então a série possui uma raiz unitária.

### 2.2.2 Modelo de Regressão Linear

Esta seção é responsável para apresentar o método Modelo de Regressão Linear.

#### I) Especificação

A teoria de regressão linear múltipla é melhor visualizada considerando o modelo em sua forma matricial

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad (2.6)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1p-1} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2p-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{np-1} \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.7)$$



Por sua vez:

$$\underline{y}_{nx1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X_{nxp} = \begin{bmatrix} 1 & t_{11} & D_{t12} & D_{t13} \\ 1 & t_{21} & D_{t22} & D_{t23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{n1} & D_{tn2} & D_{tn3} \end{bmatrix}, \underline{\beta}_{-px1} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix},$$

$$\underline{\varepsilon}_{nx1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Onde:

Antes da pandemia: Modelo -  $Y_t = \alpha_1 + \alpha_2 t$

Durante a pandemia: Modelo -  $Y_t = (\alpha_1 + \alpha_3) + (\alpha_2 + \alpha_4)t$

-  $Y_t$ : é o valor da variável resposta para a t-ésima observação, onde  $t = 1, 2, \dots, n$ ;

-  $n$  é o tamanho da amostra

-  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_k$  são parâmetros desconhecidos;

-  $\varepsilon_t$ : é o t-ésimo termo de erro aleatório com média 0 e variância  $\sigma^2$

## II) Métodos Mínimos Quadrados Ordinários (MQO)

Para determinar os estimadores de MQO  $\beta_0, \beta_1, \beta_2, \alpha_j, j = 1, 2, \dots, p-1$ , é preciso minimizar a soma dos quadrados de erros (SQE), dada por:

$$SQE = \sum_{i=1}^n (\varepsilon_i)^2 = \varepsilon_1^2 + \dots + \varepsilon_n^2 = \underline{\varepsilon}'\underline{\varepsilon} \quad (2.8)$$

Ou ainda,

$$\begin{aligned} SQE = \underline{\varepsilon}'\underline{\varepsilon} &= (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \\ &= (\underline{y}' - X'\underline{\beta}')(\underline{y} - X\underline{\beta}) \\ &= \underline{y}'\underline{y} - \underline{y}'X\underline{\beta} - \underline{\beta}'X'\underline{y} + \underline{\beta}'X'X\underline{\beta} \end{aligned} \quad (2.9)$$

Como:  $\underline{y}'X\underline{\beta}'$  e  $\underline{\beta}'X'\underline{y}$  São escalares e,  $\underline{y}'X\underline{\beta} = (\underline{\beta}'\underline{y}'\underline{y})'$  então:  $\underline{y}'X\underline{\beta} = \underline{\beta}'X'\underline{y}$

Com isso demonstra-se:

$$SQE(\underline{\beta}) = \underline{y}'\underline{y} - 2\underline{\beta}'X'\underline{y} + \underline{\beta}'X'X\underline{\beta} \quad (2.10)$$

Derivando a função 2.10 e igualando a zero, com objetivo de minimizar os erros:

$$\frac{dSQE(\underline{\beta})}{d\underline{\beta}} = -2X'\underline{y} + 2X'X\underline{\beta} = 0 \quad (2.11)$$

encontrando o vetor de mínimos quadrados do vetor de parâmetro de interesse:

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y} \quad (2.12)$$

Para obter um modelo bem ajustado, precisa-se seguir uma distribuição normal e os resíduos com média em torno de zero.

O métodos matriciais permite desenvolver formulas para a covariância entre qualquer um dos elementos de  $\hat{\underline{\beta}}$ . Afim de obter a realização da inferência estatística precisamos dessas variâncias e covariâncias.

Matriz de variância-covariância de  $\hat{\underline{\beta}}$  é

$$var - cov(\hat{\underline{\beta}}) = E\{[\hat{\underline{\beta}} - E(\hat{\underline{\beta}})][\hat{\underline{\beta}} - E(\hat{\underline{\beta}})]'\} \quad (2.13)$$

Podendo ser obtida da seguinte fórmula:

$$var - cov(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} \quad (2.14)$$

Onde  $\sigma^2$  corresponde à variância homocedástica de  $\underline{\varepsilon}$  e  $(X'X)^{-1}$  a inversa da matriz da equação 2.12 e  $\hat{\underline{\beta}}$  corresponde ao estimador MQO.

### III) Teste de Hipotese da Quebra Estrutural

Uma mudança estrutural ocorre quando os valores do modelo recebe um comportamento de oscilação durante todo intervalo. Podendo ser observado nas regressões de séries temporais, SILVA (2016). Tais mudanças podem resultar por causas externas, como por exemplo crise do petróleo, guerras, pandemia ou por mudanças internas no país, como mudanças de governo e ou políticas econômicas. Para exemplificar a ocorrência ou não de uma mudança estrutural, vamos utilizar os dados coletados na escola de línguas referente a captação de matrículas ativas em dois períodos 2017-2021 (antes e durante a pandemia) e 2020 a 2022 (durante e após a pandemia) nas filiais 1 e 2 em duas clientelas: criança/adolescente e adulto, aplica-se o teste de Chow, GUJARATI (2011), com intuito de verificar a mudança dos parâmetros ao longo do tempo, o teste de Chow observa no modelo se o intercepto e os coeficientes angulares permanecem constante no intervalo de tempo. Podendo assim sinalizar se houve uma mudança nos parâmetros durante a Covid-19.

Neste estudo, a primeira quebra estrutural foi causada pela pandemia, deixando grande parte das atividades a distância ocorrida em 2020. E na segunda quebra após causada após a redução do Covid-19, onde as atividades presenciais começaram a retornar gradativamente.

Para realização da quebra estrutural será utilizado a *dummy*  $D_1$  para a primeira quebra estrutural e posteriormente na segunda quebra estrutural será utilizada a *variável dummy*  $D_2$ . As variáveis são descritas abaixo:

$D_1$  = variável binária assumindo 1 se os meses t forem durante a pandemia e 0 caso contrário.

$D_2$  = variável binária assumindo 1 se os meses t forem após a pandemia e 0 caso contrário.

Tabela 2: Variável binária  $D_1$

Período	D1	D2
2017 a 2019	0	–
2020 a 2021	1	0
2022	–	1

### 2.2.3 Teste de Chow

O teste de Chow verifica se há mudanças significativas entre dois eventos entre as variáveis independentes e a variável dependente baseando-se na comparação da soma dos

resíduos quadrados para justificar a mudança estrutural a ser analisada. O teste de Chow resume-se em:

Dividir a amostra em dois subconjuntos: um antes e outro depois do ponto onde ocorrerá a quebra estrutural.

Onde:

Função antes do evento: Modelo:  $Y_t = \lambda_1 + \lambda_2 t + \varepsilon_{1t}$ ,  $n_1 = u$

Função durante o evento: Modelo:  $Y_t = \gamma_1 + \gamma_2 t + \varepsilon_{2t}$ ,  $n_2 = v$

Função com todo o intervalo: Modelo:  $Y_t = \alpha_1 + \alpha_2 t + \varepsilon_t$ ,  $n = (n_1 + n_2) = u + v$

A regressão com todo intervalo pressupõe que não há diferença entre os dois períodos e estima a relação entre a variável dependente e variável independente para todo o período, que consiste no intervalo de tempo observado. Ou seja essa regressão considera que o intercepto e o coeficiente angular da regressão permanecem os mesmos durante todo o período, ou seja, não se verifica mudança estrutural. Se esta for, de fato, a situação, então  $\alpha_1 = \lambda_1 = \gamma_1$  e  $\alpha_2 = \lambda_2 = \gamma_2$ .

As regressões antes do evento e depois do evento pressupõem que as regressões dos dois períodos sejam diferentes; o intercepto e os coeficientes angulares diferem, como indicado pelos parâmetros com subscritos. Nas regressões, os  $\varepsilon$  representam os termos de erro e os  $n$ , o número de observações.

Ao rodar as Funções anteriores em software estatístico, obtém-se a soma do quadrado dos resíduos. Podendo assim calcular o teste de F para verificar se há diferenças significativas entre os eventos.

$$SQR_{SR} = SQR_1 + SQR_2 \text{ com grau de liberdade } = (n_1 + n_2 - 2k)$$

Aplicando o teste de Chow tem-se:

$$F = \frac{(SQR_R - SQR_{SR})/k}{(SQR_{SR})/(n_1 + n_2 - 2K)} \quad (2.15)$$

Onde:

- $SQR_1$  : Soma do quadrado dos resíduos antes do evento
- $SQR_2$  : Soma do quadrado dos resíduos depois do evento
- $SQR_R$  : Soma do quadrado dos resíduos em todo período

- $k$ : número de regressores
- $n$  : tamanho da amostra

Obtendo o valor de F compara-se com o valor crítico da tabela F, caso o valor de F calculado não for superior ao valor de F crítico no nível de significância de 5% conclui-se que não rejeitamos a hipótese nula de estabilidade dos parâmetros (ausência de mudança estrutural). Caso contrário, se o valor de F calculado for superior ao valor de F crítico, rejeitamos a hipótese de estabilidade dos parâmetros e concluímos que as regressões antes do evento e depois do evento são diferentes.

A regressão do modelo com a quebra estrutural pode ser usada para testar diversas hipóteses. Se o coeficiente do intercepto diferencial  $\alpha_3$  for não significativo estatisticamente, podemos aceitar a hipótese de que as duas regressões têm o mesmo intercepto. Da mesma forma, se o coeficiente diferencial angular  $\alpha_4$  for não significativo estatisticamente, mas  $\alpha_3$  for significativo, podemos não rejeitar a hipótese de que as duas regressões têm a mesma inclinação; as duas retas de regressão são paralelas. O teste da estabilidade de toda a regressão ( $\alpha_3 = \alpha_4 = 0$ , simultaneamente) pode ser feito pelo teste F. Se essa hipótese não for rejeitada, as retas de regressão serão coincidentes.

Um dos modelos utilizados para entender o comportamento de séries temporais, cujas observações ( $t$ ) nos intervalos são independentes, são os modelos da classe autoregressiva, integrada e de médias móveis, também conhecidos como Modelos ARIMA.

## 2.2.4 Modelos ARIMA

Os modelos da classe ARIMA, desenvolvidos por Box & Jenkins, serão utilizados neste trabalho na análise e previsão de valores futuros para os trimestres *Inter I*, *Inter II*, *Intra I*, *Intra II*. Para análise e estimação dos parâmetros dos modelos ARIMA são necessários seguir alguns passos. A saber: Especificação; Identificação; Estimação e; diagnóstico do modelo.

### I) Especificação

#### Processo Auto-regressivo de ordem $p$ - AR( $p$ )

Seja  $Z_t$  o número de ingressantes no instante  $t$ , o Modelo AR( $p$ ) é definido pela equação (2.17).

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \alpha_t \quad (2.16)$$

Onde  $Z_t$  é uma série temporal estacionária,  $\phi_1 \dots \phi_p$  são parâmetros da estrutura autorregressiva e  $\alpha_t$  ruído branco.

### Processo de Média Móvel de ordem q - MA(q)

para o processo MA(q), o modelo está definido na equação (2.18)

$$Z_t = \alpha_t + \theta_1 \alpha_{t-1} + \dots + \theta_q \alpha_{t-q} \quad (2.17)$$

Onde  $\alpha_1, \dots, \alpha_q$  são valores defasados da série de erros,  $\theta_i$  são os parâmetros do modelo para  $i = 1, 2, 3, \dots, q$

### Processo auto-regressivo e de média móvel - ARMA(p,q)

A série temporal  $z_t$  possui características AR e MA, sendo assim ARMA (p, q):

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \alpha_t + \theta_1 \alpha_{t-1} + \theta_2 \alpha_{t-2} + \dots + \theta_q \alpha_{t-q} \quad (2.18)$$

Considerando que em uma série não estacionária é necessário fazer d diferenças para torná-la estacionária, o modelo a ser estimado é o Modelo ARIMA(p,d,q).

### Processo Autorregressivo, Integrado e de Médias Móveis - ARIMA (p,d,q)

Obtendo a série estacionária, obtem-se a forma ARIMA (p,d,q) em termos de diferença  $Z_t$  ( $\Delta Z_t$ ) representada por:

$$\Delta^d Z_t = c + \phi_1 \Delta^d Z_{t-1} + \dots + \phi_p \Delta^d z_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q} \quad (2.19)$$

Onde  $\Delta^d$  representa o operador de diferença de  $Z_t - Z_{t-1}$ .

### Função de autocorrelação (FAC)

Utiliza-se como procedimento para medir a autocorrelação contra a defasagem, dada por:

$$p_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^n (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \quad (2.20)$$

### Função de autocorrelação Parcial (FACP)

Mede-se a correlação entre duas observações de  $Z_t$  e  $Z_{t-k}$  depois de eliminada a influência de  $Z_{t-1}, \dots, Z_{t-k+1}$ , ou seja:

$$p_k = \text{Corr}(Z_t, Z_{t-k} | Z_{t-1}, \dots, Z_{t-(k+1)}) \quad (2.21)$$

### Metodologia Box & Jenkins

A metodologia Box & Jenkins para previsão, consiste em ajustar modelos ARIMA (p,d,q) a séries temporais observadas, onde dado os valores da amostra no presente e passado, obtem-se os valores futuros previstos. Criado em 1976 pelo George Box e Gwilym Jenkins, se tornou o modelo mais utilizado para previsão a curto prazo, possuindo três etapas:

- Identificação: Identifica-se os valores apropriados dos parâmetros p, d e q através da análise de correlograma e correlograma parcial;
- Diagnostico do modelo: Deve-se estimar e testar os parâmetros dos termos autorregressivos e da média móvel para obter coeficientes que representam o melhor modelo selecionado.
- Previsão: Realização da previsão dos valores futuros da série.

### II) Identificação

Segundo Bueno (2012) A FAC define defasagem da componente Média Móvel (MA). Já no FACP define a defasagem da componente autoregressivo (AR). No primeiro caso a função de autocorrelação decai exponencialmente com aumento de defasagem e a função de autocorrelação parcial é truncada a partir da defasagem p. Para o próximo caso ocorre o inverso. A função de autocorrelação é truncado na defasagem q e a função de autocorrelação parcial decai exponencialmente. No caso ARMA, as duas funções decaem exponencialmente a partir da defasagem de truncagem. A FAC decai a partir da defasagem q e a FACP decai a partir da defasagem p. A tabela 3 resume as os critérios acima.

Tabela 3: Padrões Teóricos de FAC e FACP

Modelo	FAC	FACP
AR	Decaimento exponencial	pico forte na defasagem p
MA	Pico forte da defasagem q Decaimento exponencial	Decai exponencialmente Decaimento exponencial
ARMA	após a defasagem q	após a defasagem p

### Análise dos Resíduos

Utilizando o FAC obtem-se a análise de resíduos para verificação dos dados. A correlação apresentando ausência pode-se supor que a amostra são independentes e identicamente distribuídos com variância constante e média zero. Para um bom modelo os resíduos da FAC deveriam estar em torno de zero e com nenhuma ou pouca autocorrelação fora do intervalo de confiança.

### 2.2.5 Estatísticas de Aderência

#### Raiz Quadrada do Erro Quadrático Médio (RMSE)

RMSE analisa o erro na escala do evento a ser estimado, melhor será a qualidade do ajuste próximo de zero.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{n}} \quad (2.22)$$

Onde:  $z_t$  é o valor observado,  $\hat{z}_t$  é o valor ajustado,  $n$  é o tamanho da série.

#### Critério de informação Akaike (AIC)

O critério de informação de Akaike (AIC), proposto por AKAIKE (1974), usado para estimar esta dupla esperança a partir da máxima log-verossimilhança negativa.

$$AIC = 2\ln(Lp) + 2[(p + 1) + 1] \quad (2.23)$$

Onde,  $Lp$  é a log-verossimilhança máxima e  $p$  é o número de parâmetros do modelo.

#### Desvio Médio Absoluto (MAD)



O MAD (sigla do inglês Mean Absolute Deviation ) definido como:

$$MAD = \frac{\sum_{i=1}^n |z_t - \hat{z}_t|}{n} \quad (2.24)$$

Onde:  $z_t$  é o valor observado,  $\hat{z}_t$  é o valor ajustado,  $n$  é o tamanho da série

### **Erro Percentual Médio Absoluto (MAPE)**

O MAPE sigla do inglês Mean Absolute Percentage Error, com objetivo de medir o tamanho do desvio do valor real do modelo estatístico em porcentagem, definido como:

$$MAPE = \frac{\sum_{i=1}^m |z_t - \hat{z}_t|}{|z_t|} \times \frac{1}{n} \quad (2.25)$$

$z_t$  é o valor observado,  $\hat{z}_t$  é o valor ajustado,  $n$  é o tamanho da série

## **2.2.6 Proposta**

A proposta deste trabalho consiste em gerar modelos de séries temporais para escolha do melhor modelo com base da estatística de aderência e análise de resíduos para previsão de matrículas em cada ciclo em duas filiais e com duas clientelas específicas. Como também analisar a quebra estrutural durante a COVID-19.

## 3 Resultados

Este capítulo é responsável por apresentar as análises estatísticas atendendo aos objetivos propostos comparando o número de matrículas entre os anos 2017 a 2022.

### 3.1 Quebra estrutural

Neste trabalho, dois modelos econométricos serão utilizados: o primeiro tem por objetivo analisar se houve quebra estrutural da matrícula mensal antes e durante a pandemia de covid-19. Já o segundo modelo tem por objetivo analisar se houve quebra estrutural da matrícula mensal durante e após a pandemia de covid-19.

Seja  $Y_t$  o número de matrículas ativas no mês  $t$ .

Modelo sem quebra:  $Y_t = \alpha_1 + \alpha_2 t + \varepsilon_t$

Modelo com a quebra:  $Y_t = \alpha_1 + \alpha_2 t + \alpha_3 D_t + \alpha_4 (D_t t) + \varepsilon_t$

Onde:  $D$  representa as variáveis binárias  $D_1$  e  $D_2$ , explicadas anteriormente.

### 3.2 Quebra estrutural, 2017 a 2021, antes e durante a pandemia

Inicialmente pressupõe-se que não há mudança estrutural ou diferenças nos dois períodos, entre 2017 a 2019 ( $n=36$ ) e 2020 a 2021 ( $n=24$ ). A atenção será dada ao coeficiente angular que representa a variação, em média, das matrículas ativas por mês.

#### 3.2.1 Filial 1: criança e adolescente

Obtendo o modelo sem quebra estrutural na filial 1 na clientela *criança e adolescente* usando observações de 2017 a 2021 ( $n=60$ ) com a Matrícula ativa como variável

dependente. A Tabela 4 apresenta o resultado da estimação do modelo sem quebra estrutural e Tabela 5 as estatísticas de aderência.

Tabela 4: Resultados do modelo sem quebra - Filial 1 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	26.515,0	465,580	56,95	1,32e-052 ***
tempo mensal	-275,855	13,2743	-20,78	1,49e-028 ***

Tabela 5: Resultados da estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,879556	1,49e-28	1.070,409

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 276 alunos, por mês, no período 2017 a 2021, para a clientela *crianças e adolescentes*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,8796, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão no Tabela 6.

Tabela 6: Resultados do modelo com quebra - Filial 1 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	27.324,4	581,608	46,98	1,12e-46 ***
tempo mensal	-323,049	27,4122	-11,78	8,55e-17 ***
$D_1$	-5.545,02	2.535,97	-2,187	0,0330 **
$D_1t$	146,806	57,3581	2,559	0,0132 **

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será:

$$\hat{Y}_t = 27.324,4 - 323,049t \quad (3.1)$$

Indicando que, antes da pandemia, o número de matrículas mensais de crianças e adolescentes na Filial 1 reduziu a uma média de 323 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será:

Tabela 7: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,889111	2,40e-27	1.067,344

$$\hat{Y}_t = (2.734,4 - 5.545,02) + (-323,049 + 146,806)t \quad (3.2)$$

$$\hat{Y}_t = 21.779,4 - 176,243t$$

Indicando que, durante a pandemia, a redução das matrículas foi menor, com uma média de  $-176$  matrículas por mês. O  $R^2$  ajustado foi de 0,889111, maior que o modelo sem quebra estrutural, indicando que 88,9% da variação das matrículas mensais são explicativas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está representada na Figura 9 e com quebra estrutural na Figura 10. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 1 - criança e adolescentes.

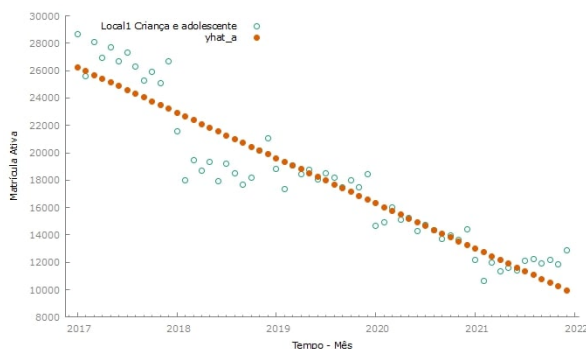


Figura 9: Modelo sem quebra estrutural

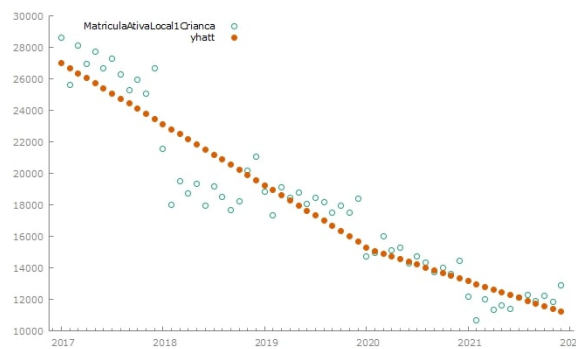


Figura 10: Modelo com quebra estrutural

Para verificar se ocorreu, de fato, uma mudança estrutural nas matrículas durante a pandemia, em comparação com antes da pandemia, Chow testa a hipótese nula de que  $\alpha_1 = \alpha_2$  e  $\beta_1 = \beta_2$ , ou seja, as regressões dos subperíodos não são diferentes, são estáveis, e não existe a quebra estrutural. A estatística do teste F encontrada foi igual a 3,49897, com um p-valor de 0,0370. Utilizando um nível de significância de 5% rejeita-se a hipótese nula de estabilidade dos parâmetros, tendo evidência de houve mudança estrutural nas matrículas de *crianças e adolescentes* na Filial 1 entre os períodos analisados.

### 3.2.2 Filial 1, Adulto

Esta seção repete a análise anterior para a clientela *adulta* da Filial 1. A Tabela 8 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 8: Resultados do modelo sem quebra - Filial 1 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	5.687,20	114,041	49,87	2,50e-049 ***
tempo mensal	-86,4240	3,25147	-26,58	3,60e-034 ***

Tabela 9: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,922825	3,60e-34	901,6023

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 86 alunos, por mês, no período 2017 – 2021, para a clientela *adulta*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,9223, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 10.

Tabela 10: Resultados do modelo com quebra - Filial 1 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	5.855,10	146,061	40,09	6,28e-43 ***
tempo mensal	-97,9710	6,88410	-14,23	2,99e-20 ***
$D_1$	90,9536	636,865	0,1428	0,8869
$D_1t$	7,62365	14,4045	0,5293	0,5987

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 5855,10 - 97,9710t$ , indicando que, antes da pandemia, o número de matrículas mensais de *adulto* na Filial 1 reduziu a uma média de 98 por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = 5946,06 - 90,3474t$ , indicando que durante a pandemia, a redução das matrículas foi menor, com uma média de -90 matrículas por mês. O  $R^2$  ajustado foi de 0,925312, maior do que o modelo sem quebra estrutural, indicando que 92,5% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

Tabela 11: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,925312	3,83e-32	901,5315

As Figuras 11 e 12 apresentam os gráficos com as retas estimadas nos diferentes períodos e modelos. Observe a diferença entre as inclinações das retas no modelo com a quebra estrutural (Figura 12) é pequena, visualmente

Valores observados e ajustados da variável matrículas ativas - Filial 1 - Adultos

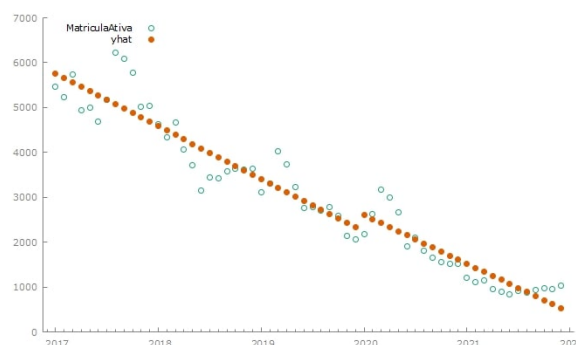
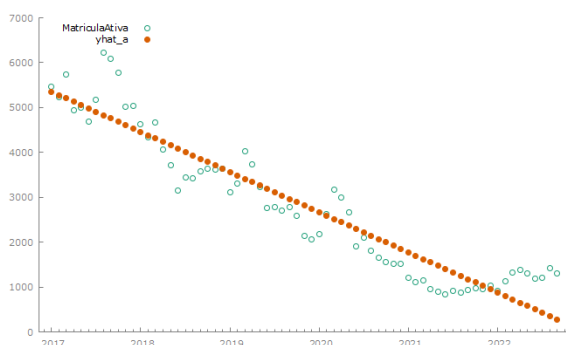


Figura 11: Modelo sem quebra estrutural      Figura 12: Modelo com a queda estrutural

A estatística do teste F encontrada foi igual a 1,96563, com um p-valor de 0,1496. Utilizando um nível de significância de 5%, não rejeita-se a hipótese nula de estabilidade dos parâmetros, não há evidência suficiente de mudança estrutural nas matrículas de *adultos* na Filial 1 entre os períodos analisados.

### 3.2.3 Filial 2, Criança e adolescente

Esta seção repete a análise das seções anteriores para a Filial 2 da *clientela criança e adolescente*. A Tabela 12 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 12: Resultados do modelo sem quebra - Filial 2 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	21.851,1	292,151	74,79	2,22e-059 ***
tempo mensal	-161,008	8,32962	-19,33	5,88e-027 ***

Tabela 13: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,863310	5,88e-27	1.014,488

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 161 alunos, por mês, no período 2017 - 2021, para a *clientela crianças e adolescentes*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,8633, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 14.

Tabela 14: Resultados do modelo com quebra - Filial 2 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	21.569,0	361,394	59,68	2,19e-052 ***
tempo mensal	-148,211	17,0331	-8,701	5,50e-012 ***
$D_1$	4.513,38	1.575,78	2,864	0,0059 ***
$D_1t$	-98,6366	35,6406	-2,768	0,0076 ***

Tabela 15: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,882875	4,76e-26	1.010,245

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 21569 - 148,211t$ , indicando que, antes da pandemia, o número de matrículas mensais de *crianças e adolescentes* na Filial 2 reduziu a uma média de 148 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = 26082,4 - 161,008t$ , indicando que, durante a pandemia, a redução das matrículas foi maior, com uma média de -161 matrículas por mês. O  $R^2$  ajustado foi de 0,876601 maior do que o modelo sem quebra estrutural, indicando que 87,6% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está na Figura 13 e com quebra estrutural na Figura 14. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 2 - Criança e Adolescente

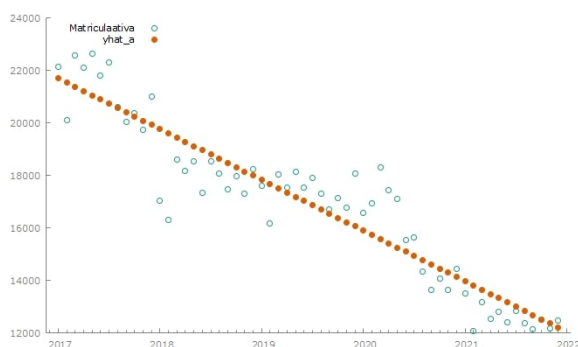


Figura 13: Modelo sem quebra estrutural

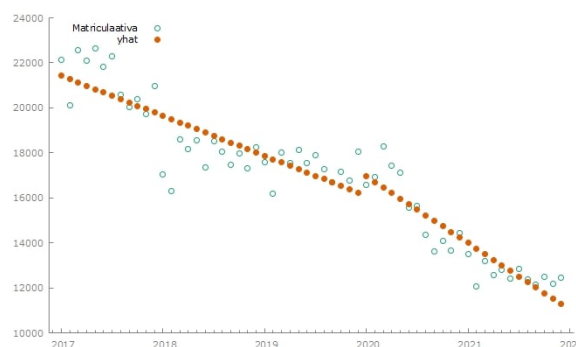


Figura 14: Modelo com quebra estrutural

A estatística do teste F encontrada foi igual a 4,12332, com um p-valor de 0,0214. Utilizando um nível de significância de 5%, rejeita-se a hipótese nula de estabilidade dos parâmetros, há evidência suficiente de mudança estrutural nas matrículas de *criança e adolescente* na Filial 2 entre os períodos analisados.

### 3.2.4 Filial 2: Adulto

Esta seção repete a análise anterior para a clientela *adulta* da Filial 2. A Tabela 16 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 16: Resultados do modelo sem quebra - Filial 2 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	5.786,34	105,124	55,04	9,16e-052
tempo mensal	-52,9577	2,99722	-17,67	5,12e-025

Tabela 17: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,840624	5,12e-25	893,4699

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 53 alunos, por mês, no período 2017 – 2021, para a clientela *adulta*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,840624, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 18.



Tabela 18: Resultados do modelo com quebra - Filial 2 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	5.987,94	128,602	46,56	1,82e-046
tempo mensal	-67,3964	6,06122	-11,12	8,51e-016
$D_1$	515,010	560,739	0,9184	0,3623
$D_1t$	1,68944	12,6827	0,1332	0,8945

Tabela 19: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,859283	1,87e-24	886,2551

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 5987,94 - 67,3964t$ , indicando que, antes da pandemia, o número de matrículas mensais de *adulto* na Filial 2 reduziu a uma média de 67 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = 6502,95 - 65,7070t$ , indicando que, durante a pandemia, a redução das matrículas foi menor, com uma média de -66 matrículas por mês. O  $R^2$  ajustado foi de 0,859283 maior do que o modelo sem quebra estrutural, indicando que 85,9% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está representada na Figura 15 e com quebra estrutural na Figura 16. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 2 - Adultos

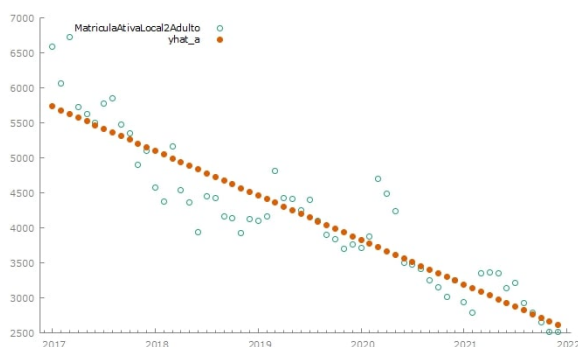


Figura 15: Modelo sem quebra estrutural

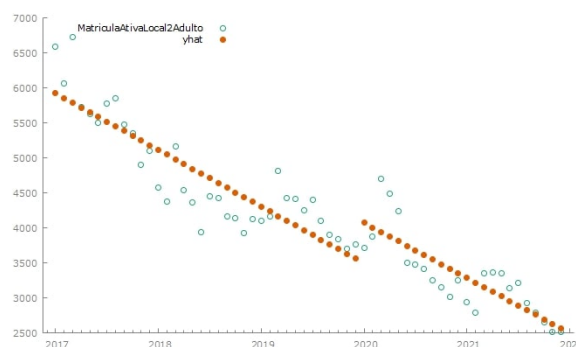


Figura 16: Modelo com quebra estrutural

A estatística do teste F encontrada foi igual a 4,84538, com um p-valor de 0,0115. Utilizando um nível de significância de 5%, rejeita-se a hipótese nula de estabilidade dos parâmetros, há evidência suficiente de mudança estrutural nas matrículas de *adultos* na Filial 2 entre os períodos analisados.

### Considerações finais da análise da quebra estrutural antes e durante da pandemia

A Tabela 20 resume a análise da quebra estrutural das matrículas nos períodos antes e durante a pandemia e também traz a informação se a tendência das matrículas era crescente, decrescente ou constante.

Tabela 20: Resumo da análise da quebra estrutural antes e durante a pandemia

<b>População analisada</b>	<b>Houve quebra estrutural antes e durante a pandemia?</b>
Filial 1, criança e adolescente	Sim, com tendência decrescente nos dois períodos
Filial 1, adulto	Não, com tendência decrescente nos dois períodos
filial 2, criança e adolscente	Sim, com tendência decrescente nos dois períodos
filial 2, adulto	Sim, com tendência decrescente nos dois períodos

Pode-se concluir que as matrículas médias mensais foram impactadas pela pandemia, em comparação com o período anterior, em todas as filiais e clientelas, exceto na Filial 1 para as clientelas *adultas*, que se manteve estatisticamente igual.

### 3.3 Quebra estrutural, 2020 a 2022: durante e após a pandemia

O objetivo desta seção é analisar se, no ano de 2022, considerando aqui como após a pandemia ou o período em que muitas pessoas já haviam sido vacinadas e o número de óbitos diretos pela doença estava muito menor, as matrículas tiveram comportamento igual ou diferente ao período imediatamente anterior, durante a pandemia. Em resumo, o objetivo é verificar se em 2022 houve um comportamento diferente das matrículas em relação aos dois primeiros anos da pandemia (2020 e 2021). A análise também será feita segundo a filial e clientela.

Inicialmente pressupõe que não há mudança estrutural ou diferenças nos dois períodos, entre 2020-2021 ( $n=24$ ) e 2022 ( $n=9$ ). A atenção será dada ao coeficiente angular que representa a variação, em média, das matrículas ativas por mês.

### 3.3.1 Filial 1: Criança e adolescente

A Tabela 21 apresenta o resultado da estimação do modelo sem quebra estrutural na clientela *Criança e adolescente* no Filial 1.

Tabela 21: Resultados do modelo sem quebra - Filial 1 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	16.115,4	1.265,08	12,74	7,39e-014
tempo mensal	-53,0421	23,4934	-2,258	0,0311

Tabela 22: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,113510	0,031149	568,0522

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 53 alunos, por mês, no período 2020 – 2022, para a clientela *crianças e adolescentes*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,113510, indicando que este modelo de tendência linear possui baixa capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 23.

Tabela 23: Resultados do modelo com quebra - Filial 1 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	21.779,4	1.186,45	18,36	1,65e-017
tempo mensal	-176,243	24,2175	-7,278	5,15e-08
$D_2$	-24.784,1	6.998,28	-3,541	0,0014
$D_2t$	430,126	108,754	3,955	0,0005

Tabela 24: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,637942	3,52e-07	540,3010

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 21779,4 - 176,243t$ , indicando que, durante a pandemia, o número de matrículas mensais de *crianças e adolescentes* no Filial 1 reduziu a uma média de 176 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será  $\hat{Y}_t = -3004,64 + 253,883t$ , indicando que, após a pandemia, houve um aumento no número das matrículas, com uma média de +254 matrículas por mês. O  $R^2$  ajustado foi

de 0,637942, maior que o modelo sem quebra estrutural, indicando que 63,8% da variação das matrículas mensais são explicativas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está representada na Figura 17 e com quebra estrutural na Figura 18. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 1 - criança e adolescentes.

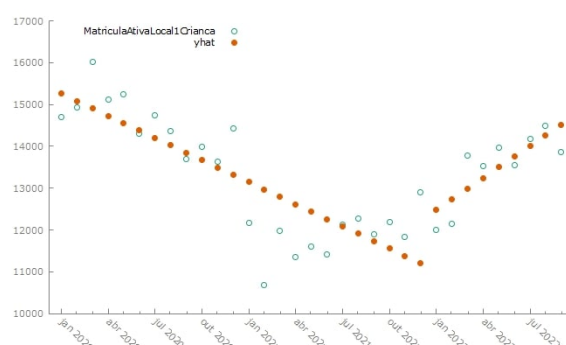
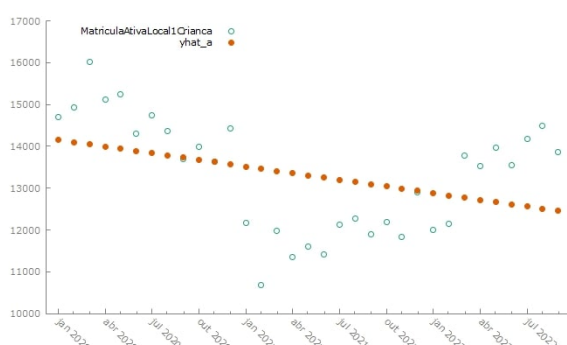


Figura 17: Modelo sem quebra estrutural

Figura 18: Modelo com quebra estrutural

A estatística do teste F encontrada foi igual a 23,4513, com um p-valor de  $8,7306e-7$ . Utilizando um nível de significância de 5% rejeita-se a hipótese nula de estabilidade dos parâmetros, tendo evidência de houve mudança estrutural nas matrículas de crianças e adolescentes na filial 1 entre os períodos analisados.

### 3.3.2 Filial 1: Adulto

Esta seção repete a análise anterior para a clientela *adulto* da filial 1. A Tabela 25 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 25: Resultados do modelo sem quebra - Filial 1 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	3.899,92	450,957	8,648	9,12e-010
tempo mensal	-45,7343	8,37454	-5,461	5,72e-06

Tabela 26: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,473890	5,72e-06	499,9717

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 46 alunos, por mês, no período 2020a2022, para a clientela *adulto*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,473890, indicando que este modelo de tendência linear possui média capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 27.

Tabela 27: Resultados do modelo com quebra - Filial 1 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	5.964,06	434,661	13,68	3,54e-014
tempo mensal	-90,3474	8,87217	-10,18	4,36e-011
$D_2$	-6.864,36	2563,84	-2,677	0,0121
$D_2t$	123,564	39,8425	3,101	0,0043

Tabela 28: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,773039	4,36e-10	474,0267

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 5946,06 - 90,3474t$ , indicando que, durante a pandemia, o número de matrículas mensais de *adulto* na Filial 1 reduziu a uma média de 90 por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = -918,306 + 33,2167t$ , indicando que após a pandemia, houve aumento das matrículas, com uma média de +33 matrículas por mês. O  $R^2$  ajustado foi de 0,773039, maior do que o modelo sem quebra estrutural, indicando que 77,3% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As Figuras 19 e 20 apresentam os gráficos com as retas estimadas nos diferentes períodos e modelos. Observe a diferença entre as inclinações das retas no modelo com a quebra estrutural (Figura 20) é positiva, visualmente:

## Valores observados e ajustados da variável matrículas ativas - Filial 1 - Adultos

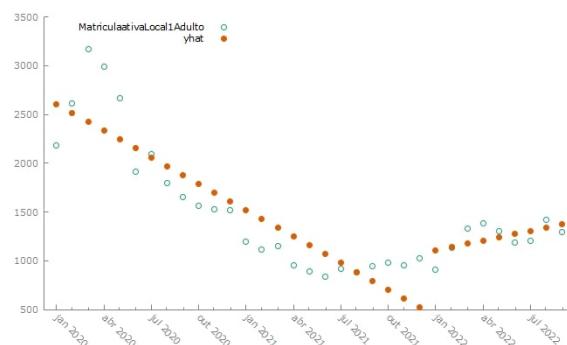
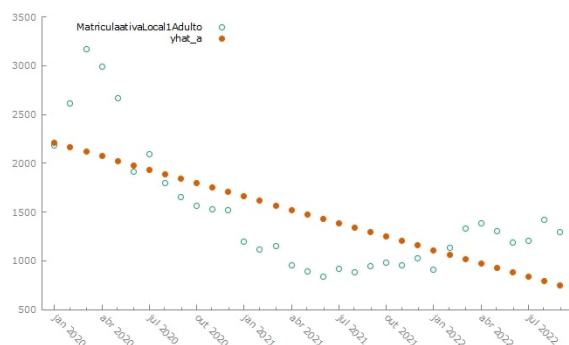


Figura 19: Modelo sem quebra estrutural      Figura 20: Modelo com a quebra estrutural

A estatística do teste F encontrada foi igual a 21,43, com um p-valor de  $1,93065e-06$ . Utilizando um nível de significância de 5%, rejeita-se a hipótese nula de estabilidade dos parâmetros, há evidência suficiente de mudança estrutural nas matrículas de *adultos* na filial 1 entre os períodos analisados.

### 3.3.3 Filial 2: Criança e adolescente

Esta seção repete a análise das seções anteriores para a filial 2 da *clientela criança e adolescente*. A Tabela 29 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 29: Resultados do modelo sem quebra - Filial 2 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	22.209,1	1.022,39	21,72	2,57e-010
tempo mensal	-162,511	18,9863	-8,559	1,15e-09

Tabela 30: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,693084	1,15e-09	553,9942

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 163 alunos, por mês, no período 2020 - 2022, para a *clientela crianças e adolescentes*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,693084, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 31.

Tabela 31: Resultados do modelo com quebra - Filial 2 - Criança e Adolescente

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	26.082,4	1.171,93	22,26	8,89e-020
tempo mensal	-246,848	23,9211	-10,32	3,22e-011
$D_2$	-15.236,7	6.912,61	-2,204	0,0356
$D_2t$	268,064	107,423	2,495	0,0185

Tabela 32: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,812744	2,74e-11	539,4881

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$  for igual a zero, a reta estimada será de  $\hat{Y}_t = 26082 - 246,949t$ , indicando que, durante a pandemia, o número de matrículas mensais de *crianças e adolescentes* na Filial 2 reduziu a uma média de 247 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = 10845,7 + 21,2167t$ , indicando que, após a pandemia, o aumento das matrículas foi maior, com uma média de +21 matrículas por mês. O  $R^2$  ajustado foi de 0,812744 maior do que o modelo sem quebra estrutural, indicando que 81,3% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está na Figura 21 e com quebra estrutural na Figura 22. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 2 - Criança e Adolescente

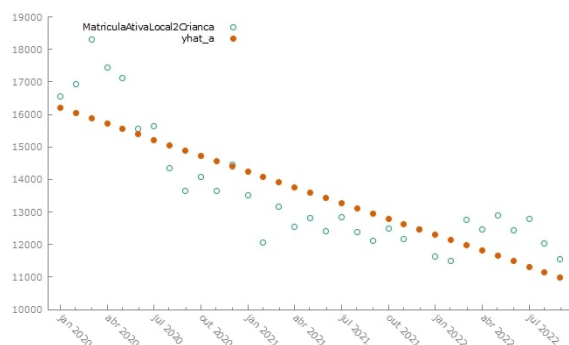


Figura 21: Modelo sem quebra estrutural

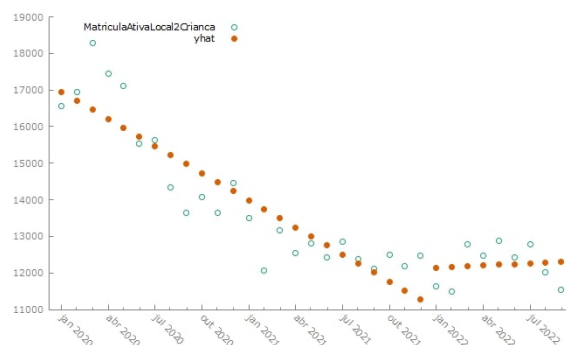


Figura 22: Modelo com quebra estrutural

A estatística do teste F encontrada foi igual a 10,9048, com um p-valor de 0,0003. Utilizando um nível de significância de 5%, rejeita-se a hipótese nula de estabilidade dos parâmetros, há evidência suficiente de mudança estrutural nas matrículas de *criança e adolescente* na filial 2 entre os períodos analisados.

### 3.3.4 Filial 2: Adulto

Esta seção repete a análise anterior para a clientela *adulta* da filial 2. A Tabela 33 apresenta o resultado da estimação do modelo sem quebra estrutural.

Tabela 33: Resultados do modelo sem quebra - Filial 2 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	7.234,17	312,131	23,18	3,85e-021
tempo mensal	-82,8386	5,79645	-14,12	4,80e-015

Tabela 34: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,861074	4,80e-15	475,6870

O coeficiente angular estimado indica que as matrículas reduziram, em média, aproximadamente 82 alunos, por mês, no período 2020 – 2022, para a clientela *adulta*. Ao nível de significância de 5%, os parâmetros estimados foram significativos. O  $R^2$  ajustado foi de 0,861074, indicando que este modelo de tendência linear possui alta capacidade explicativa. Os resultados do modelo estimado com a quebra estrutural estão na Tabela 35.

Tabela 35: Resultados do modelo com quebra - Filial 2 - Adulto

Variável	Coefficiente	Erro padrão	razão-t	p-valor
Constante	6.502,95	427,382	15,22	2,32e-015
tempo mensal	-65,7070	8,72539	-7,532	2,65e-08
$D_2$	-1.258,07	2.520,91	-0,4991	0,6215
$D_2t$	12,3736	39,1753	0,3159	0,7544

Tabela 36: Resultados estatística de aderência

R-quadrado ajustado	P-valor(F)	Critério de Akaike
0,879056	4,99e-14	472,9120

De acordo com estes resultados, tanto os interceptos diferenciais quanto os coeficientes angulares são estatisticamente significativos ao nível de significância de 5%. Quando  $D_1$



for igual a zero, a reta estimada será de  $\hat{Y}_t = 6502,95 - 65,7070t$ , indicando que, durante a pandemia, o número de matrículas mensais de *adulto* na Filial 2 reduziu a uma média de 66 alunos por mês. Quando  $D_1$  for igual a um, a reta estimada será de  $\hat{Y}_t = 5244,89 - 53,333t$ , indicando que, após a pandemia, a redução das matrículas foi menor, com uma média de -53 matrículas por mês. O  $R^2$  ajustado foi de 0,879056 maior do que o modelo sem quebra estrutural, indicando que 87,9% da variação das matrículas mensais são explicadas pelas variáveis do modelo, quando se inclui a variável indicadora de período. O critério de informação de Akaike foi menor, indicando também o melhor ajuste.

As retas estimadas para os dois modelos sem quebra estrutural está representada na Figura 23 e com quebra estrutural na Figura 24. As diferenças dos coeficientes angulares nos períodos podem ser visualizados entre os períodos.

Valores observados e ajustados da variável matrículas ativas - Filial 2 - Adultos

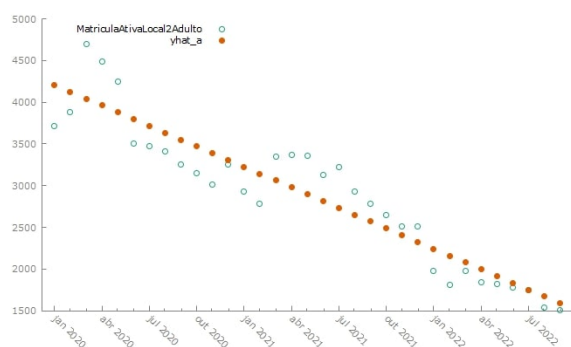


Figura 23: Modelo sem quebra estrutural

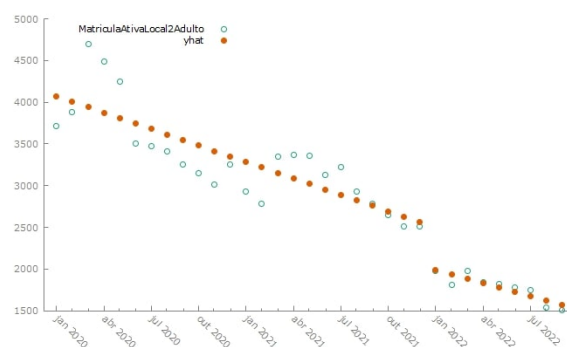


Figura 24: Modelo com quebra estrutural

A estatística do teste F encontrada foi igual a 3,30451, com um p-valor de 0,0510. Utilizando um nível de significância de 5%, não rejeita-se a hipótese nula de estabilidade dos parâmetros, não há evidência suficiente de mudança estrutural nas matrículas de *adultos* na Filial 2 entre os períodos analisados.

### Considerações finais da análise da quebra estrutural durante e após a pandemia

A Tabela 37 resume a análise da quebra estrutural das matrículas nos períodos durante e após a pandemia e também traz a informação se a tendência das matrículas era crescente, decrescente ou constante.

Tabela 37: Resumo da análise da quebra estrutural durante e após a pandemia

População analisada	Houve quebra estrutural durante e após a pandemia?
Filial 1, criança e adolescente	Sim, com tendência decrescente durante e crescente após a pandemia
Filial 1, adulto	Sim, com tendência decrescente durante e crescente após a pandemia
Filial 2, criança e adolescente	Sim, com tendência decrescente durante e levemente crescente após a pandemia
Filial 2, adulto	Não, com tendência decrescente nos dois períodos

Afirma-se que existe uma tendência crescente nas matrículas se o coeficiente de inclinação foi positivo. Segundo Gujarate (2011), entende-se por tendência um movimento sustentado crescente ou decrescente no comportamento de uma variável. Com base nas análises, pode-se concluir que após a pandemia, houve uma recuperação acentuada, decrescente ou da filial 1 para ambas as clientelas e uma leve ou suave recuperação na filial 2 para a clientela *criança e adolescente* em comparação com o período anterior, durante a pandemia.

### 3.4 Previsão, Introdução

A implementação de séries temporais na área de educação pode ser benéfica para prever o número de ingressantes em instituições educacionais. Com a análise de dados históricos, é possível identificar tendências e padrões sazonais que ajudam a prever o número de alunos que ingressarão em uma instituição de ensino em um período futuro.

Essas previsões podem ajudar as instituições de ensino a se preparar melhor para o número de alunos que espera-se ingressar. Implementando um planejamento de alocação de recursos e equipamentos, bem como horários e cronogramas de aulas. Evitando assim a superlotação de salas de aula e garantir que haja professores para atender à demanda. Além disso, essa análise pode ajudar no planejamento de campanhas de marketing para atrair novos alunos. Com a informação de quantos alunos espera-se obter em determinado período.

Nas próximas sessões a variável de interesse é o número de ingressantes por trimestre. Foram estimados modelos ARIMA para previsão do número de ingressantes para o último trimestre de 2022 (*Intra II*) e para o primeiro trimestre de 2023 (*Inter I*), para cada filial, considerando todas as clientelas.

### 3.4.1 Modelagem ARIMA para o Filial 1

#### a) Escolha do melhor modelo de previsão

Para a amostra treino, foram considerados o período do primeiro trimestre de 2016 até o primeiro trimestre de 2021, totalizando 21 observações, que corresponde a 80,7% do total de observações disponíveis até a elaboração desta pesquisa.

Para identificar os possíveis modelos ARIMA a serem ajustados à amostra treino, foram calculadas as funções de autocorrelação e autocorrelação parcial associadas ao ingresso trimestral, apresentadas na Figura 25.

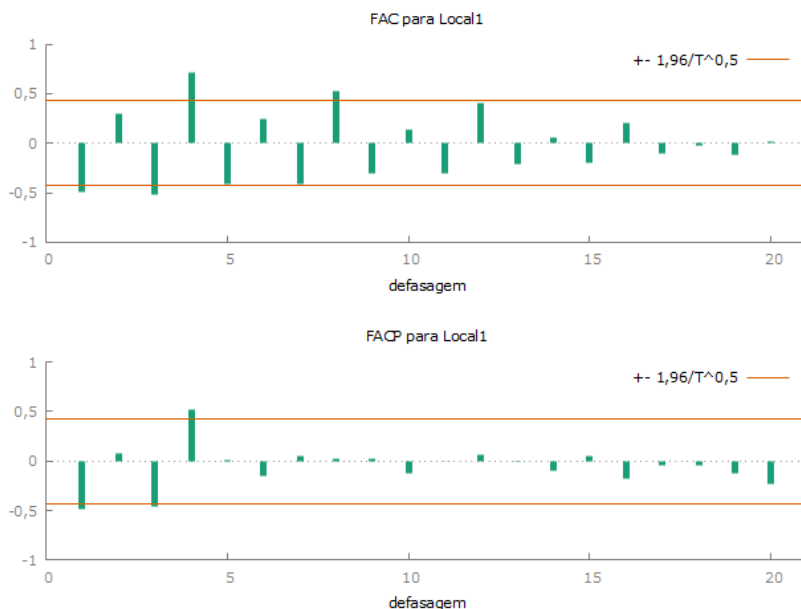


Figura 25: Correlograma e correlograma parcial - Filial 1

Utilizando o Teste de Dickey-Fuller o valor crítico calculado é de  $-9,48183$  com p-valor de  $1,798e-08$  rejeitando  $H_0$  com nível de significância de 5%, portanto temos evidências de que a série é estacionária. Com o pico significativo na defasagem 4 no correlograma parcial, pode-se supor que um ARIMA (4,0,0) seja um bom modelo. Também será estimado um ARIMA (3,0,2).

A Tabela 38 apresenta os resultados da estimação dos modelos ARIMA identificados no parágrafo anterior. Analisando visualmente a aderência dos dois modelos para a amostra treino, pode-se concluir que o AR(4) possui maior aderência aos dados observados:

Tabela 38: Resultado do Modelo ARIMA - Filial 1

	<b>Coefficiente</b>	<b>Erro padrão</b>	<b>z</b>	<b>p-valor</b>
<b>AR(4)</b>				
Constante	94,2801	30,0473	3,138	0,0017 ***
$\phi_1$	-0,0340511	0,121813	-0,2795	0,7798
$\phi_2$	0,00177965	0,0850965	0,02091	0,9833
$\phi_3$	-0,0504170	0,123636	-0,4078	0,6834
$\phi_4$	0,878952	0,0940331	9,347	8,99e-021 ***
<b>ARMA(3,2)</b>				
Constante	99,0243	9,40679	10,53	6,49e-026 ***
$\phi_1$	-1,00377	0,178189	-5,633	1,77e-08 ***
$\phi_2$	-0,821588	0,222365	-3,695	0,0002 ***
$\phi_3$	-0,787203	0,143759	-5,476	4,35e-08 ***
$\theta_1$	1,40349	0,372546	3,767	0,0002 ***
$\theta_2$	0,999974	0,499566	2,002	0,0453 **

Tabela 39: Estatística de Aderência

	<b>R-quadrado ajustado</b>	<b>Crítério de Akaike</b>
AR(4)	0,821251	237,8027
ARMA(3,2)	0,802239	242,3561

Observa-se, no primeiro modelo, o coeficiente referente ao termo autorregressivo de ordem 4 foi significativo ao nível de significância de 5%. No segundo modelo, os coeficientes  $\phi_3$  e  $\theta_2$  foram significativos ao nível de significância de 5%. Analisando as medidas de qualidade dos modelos, o que possui o maior  $R^2$  ajustado e o menor critério de informação é o modelo ARIMA(4,0,0). Na figura 26 e 27 apresenta os gráficos observados (efetivos) e ajustados de Ingressantes na filial 1 com ARIMA (4,0,0) e ARIMA (3,0,2).

Valores Observados (efetivos) e Ajustados de Ingressantes - Filial 1

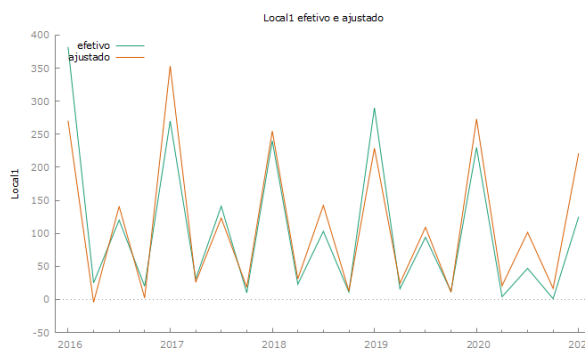


Figura 26: ARIMA (4,0,0)

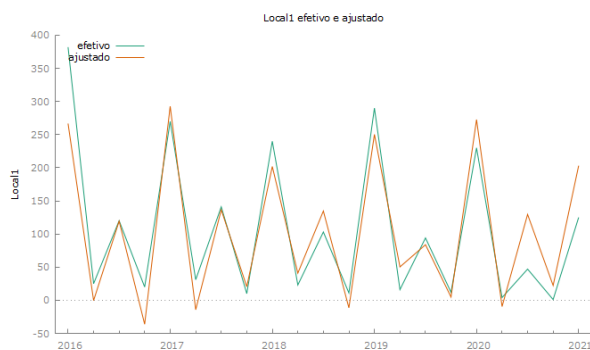


Figura 27: ARIMA (3,0,2)

Restringindo as observações para a amostra teste, que vai do segundo trimestre de 2021 (*Intra I*) até o terceiro trimestre de 2022 (*Inter II*), foram estimados os modelos ARIMA (4,0,0) e o ARIMA (3,0,2) nessa amostra e registrados os valores ajustados para calcular as estatísticas de aderência.

Tabela 40: Estatística de Aderência - Filial 1

	<b>EQM</b>	<b>MAD</b>	<b>MAPE</b>
AR(4)	38,869	23,82	123,01
ARMA(3,2)	68,05	63,935	733,25

conforme observa-se na Tabela 40, pode-se concluir que o ARIMA(4,0,0) possui maior aderência aos dados observados, uma vez que possui as menores medidas de aderência. Por esse motivo, o ARIMA(4,0,0) é o modelo escolhido para previsão.

### b) Análise do melhor modelo de previsão

A Tabela 41 apresenta os resultados da estimação do modelo ARIMA(4,0,0) para todos os dados disponíveis até o terceiro trimestre de 2022, isto é, as 27 observações.

Tabela 41: Resultado do modelo ARIMA - Filial 1

	<b>Coefficiente</b>	<b>Erro padrão</b>	<b>z</b>	<b>p-valor</b>
AR(4)				
Constante	98,3975	20,4516	4,811	1,50e-06 ***
$\phi_1$	-0,0671606	0,119875	-0,5603	0,5753
$\phi_2$	-0,0339453	0,0861163	-0,3942	0,6934
$\phi_3$	-0,0774923	0,112976	-0,6859	0,4928
$\phi_4$	0,860780	0,0892203	9,648	5,02e-022 ***

Observa-se que o coeficiente que representa a quarta autocorrelação é significativo ao nível 5%. O  $R^2$  ajustado foi 0,81, indicando alta explicação do modelo.

Tabela 42: Estatística de Aderência

	<b>R-quadrado ajustado</b>	<b>Critério de Akaike</b>
AR(4)	0,818481	298,4803

O diagnóstico dos resíduos do modelo ARIMA(4,0,0) inicia com o correlograma dos resíduos, apresentados na Figura 28.

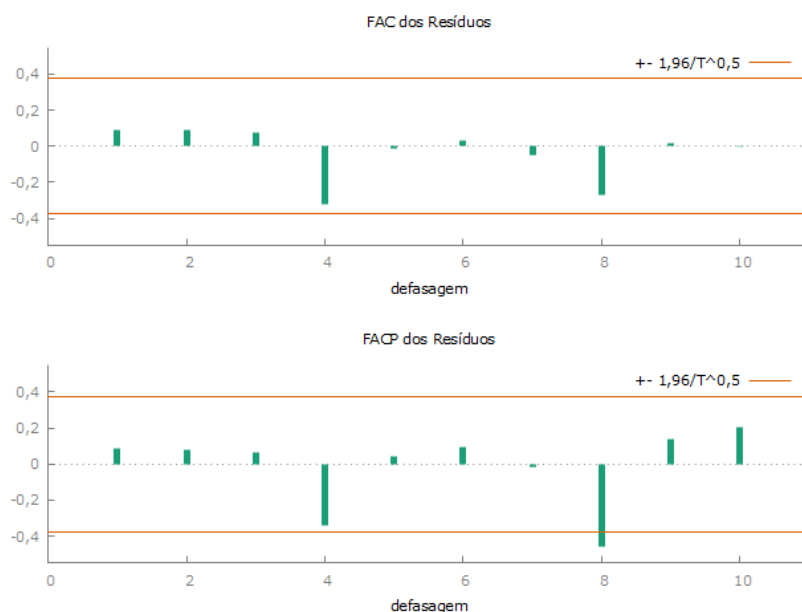


Figura 28: Correlograma dos resíduos - Filial 1

Observa-se que todas as 10 correlações estão dentro do intervalo de confiança, padrão típico de um ruído branco. O teste de Ljung-Box apresentou a estatística  $Q$  igual a 7,3929, com um  $p$ -valor de 0,286. Utilizando um nível de significância de 5%, não rejeita-se a hipótese nula de que o erro é um ruído branco.

### c) Previsão de valores futuros

A previsão dos ingressantes na Filial 1, utilizando o ARIMA(4,0,0), dois passos à frente, pode ser observado na Tabela 43.

Tabela 43: Previsão dos Ingressantes - Filial 1

<b>Observação</b>	<b>Previsão</b>	<b>Erro-padrão</b>	<b>Intervalo</b>
2022:4	13,05	43,018	-71,26 97,36
2023:1	216,21	43,115	131,17 300,72

Segundo este modelo, o número de ingressantes na Filial 1 no quarto trimestre de 2022 é de 13 ingressantes, podendo chegar até 97 ingressantes. Já no período Inter I de

2023, a previsão será de 216 matrículas, com um erro padrão de 43,1115, que resulta em um intervalo de 95% de confiança variando de 132 ingressantes até 301.

A Figura 29 apresenta estes valores previstos e os intervalos de 95% de confiança.

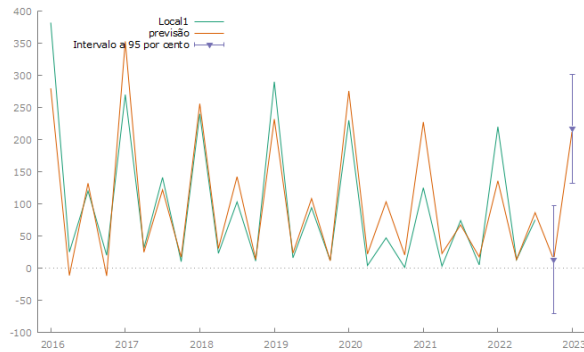


Figura 29: Previsão 2 trimestres - Filial 1

### 3.4.2 Modelagem ARIMA para a Filial 2

#### a) Escolha do melhor modelo de previsão

Para a amostra treino, foram considerados os mesmos períodos da análise anterior.

Para identificar os possíveis modelos ARIMA a serem ajustados à amostra treino, foram calculadas as funções de autocorrelação e autocorrelação parcial associadas ao ingresso trimestral, apresentadas na Figura 30.

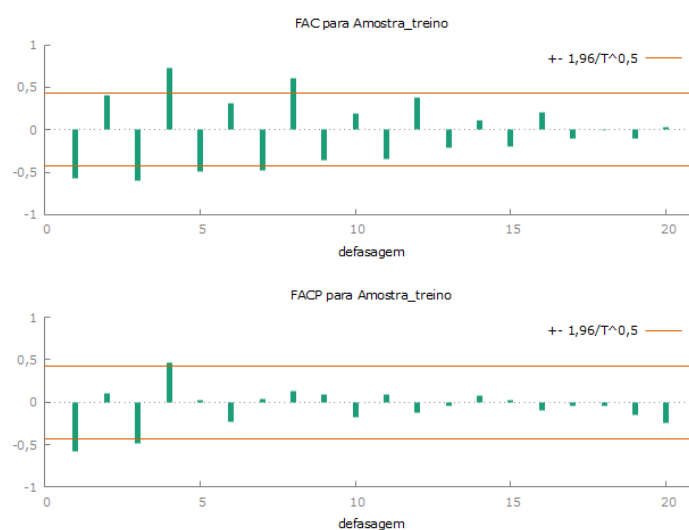


Figura 30: Correlograma e correlograma parcial - Filial 2

Utilizando o Teste de Dickey-Fuller o valor crítico calculado é de -10,1512 com p-valor de 1,648e-09 rejeitando  $H_0$  ao nível de significância de 5%, portanto temos evidências de

que a série é estacionária. Com o pico significativo na defasagem 4 no correlograma parcial, pode-se supor que um ARIMA (4,0,0) seja um bom modelo. Também será estimado um ARIMA (3,0,0).

A Tabela 44 a seguir apresenta os resultados da estimação dos modelos ARIMA identificados no parágrafo anterior.

Tabela 44: Resultado do Modelo ARIMA - Filial 2

	<b>Coefficiente</b>	<b>Erro padrão</b>	<b>z</b>	<b>p-valor</b>
<b>AR(4)</b>				
Constante	106,594	25,6039	4,163	3,14e-05 ***
$\phi_1$	-0,0611389	0,137020	-0,4462	0,6554
$\phi_2$	0,000285558	0,110300	0,002589	0,9979
$\phi_3$	-0,0719261	0,143566	-0,5010	0,6164
$\phi_4$	0,832604	0,122472	6,789	1,06e-011 ***
<b>AR(3)</b>				
Constante	115,120	6,11237	18,83	3,98e-079 ***
$\phi_1$	-0,602196	0,188759	-3,190	0,0014 ***
$\phi_2$	-0,395127	0,224533	-1,760	0,0784 *
$\phi_3$	-0,682065	0,173213	-3,398	8,23e-05 ***

Tabela 45: Estatística de Aderência

	<b>R-quadrado ajustado</b>	<b>Critério de Akaike</b>
AR(4)	0,816392	240,2622
AR(3)	0,632856	250,3125

Observa-se, no primeiro modelo, o coeficiente referente ao termo autorregressivo de ordem 4 foi significativo ao nível de significância de 5%. No segundo modelo, o coeficiente  $\phi_3$  foi significativo ao nível de significância de 5%. Analisando as medidas de qualidade dos modelos, o que possui o maior  $R^2$  ajustado e o menor critério de informação é o modelo ARIMA(4,0,0).

Na figura 31 e 32 apresenta os gráficos observados (efetivos) e ajustados de Ingressantes na filial 1 com ARIMA (4,0,0) e ARIMA (3,0,0).



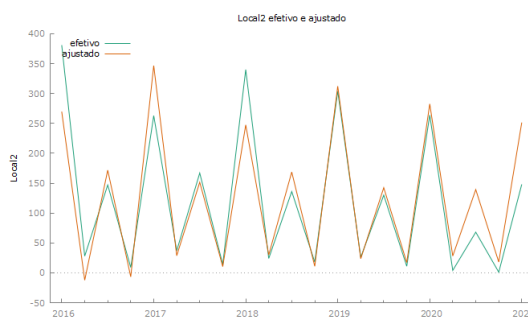


Figura 31: ARIMA(4,0,0)

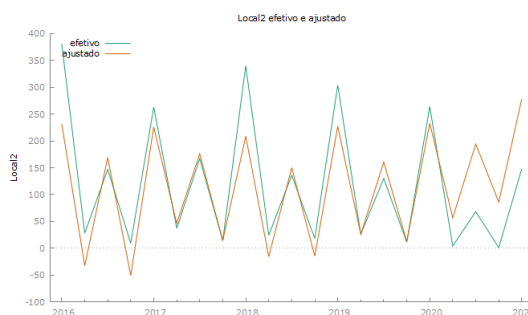


Figura 32: ARIMA(3,0,0)

Restringindo as observações para a amostra teste, que vai do segundo trimestre de 2021 (*Intra I*) até o terceiro trimestre de 2022 (*Inter II*), foram estimados os modelos ARIMA (4,0,0) e o ARIMA (3,0,0) nessa amostra e registrados os valores ajustados para calcular as estatísticas de aderência. Conforme observa-se na Tabela 47:

Tabela 46: Estatística de Aderência - Filial 2

	EQM	MAD	MAPE
AR(4)	20,418	17,148	334,95
AR(3)	98,442	90,001	1353,3

pode-se concluir que o ARIMA(4,0,0) possui maior aderência aos dados observados, uma vez que possui as menores medidas de aderência. Por esse motivo, o ARIMA(4,0,0) é o modelo escolhido para previsão.

#### b) Análise do melhor modelo de previsão

A Tabela 47 apresenta os resultados da estimação do modelo ARIMA(4,0,0) para todos os dados disponíveis até o terceiro trimestre de 2022, isto é, as 27 observações.

Tabela 47: Resultado do melhor modelo ARIMA - Filial 2

	Coefficiente	Erro padrão	z	p-valor
AR(4)				
Constante	101,412	28,1119	3,607	0,0003 ***
$\phi_1$	-0,0412742	0,105209	-0,3923	0,6948
$\phi_2$	-0,0118083	0,0753778	0,1567	0,8755
$\phi_3$	-0,0415949	0,102527	-0,4057	0,6850
$\phi_4$	0,878727	0,0783617	11,21	3,49e-029 ***

Tabela 48: Estatística de Aderência

	R-quadrado ajustado	Critério de Akaike
AR(4)	0,846537	297,9921

Observa-se que o coeficiente que representa a quarta autocorrelação é significativo ao nível 5%. O  $R^2$  ajustado foi 0,84, indicando alta explicação do modelo. O diagnóstico dos resíduos do modelo ARIMA(4,0,0) inicia com o correlograma dos resíduos, apresentados na Figura 33.

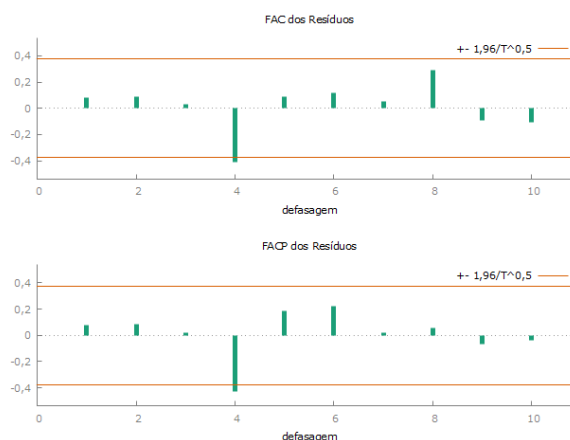


Figura 33: Correlograma dos resíduos - Filial 2

Observa-se que uma correlação está fora do intervalo de confiança, logo não é um ruído branco. O teste de Ljung-Box apresentou a estatística Q igual a 11,5053, com um p-valor de 0,074. Utilizando um nível de significância de 5%, rejeita-se a hipótese nula de que o erro é um ruído branco.

### c) Previsão de valores futuros

A previsão dos ingressantes na Filial 1, utilizando o ARIMA(4,0,0), dois passos à frente, pode ser observado na Tabela 49.

Tabela 49: Previsão dos Ingressantes - Filial 2

Observação	Previsão	Erro-padrão	Intervalo
2022:4	11,04	42,586	-72,43 94,51
2023:1	163,50	42,622	79,96 247,04

Segundo este modelo, o número de ingressantes na Filial 1 no quarto trimestre de 2022 é de 11 ingressantes, podendo chegar até 95 ingressantes. Já no período Inter I de 2023, a previsão será de 164 matrículas, com um erro padrão de 42,622, que resulta em um intervalo de 95% de confiança variando de 79,96 ingressantes até 247,04.

A Figura 34 apresenta estes valores previstos e os intervalos de 95% de confiança.



Figura 34: Previsão próximos 2 trimestres - Filial 2

## 4 Conclusão

Diante da situação de pandemia de Covid-19 decretada pela Organização Mundial da Saúde em 2020, as instituições de ensino se viram diante da necessidade de adequarem ao modelo remoto de ensino. Em alguns casos, isso pode ter influenciado em uma quebra estrutural no número de alunos matriculados. Por este motivo o presente trabalho visou entender o comportamento da quebra estrutural no número de alunos ativos e prever o número de ingressantes para os Ciclos *Intra II* de 2022 e *inter I* de 2023 em duas filiais de uma escola de idiomas.

Com os dados fornecidos pela escola de ensino de idiomas de 2016 a 2022, dois modelos estatísticos foram estimados para análise da quebra estrutural durante a pandemia e após a pandemia, para o número de alunos ativos mensais em duas etapas: o modelo de regressão linear, com a modelagem de Mínimos Quadrados Ordinários (MQO), sem as variáveis binárias e com elas e o teste de Chow.

Com isso, verificou-se durante os anos de 2017 a 2021, a filial 1 apresentou uma quebra estrutural na clientela de *criança e adolescente* com uma redução 323 matrículas ativas antes da pandemia e uma redução de 176 matrículas ativas mensais durante a pandemia e, na filial 2, a quebra estrutural antes da pandemia teve redução de 148 matrículas ativas e redução de 247 matrículas ativas durante a pandemia. Olhando para a quebra estrutural após a pandemia, na filial 1 houve quebra estrutural na clientela *criança e adolescente* com redução de 176 matrículas durante a pandemia e após a pandemia um aumento de 253 matrículas mensais, enquanto que, na filial 2, na clientela Criança/Adolescente, diante da redução de 247 matrículas anotadas no período durante a pandemia, há um aumento de 21 matrículas ativas mensais no período após pandemia. Na clientela *adulta* não houve quebra estrutural.

Para análise de previsão de novos alunos ingressantes trimestrais nos anos de 2022 a 2023, foi utilizado a metodologia de Box & Jenkins com a modelagem ARIMA. Após estimados os modelos com a amostra treino, foram feitas algumas estatísticas de aderência

para a escolha do melhor modelo ajustado para a amostra teste. Foram realizadas previsões pontuais com intervalos de confiança para o número de novas matrículas.

Obtendo os dados realizados no períodos *Intra II* em 2022 e *Inter I* em 2023 pode-se comparar a previsão com o realizado, onde na filial 1 a previsão no *Intra II* foi 13 ingressantes, com um intervalo de 95% de confiança variando de 0 a 97 alunos foram observados 5 ingressantes, dentro do intervalo prevista. Já no período *inter I* de 2023, a previsão era de 216 ingressantes variando de 132 a 300 e o realizado foi 152. Já na filial 2 o *Intra II* em 2022 a previsão foi de 11 ingressantes somente variando de 0 a 95 pessoas e o realizado 12 ingressantes, já em *Inter I* de 2023 a previsão era de 163 variando de 80 a 247 matrículas novas e foram realizadas 160.

O trabalho apresentou resultados importantes para a escola de ensino de idiomas. Os resultados alcançados não apenas corroboram os objetivos, mas também oferecem uma base sólida para que essa metodologia possa ser aplicada e adaptada em outras áreas de conhecimento.

# Referências

AKAIKE, H.

*"A new look at the statistical model identification"*, 1974.

BANFIELD, M. E. . N. W. R. *Product Leadership: How Top Product Managers Launch Awesome Products and Build Successful Teams*. 2017.

BOX G. E. P., J. G. M.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 1994.

BUENO, R. D. L. da S.

*Econometria de Séries Temporais*, 2012.

GUJARATI, D. N.

*Econometria básica*, 2011.

LEWIS, E. S. E.

*Getting the Most out of Business; Observations of the Application of the Scientific Method to Business Practice*, 1915.

MORAES, P. C. de.

*Inbound Market Como Estratégia de Atração dos Novos Consumidores* — Universidade Federal Fluminense, 2018.

SILVA, A. A. D. D.

*Testando Quebras Estruturais em uma Regra de Política Monetária com Regressores endógenos: um Estudo Empírico para o Brasil (2000-2016)* — Universidade Federal da Paraíba, 2016.

SOUZA, V. F. D.

*Investimento no Mercado de Capitais: Análise Fundamentalista na Bolsa de Valores B3* — Faculdades Integradas de Caratinga, 2018.