



Universidade de Brasília
Departamento de Estatística

**UMA ANÁLISE ESTATÍSTICA DO CENÁRIO DE CONSTRUÇÃO CIVIL NO DF
(DISTRITO FEDERAL) ANTES E DURANTE O PERÍODO PANDÊMICO DA COVID-
19: UMA COMPARAÇÃO COM UM CASO ESPECÍFICO**

Vinícius Paiva de Araujo

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília - DF

2022

Vinícius Paiva de Araujo

**UMA ANÁLISE ESTATÍSTICA DO CENÁRIO DE CONSTRUÇÃO CIVIL NO DF
(DISTRITO FEDERAL) ANTES E DURANTE O PERÍODO PANDÊMICO DA COVID-
19: UMA COMPARAÇÃO COM UM CASO ESPECÍFICO**

Orientador: Prof. Gladston Luiz da Silva

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília - DF

2022

Agradecimentos

Agradeço à toda minha família por sempre estar ao meu lado, com suas palavras de afeto e suas atitudes carinhosas que me impulsionaram para completar minha graduação.

Agradeço à Universidade de Brasília por disponibilizar uma multitude de oportunidades para a especialização de seus estudantes no mercado de trabalho.

Agradeço às entidades que me acolheram durante meus estágios e possibilitaram o desenvolvimento dos meus conhecimentos em Estatística adquiridos durante a minha graduação.

Finalmente, agradeço aos cientistas do mundo, por combaterem diariamente os preconceitos e tentarem fazer do mundo um lugar melhor.

Resumo

Este trabalho pretende analisar o comportamento do setor de Construção Civil no Distrito Federal ao longo de 4 anos por meio de dados fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Essa análise busca verificar o grau de correlação entre as variáveis por meio da técnica de Regressão e também realizar previsões sobre o comportamento desse setor com o uso da técnica de Séries Temporais com o modelo ARIMA. Foram feitos estudos sobre o comportamento das variáveis ao longo dos anos, especialmente com a presença da COVID-19, que ocasionou uma pandemia a partir de 2020. Observou-se que a variável número de empregados não ajuda a explicar a variável salário médio, enquanto que a variável número de empresas contribui negativamente para a explicação dessa. Além disso, não foi possível realizar as previsões pontuais devido à série temporal se assemelhar a um ruído branco.

Palavras-chaves: IBGE, Autorregressivo Integrado e de Médias Móveis, ARIMA, Correlação, Regressão, Pandemia, Distrito Federal, Construção Civil.

Abstract

This work aims to analyze the behavior of the Construction Sector in Distrito Federal (Federal District) over a period of 4 years using data provided by the Brazilian Institute of Geography and Statistics (IBGE). This analysis seeks to verify the degree of correlation between the variables through the Regression technique and also make predictions about the behavior of this sector using the Time Series technique with the ARIMA model. Studies have been conducted on the behavior of the variables over the years, specially with the presence of COVID-19, which caused a pandemic starting in 2020. It was observed that the variable "number of employees" does not help to explain the variable "average employee salary," while the variable "number of companies" contributes negatively to the explanation of the latter. In addition, it was not possible to make pontual forecasts due to the time series being a white noise.

Key-Words: IBGE, Autoregressive Integrated Moving Average, ARIMA, Correlation, Regression, Pandemic, Distrito Federal, Civil Construction.

SUMÁRIO

1. Introdução	6
2. Revisão Bibliográfica	8
3. Metodologia.....	20
4. Resultados	22
5. Considerações Finais	42
6. Referências Bibliográficas	43

1 Introdução

Conforme dados do Banco Central, a economia brasileira cresceu 2,9% em 2022 ante um avanço de 5% registrado no ano anterior. A desaceleração do Produto Interno Bruto (PIB) é explicada, sobretudo, pelo efeito pandemia, que atingiu a economia global em 2020. A pandemia causada pela corona vírus (Covid-19) foi um determinante para a falência de inúmeras empresas ao redor do mundo. Porém, um setor que continuou com suas atividades de forma relativamente regular, especialmente no Distrito Federal (DF), foi o de construção civil, levando-se em conta tanto grandes quanto pequenos empresários. Isso se deve, entre outros fatores, pelo número médio de funcionários imunizados neste ramo (SOUZA et al., 2022).

Foi possível observar que os microempreendedores tiveram as maiores taxas de falência quando comparadas às de grandes empresas. Estudos prévios mostram que a taxa de sobrevivência para as microempresas no Brasil após 5 anos é de apenas 21,5%. Essa taxa é mais crítica quando se considera as empresas de pequeno porte brasileiras, representando 17%. (GUERRA, 2021)

Ademais, a importância do setor de Construção Civil se aplica para todas as regiões brasileiras, com análises levantadas na região Sul em determinados tipos de mercado, por exemplo (PERUFO; GODOY, 2019). No caso do Distrito Federal, ele é extremamente presente no setor profissional da capital do país, apresentando um dos maiores números de habitantes por engenheiros, com 179,14 habitantes por engenheiro, enquanto que a média nacional se encontra em 267,62 habitantes por engenheiro no ano de 2013 (FÁBIO LIMA, 2013).

Como mencionado anteriormente, o perfil comum da microempresa no Brasil é de um curto período de vida, sendo o menos provável que continue em atividade após 5 anos desde sua criação. Isso se deve, entre outros fatores, à falta de gestão financeira dos negócios, o que leva ao fim prematuro da companhia (Palmeira et al., 2011). Além da falta de gestão, os pequenos empresários não focam no controle estatístico dos processos, resultado em produtos de qualidade heterogêneas e, conseqüentemente, insatisfação dos clientes. Em suma, a não observação dessas práticas afeta o desempenho do empreendedor e pode levar ao fracasso no mercado.

Este trabalho busca explorar especificidades do cenário de construção civil que ajudem a explicar os motivos deste setor não ter sido tão afetado pela Covid-19 no Distrito Federal por meio das análises estatísticas. Também se pretende trazer previsões sobre alguns temas importantes deste setor para o futuro próximo, como, por exemplo, a média salarial e número de empresas do ramo.

O objetivo deste trabalho é analisar o comportamento da Empresa M, que é uma empresa real e pediu para que sua identidade fosse ocultada para a realização deste trabalho, no cenário de construção civil no período de 2017 até 2021, levando-se em consideração o período pandêmico vivenciado devido ao vírus da Covid-19 e comparar seu desempenho com o cenário geral desse setor no Distrito Federal. Serão utilizados métodos de Regressão Linear Simples e Múltipla para verificar se há associação entre as variáveis, além de Análise de Séries Temporais para realizar previsões do cenário de construção civil no DF a pequeno e médios prazos.

Para alcançar este objetivo geral, seguem os seguintes objetivos específicos realizados:

- Analisar o desempenho da Empresa M de 2017 até 2021 no cenário geral de construção civil no DF;
- Ajustar modelos de regressão linear para identificar variáveis que ajudem a explicar o desempenho sobre o desempenho da Empresa M, levando-se em conta a pandemia vivenciada pelo vírus da Covid-19;
- Realizar previsões sobre o futuro econômico do setor de construção civil a pequeno e médios prazos.

Na seção 2, as metodologias utilizadas para alcançar esses objetivos são demonstradas. Na seção 3, os resultados obtidos são apresentados, junto com suas devidas explicações. Na seção 4, foram feitas as considerações finais para o cunho de ainda ser possível dar continuidade sobre o assunto deste trabalho. Finalmente, na seção 5, estão disponíveis as referências bibliográficas utilizadas para a confecção deste trabalho.

2 Revisão Bibliográfica

A Análise Exploratória é uma etapa crucial na organização e apresentação dos dados, pois busca identificar a tipologia das variáveis dos dados para então aplicar os melhores testes que explicam a situação problema, além de nortear ao pesquisados quais os melhores tipos de gráficos a serem utilizados para a apresentação das variáveis e seus comportamentos (MEDRI, 2011). Os tipos de variáveis seguem como:

- Qualitativa Nominal: Variável que assume “valores” em categorias, rótulos ou classes. Não existe relação de ordem entre elas. Exemplos: Etnia, nacionalidade, estado civil;
- Qualitativa Ordinal: Variável que assume “valores” em categorias, rótulos ou classes. Existe algum tipo de ordenação entre elas. Exemplos: Colocação de premiação, nível de escolaridade e satisfação média de um produto;
- Quantitativa Discreta: Variável que assume valores na escala numérica. Pode ser enumerável e existe no conjunto dos números inteiros. Exemplos: Idade, número de televisões na residência e número de integrantes da família;
- Quantitativa Contínua: Variável que assume valores na escala numérica. Não pode ser enumerável e existe no conjunto dos números reais geralmente. Exemplos: Média salarial, altura e peso.

2.1 Estudos Comparativos de Dados

Os Estudos Comparativos são de grande importância nas etapas iniciais do estudo de comportamento de variáveis, pois buscam estabelecer a correlação entre as variáveis através de um coeficiente para melhor compreensão da natureza dos dados. Além disso, também é comum montar um gráfico de dispersão para explicitar essa correlação. O indicador que é usado com mais frequência em dados de natureza linear é o Coeficiente de Correlação de Pearson (GRÁCIO; DE OLIVEIRAM, 2015).

Esse coeficiente traz um valor para a correlação entre os conjuntos “X” e “Y” de dados, podendo variar entre -1 a $+1$ e é mais aconselhável para conjuntos com associação linear. Algo notável de se apontar é que esse coeficiente detecta a tendência de crescimento dos valores, porém não consegue detectar a magnitude de tal crescimento. Isso significa que, caso as observações do conjunto “X” fossem multiplicadas por uma constante, o resultado do coeficiente continuaria o mesmo, não detectando a proximidade dos valores dos pares de “X” e “Y”. O Coeficiente de Correlação de Pearson é definido pela fórmula:

$$r = \frac{\left(\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}\right)}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$$

2.2 Testes de Normalidade de Shapiro-Wilk e Kolmogorov-Smirnov

Para a aplicação de diversas técnicas estatísticas, deve-se validar se os dados estudados seguem uma distribuição normal. Para esse propósito, são aplicados testes para validar tal informação, sendo os mais comuns os testes de Shapiro-Wilk e Kolmogorov-Smirnov (Razali, 2011).

A estatística do teste de Shapiro-Wilk é dada por:

$$W = \frac{b^2}{\sum_{i=1}^n (X_i - \bar{X})}$$

Onde b é uma constante dada por:

- $b = \sum_{i=1}^{n/2} a_{n-i+1} (X_{n-i+1} - X_i)$ caso o número de observações seja par;
- $b = \sum_{i=1}^{(n+1)/2} a_{n-i+1} (X_{n-i+1} - X_i)$ caso o número de observações seja ímpar.

Os valores de X_i são obtidos na ordenação crescente da amostra, enquanto que os valores de a_i são constantes calculadas a partir de:

$$(a_1, a_2, \dots, a_n) = \frac{(m^T V^{-1})}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

Sendo que m é uma matriz $1 \times n$ composta pelos valores esperados da estatística de ordem da amostra e V é a matriz de covariância de tais estatísticas.

As hipóteses desse teste são:

- $H_0 =$ A amostra segue uma distribuição Normal;
- $H_1 =$ A amostra não segue uma distribuição Normal.

A hipótese nula é rejeitada caso $W_{calculado} < W_{tabelado}$ ou caso o p-valor calculado seja menor que o grau de significância estabelecido.

A estatística do teste de Kolmogorov-Smirnov é dada por:

$$T = \sup_x | F^*(x) - F_n(x) |$$

Onde “sup” representa o supremo. $F^*(x)$ é a função de distribuição acumulada e $F_n(x)$ é a função de distribuição empírica estimada de acordo com a amostra. Para este teste de normalidade, $F_n(x)$ segue uma distribuição Normal com média μ e desvio-padrão σ .

As hipóteses são dadas por:

- $H_0 = F(x) = F_n(x), \forall x \in (-\infty, \infty);$
- $H_1 = F(x) \neq F_n(x), \forall x \in (-\infty, \infty).$

A hipótese nula é rejeitada quando $T_{calculado} <$ valor crítico, sendo este valor crítico calculado pelo nível de significância estabelecido.

2.3 Regressão Linear Simples e Múltipla

A Regressão Linear é uma técnica estatística para investigar e modelar a relação entre variáveis. Este tipo de técnica estabelece um conjunto de variáveis “X” como variáveis preditoras e outro conjunto de variáveis “Y” como variáveis resposta (MONTGOMERY; PECK; VINING,

2021).

Para o caso da Regressão Linear Múltipla, há mais de uma variável preditora para tentar explicar o fenômeno do conjunto “Y”. É importante destacar que dizemos que a regressão é linear devido à linearidade do modelo dos parâmetros e não devido à suposição que o conjunto “Y” é uma função linear das observações em “X”. A sua equação segue como:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Similarmente ao caso simples, a Regressão Linear Múltipla pressupõe algumas afirmações para que sua aplicação seja devida:

- A variável dependente Y é função linear das variáveis independentes X_1, X_2, \dots, X_k ;
- Os valores de X_1, X_2, \dots, X_k são independentes;
- Os erros do modelo são independentes, homocedásticos e seguem uma distribuição Normal padrão.

2.3.1 Método dos Mínimos Quadrados Ordinários (MMQO)

O objetivo de aplicar a Regressão Linear é poder estimar uma equação que melhor resume os dados vistos e conseguir ter um palpite para futuras observações. Para conseguirmos ter um palpite razoável, podemos estimar os parâmetros dessa equação pelo método dos mínimos quadrados.

Esse método de estimação consiste em escolher os parâmetros que minimizam a soma dos quadrados dos resíduos da amostra (FERNANDEZ; RIBOLDI, 1995). Isso pode ser expresso por:

$$SQ = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$

Podemos simplificar essa expressão para obtermos diretamente os parâmetros β_0, β_1 de interesse:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Onde:

- \bar{x} = média amostral de x_i ;
- \bar{y} = média amostral de y_i .

Estas expressões dos parâmetros são o que denominamos por estimativas de mínimos quadrados ordinários (MQO).

Ainda mais, a expressão do parâmetro β_1 pode ser vista como sendo a covariância amostral entre x_i e y_i . Então, podemos reescrevê-la como:

$$\beta_1 = \rho_{x_y} \left(\frac{\sigma_x}{\sigma_y} \right)$$

Onde:

- ρ_{x_y} é a correlação entre x_i e y_i ;
- σ_x e σ_y denotam os desvios-padrão da amostra.

Com isso, temos a implicação de que, caso x_i e y_i sejam correlacionados, $\beta_1 \neq 0$, sendo que este parâmetro pode ser positivo ou negativo, dependendo do tipo de correlação entre as variáveis.

Quanto aos resíduos do modelo, começaremos com sua definição. O resíduo é a diferença entre a observação “i” e o valor estimado para a essa observação. Podemos expressar isso por:

$$\epsilon_i = y_i - y_{\text{estimado}} = y_i - \beta_0 - \beta_1 x_i$$

Agora, como dito anteriormente, escolheremos os parâmetros que minimizem a soma dos quadrados dos resíduos dada por:

$$SQ = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$

2.4 Coeficiente de Variação

O coeficiente de Variação é uma técnica estatística utilizada para analisar a dispersão de dois ou mais conjuntos de dados em termos relativos a seu valor médio. Essa técnica se mostra especialmente útil quando os dados apresentam unidades de medidas diferentes (MONTGOMERY; PECK; VINING, 2021).

O cálculo desse coeficiente pode ser feito através da fórmula:

$$C.V = \frac{\hat{\sigma}}{\bar{X}} * 100\%$$

Onde $\hat{\sigma}$ é o desvio-padrão estimado e \bar{X} é a média amostral.

O valor desse coeficiente pode variar entre 0 a 1, por se tratar de uma porcentagem. A sua interpretação pode ser entendida como, quanto maior o valor obtido, maior a dispersão dos dados.

Para o estudo de Regressão Linear, esse coeficiente é utilizado no estudo de resíduos do modelo para verificar a sua homocedasticidade e ser possível ajustar um modelo linear para os dados.

2.5 Teste de Breusch-Pagan-Godfrey

O teste de Breusch-Pagan-Godfrey é utilizado para verificar a presença de heterocedasticidade nos dados. Para o caso de Regressão Linear, esse teste verificar a heterocedasticidade nos resíduos do modelo ajustado de regressão (MONTGOMERY; PECK; VINING, 2021).

As hipóteses para essa técnica são as seguintes:

- Hipótese Nula: $\delta_1 = \delta_2 = \dots = \delta_k = 0$;
- Hipótese Alternativa: $\delta_j \neq 0$.

O cálculo desse teste se dá pela seguinte fórmula:

$$\hat{\epsilon}_i^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + \mu_i$$

Onde o p-valor dessa estatística pode ser verificado na tabela Qui-Quadrado e δ_k é o coeficiente estimado da k-ésima variável preditora.

A interpretação desse teste pode ser feita por, ao estabelecer um nível de significância, a hipótese nula será rejeitada quando o valor da estatística χ^2 for maior que o valor crítico, então, seguirá que os resíduos são homocedásticos.

2.4 Análise de Séries Temporais

A análise de Séries Temporais é uma técnica que busca identificar tendências e sazonalidade em um conjunto de dados em um período de tempo (MORETTIN; CASTRO, 2008). Esses dados podem ser de natureza discreta ou contínua, porém com a sua principal característica sendo de suas observações vizinhas serem dependentes, construindo, assim, uma ideia de tempo.

Escrevendo de forma mais formal, seja $\{X(t), t \in T\}$ uma série temporal, onde T é o conjunto de tempos da série $T = t_1, t_2, \dots, t_n$. Uma série temporal será classificada como contínua quando os dados são observados ao longo do tempo T sem espaçamentos específicos, onde $T = t_1 < t_2 < \dots < t_n$. Alternativamente, a série pode ser classificada como discreta quando os dados são observados em intervalos de tempo iguais e/ou específicos.

O objetivo de realizar a análise de Séries Temporais, a princípio, é poder fazer previsões futuras com um nível de confiança satisfatório, detectar o comportamento que a série apresenta e procurar periodicidade relevante dos dados.

Existem dois tipos de séries temporais: as estacionárias e as não-estacionárias. A estacionariedade de uma série de segunda ordem é detectada a partir das seguintes características:

- $E[X(t)] = \mu$;
- $Cov[X(t), X(t+k)]$ depende apenas da diferença entre os tempos.

Essas características mostram que as séries aleatórias se desenvolvem ao redor de uma média constante. Porém, vale ressaltar que é mais comum encontrar séries que não sejam estacionárias, demandando a aplicação de diferenciações para cálculos futuros. De forma geral, pode-se escrever um processo $X(t)$ como:

$$X(t) = T_t + S_t + a_t$$

- a_t sendo um Ruído Branco, uma componente aleatória que possui distribuição Ruído Branco $\sim N(0, \theta_a^2)$;

- T_t é a componente de tendência;
- S_t é a componente de sazonalidade.

Contudo, as componentes de tendência e sazonalidade são bastante correlacionadas e, por isso, os métodos de estimação dessas componentes podem ser afetados caso alguma delas seja ignorada (MORETTIN; CASTRO, 2008). Com isso, não é aconselhado isolar uma componente da outra.

Caso não seja observado estacionariedade em uma série, é possível realizar transformações nos dados para que isso seja ajustado. Essas transformações causam na estabilização da variância e o efeito sazonal torna-se aditivo. O procedimento comumente utilizado é o de diferenças sucessivas dos dados originais da maneira:

$$\Delta X_t = X_t - X_{t-1}$$

sendo está a primeira diferença. Dando continuidade ao cálculo, tem-se:

$$\begin{aligned}\Delta^2 X_t &= \Delta [\Delta X_t] = \Delta X_t = \Delta [X_t - X_{t-1}] \\ \Delta^2 X_t &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

Agora, aplicando n diferenças sucessivas, obtém-se a fórmula geral:

$$\Delta^n = \Delta[\Delta^{n-1} X_t]$$

Vale ressaltar que também é comum a aplicação de transformações não-lineares em séries, caso isso se torne necessário. Esse tipo de procedimento é mais usual em séries de natureza econômica e financeira. Para esse tipo de aplicação, a transformação de Box-Cox (1964) é a mais utilizada, dada por:

- Caso $\lambda \neq 0$: $X_t^\lambda = \frac{\{X_t^\lambda - c\}}{\lambda}$,
- Caso $\lambda = 0$: $\log X_t$.

onde c e λ são parâmetros a serem estimados. Esse procedimento é adequado quando o desvio-padrão da série original for proporcional à média.

2.4.1 Modelos ARIMA e SARIMA

O modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade (SARIMA) é dado por:

$$\Phi_p(B^s) \Phi_p(B) \nabla_s^D \nabla^d = \Theta_Q(B^s) \Theta_q(B) \epsilon_t + c$$

Este modelo generaliza todos os modelos do modelo autorregressivo integrado de médias móveis (ARIMA) e este permite modelar séries estacionárias como também séries sazonais. Também segue a explicação dos parâmetros usados na equação (3.4.1):

- $\Phi_p =$ Termo autorregressivo sazonal;
- $\Phi_p =$ Termo autorregressivo;
- $\nabla_D =$ Diferenciação sazonal (1-B);
- $\nabla_d =$ Diferenciação da variável (1-B);
- $\Theta_Q =$ Média móvel sazonal;
- $\Theta_q =$ Média móvel;
- $\epsilon =$ Erro do modelo.

A seguir, tem-se um exemplo de aplicação de um modelo SARIMA (FIORUCCI, 2020) onde $x(t)$ segue um processo SARIMA (1,0,1) X (1,0,1), deseja-se descobrir a função da média condicional e dos resíduos do modelo. Para obter esta resposta, basta abrir a equação proposta em (3.4.1). Aplicando a equação, obtemos:

$$\Theta_1(B^s) \Phi_1(B) x(t) = \Theta_1(B^s) \Phi_1(B) \epsilon_t$$

$$\Rightarrow (1 - \varphi B^s)(1 - \Phi B) x(t) = (1 + \varphi B^s) (1 + \varphi B) \epsilon_t$$

$$\begin{aligned}
&\Rightarrow (1 - \Phi B - \varphi B^s + \Phi \varphi B^{s+1}) x(t) = (1 + \varphi B + \varphi B^s + \varphi \varphi B^{s+1}) \epsilon_t \\
&\Rightarrow X_t - \Phi X_{t-1} - \varphi X_{t-s} + \Phi \varphi X_{t-s-1} = \epsilon_t + \vartheta \epsilon_{t-1} + \vartheta X \epsilon_{t-1} + \vartheta \theta \epsilon_{t-s-1} \\
&\Rightarrow X_t = \Phi X_{t-1} + \varphi X_{t-s} + \Phi \varphi X_{t-s-1} + \epsilon_t + \vartheta \epsilon_{t-1} + \vartheta X \epsilon_{t-1} + \vartheta \theta X \epsilon_{t-s-1}
\end{aligned}$$

Com isso, temos que a média condicional é dada por:

$$\mu_{t|t-1} = \Phi X_{t-1} + \varphi X_{t-s} - \Phi \varphi X_{t-s-1} + \vartheta \epsilon_{t-1} + \vartheta \epsilon_{t-s} + \vartheta \vartheta \epsilon_{t-s-1}$$

Para os resíduos do modelo, se $\hat{\beta}$ é a estimativa do vetor de parâmetros, temos que:

$$\epsilon_t(B) = X_t - \mu_{t|t-1}(B), t = 1, 2, \dots, n$$

2.4.2 Diagnóstico do Modelo

Com a definição do modelo, é preciso verificar se este é adequado para o conjunto de dados em estudo. Para tanto, é comum se utilizar o teste de Ljung-Box para verificar se os resíduos do modelo são correlacionados ou não (Burns, 2002). Este teste propões as seguintes hipóteses:

- H_0 = Os resíduos não são correlacionados;
- H_1 = Os resíduos são correlacionados.

O cálculo para as estimativas das autocorrelações é feito por meio das fórmulas:

$$r_k = \frac{\sum_{t=k+1}^n a_t a_{t-1}}{\sum_{t=1}^n (a_t)^2}$$

A estatística de referência do teste é:

$$Q(k) = n(n-2) \sum_{j=1}^k \frac{(r_j^2)}{(n-j)}$$

Onde k é o número de defasagens usadas e $Q(k)$ se aproxima de uma distribuição Qui-quadrado com $(K - p - q)$ graus de liberdade. A hipótese nula é rejeitada caso $Q(k) > \chi^2_{1-\alpha, K-p-q}$.

3 Metodologia

Este trabalho utiliza dados disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) referentes à média salarial, número de funcionários no setor, número de empresas locais e PIB de estados fornecidos pelo Cadastro Central de Empresas (CEMPRE) e organizados pela Classificação Nacional de Atividades Econômicas (CNAE). Os dados estão disponibilizados no formato .csv e analisados no software R, utilizando o ambiente de desenvolvimento R-Studio e Posit Cloud, juntamente com o auxílio do software Excel. A base de dados contém informações sobre diversas características de empresas no Distrito Federal que serão listadas mais à frente.

Inicialmente foi estudado o setor de construção civil no DF, com foco aos anos de 2017 até 2019. Foram utilizadas as ferramentas de Séries Temporais e Regressão Linear para as análises estatísticas, sendo que na análise de Series Temporais foram verificadas as existências de sazonalidade e de tendência, enquanto, na análise de Regressão, foram identificadas as variáveis significativas do modelo.

Após a coleta de dados, foram consideradas apenas dados referentes às empresas do DF que atuam no setor da construção civil.

Após a análise exploratória dos dados, foi realizada a análise de Séries Temporais do tipo SARIMA para verificar se houve alguma mudança significativa a partir de 2020 no comportamento das variáveis devido ao impacto da Covid-19.

3.1 Conjunto de Dados

Para o estudo de caso de empresas no contexto do DF, o banco de dados contém informações sobre o salário médio do setor, o número de pessoas que trabalham na área e o número de empresas locais. Além disso, também foi utilizado um conjunto de dados sobre a variação percentual do PIB (Produto Interno Bruto) por trimestre ao longo dos anos de 2017 até 2021.

Para o estudo de caso específico, foram consideradas as seguintes variáveis disponibilizadas pela Empresa M:

- Despesas: Valor monetário gasto durante o mês;
- Receitas: Valor monetário adquirido durante o mês;
- Lucro Líquido: Diferença do total de receitas e despesas do mês.

4 Resultados

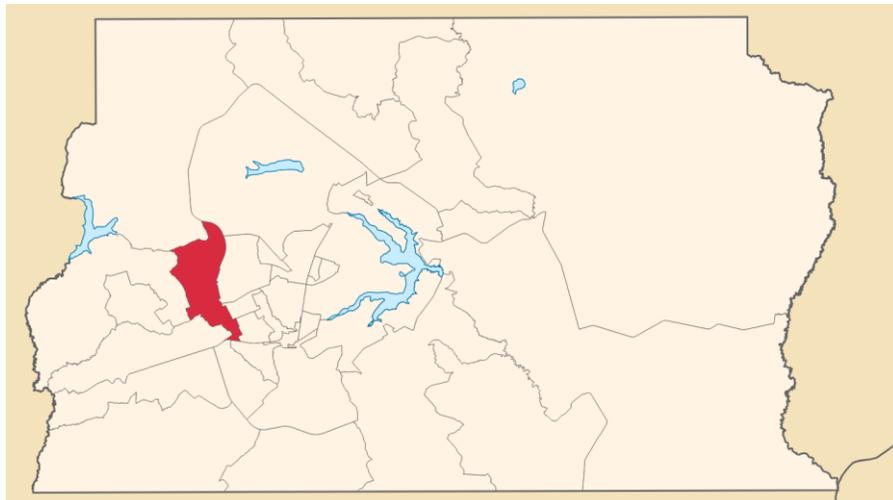
4.1 Contextualização

Esta seção do trabalho visa contextualizar a natureza do banco de dados, trazer o histórico da Empresa M no Distrito Federal e dar uma noção de como era o cenário de Construção Civil de 2017 até 2021. O código utilizado para a confecção desta seção e os bancos de dados públicos podem ser visualizados neste [hyperlink](#).

Esta empresa se especializou em produtos e serviços voltados para Construção Civil e têm como público-alvo fornecer serviços para outros negócios locais, geralmente pequenos e médios empreendedores, porém também já prestou serviços para obras públicas em Brasília. Foi inaugurada em 1992, possui 3 filiais no Distrito Federal em 2021, sendo a filial de Taguatinga a mais lucrativa e a que será utilizada como referência sobre para as próximas análises.

Sobre a área de atuação da Empresa M, ela realiza a maioria de seus serviços e fornecimentos para o entorno de Brasília, fazendo a grande parte de suas atividades longe da capital do Distrito Federal. Como referência, a Região Administrativa (R.A) de Taguatinga está localizada a mais de 20 quilômetros distante da capital.

Figura 1: Mapa de Brasília, com a R.A de Taguatinga destacada



FONTE: Administração de Taguatinga, 2020

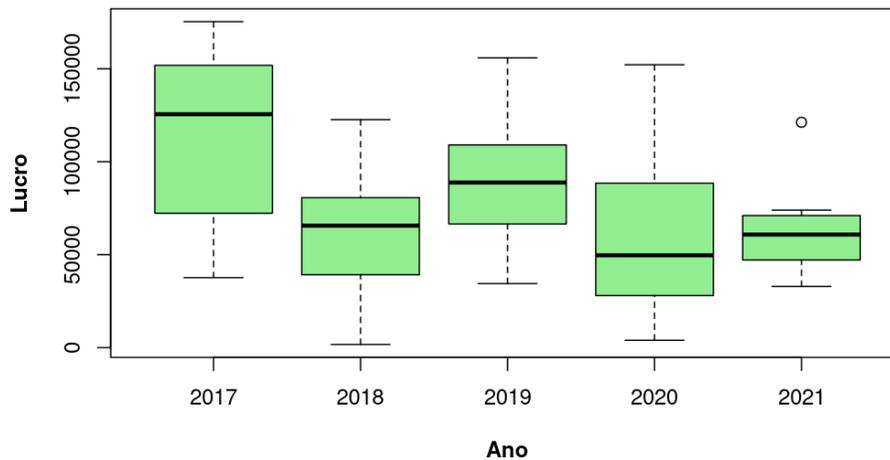
Essa R.A possuía mais de 220 mil habitantes no final de 2020, sendo a terceira maior cidade do DF (Administração de Taguatinga, 2020). Isso mostra como é uma região que cresce bastante

na capital do país e possui suas peculiaridades específicas de mercado. Por não ser uma cidade planejada, os moradores do distrito demandam cada vez mais locais de residência devido ao crescimento da população. Com esses fatores, o setor de Construção Civil se demonstra extremamente útil para suprir essa demanda, mostrando como ele é de grande interesse para essa R.A.

Complementando o fato de estar localizada em Taguatinga, a Empresa M também tem a vantagem estratégica de fornecimento para seus clientes. Devido à sua localização, a empresa pode atender seus clientes principais, que estão no entorno de Brasília, sem muita dificuldade de locomoção. Assim, eles reduzem os custos de mobilidade e oferecem filiais de atendimento próximas, o que torna seus serviços mais chamativos no mercado.

Para dar um pouco de contexto sobre a evolução da empresa M entre 2017 até 2021, a Gráfico 1 apresenta o lucro no período:

Gráfico 1: Lucro de 2017 até 2021



FONTE: IBGE, 2023

O *boxplot* do Gráfico 1 foi feito a partir dos valores de lucro dos meses de cada ano, sendo o lucro calculado a partir da diferença entre as receitas e as despesas do mês. Verifica-se que existe uma tendência negativa dos dados ao longo dos anos, porém a empresa M não registrou um mês de déficit nesse espaço de tempo. Também é importante notar que 2021 foi o ano que mostrou a

menor amplitude de dados, porém apresentou um *outlier*, sendo este o representante do mês de abril devido a um maior fluxo de transações do que o usual.

Também é possível analisar a média e o desvio-padrão dessa margem de lucros a partir da Tabela 1:

Tabela 1: Média e Desvio-padrão anual dos lucros da Empresa M

Anos	Média (R\$)	Desvio-Padrão (R\$)
2017	115.486,15	44.4115,63
2018	60.480,07	34.148,40
2019	89.446,67	32.007,53
2020	61.567,95	46.025,02
2021	61.744,55	23.382,40

FONTE: IBGE, 2023.

4.2 Histórico da Empresa

A Empresa M começou atuando com 12 funcionários efetivos em 1992, fazendo serviços para outros negócios locais. Ao longo dos anos, a empresa foi ganhando espaço na região de Taguatinga e conseguiu realizar serviços em obras maiores por meio de licitações públicas.

Surgiu em Brasília devido ao desmembramento de outro estabelecimento do mesmo nicho e estava planejando em sair do DF por motivos desconhecidos. Contudo, os funcionários desta filial decidiram montar um acordo com seu proprietário para poderem continuar operando no DF por meio de um novo CNPJ e nome fantasia. Após a finalização dessa transição, a Empresa M passou a ser propriedade de uma mesma família desde então.

Esse breve resumo de suas origens mostra que os funcionários dessa empresa já possuíam conhecimento prévio sobre os seus principais clientes e as especificidades do ramo de construção civil. Esses fatores podem ter auxiliado enormemente no seu sucesso no mercado e também para obtenção de novos contatos para serviços.

Esse êxito no mercado possibilitou a abertura de novas filiais nas áreas do Setor de Indústria e Reabastecimento (SIA) e Valparaíso e estão trabalhando para abrir outra unidade na BR 060 até

2024. Essas outras filiais operam da mesma forma que a unidade principal de Taguatinga, porém contam com um número de funcionários menor e também atendem menos clientes em média. Enquanto a unidade principal possui em torno de 30 empregados, as outras possuem cerca de 15 funcionários cada.

Essas filiais tentam manter um bom ambiente entre si. A Empresa M costuma celebrar datas comemorativas e faz reuniões entre seus funcionários e permite que tragam familiares e amigos para esses eventos. Por exemplo, os funcionários costumam trazer roupas festivas para em junho para fazerem uma festa de São João após o expediente em junho.

Além disso, também é comum os funcionários comemorarem aniversários e apoiar uns aos outros em momentos de dificuldade. Por exemplo, quando uma das funcionárias foi diagnosticada com câncer, os seus colegas de trabalho fizeram mensagens de apoio para ela e tentaram ajudá-la de outras formas enquanto fazia seu tratamento.

Essas atitudes mostram que a Empresa M tenta criar um ambiente amigável entre seus empregados e dar um sentimento de pertencimento a eles. Então fica claro que é uma empresa que tem como objetivo realizar seus serviços, mas sem perder a essência humana nas suas atividades.

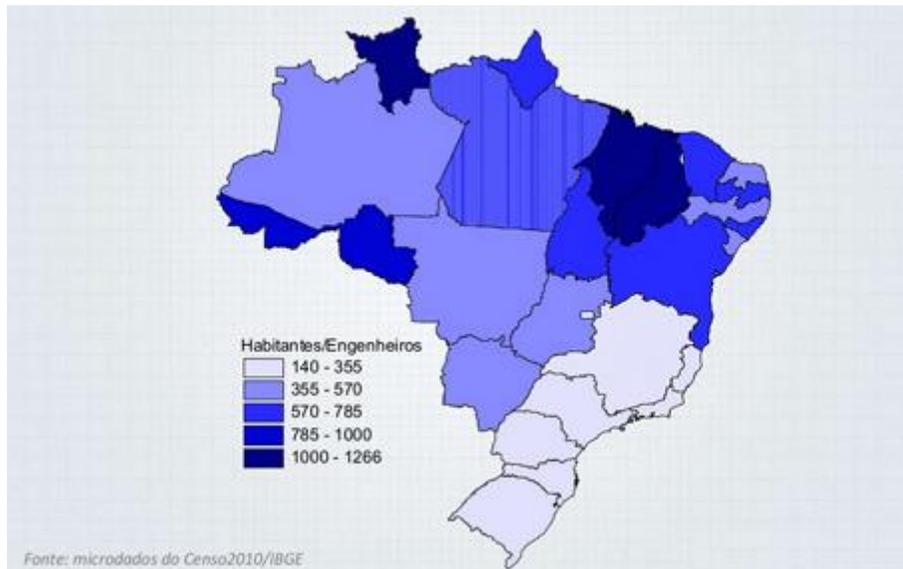
4.3 Cenário Econômico do Brasil e Distrito Federal

Nesta seção do trabalho, foram realizadas análises exploratórias e o uso de métodos estatísticos para a modelagem dos dados, conforme descrito anteriormente. As variáveis consideradas foram as seguintes:

- Média Salarial;
- Número de funcionários no setor;
- Número de empresas locais;
- PIB;
- Despesas;
- Receitas;
- Lucro Líquido.

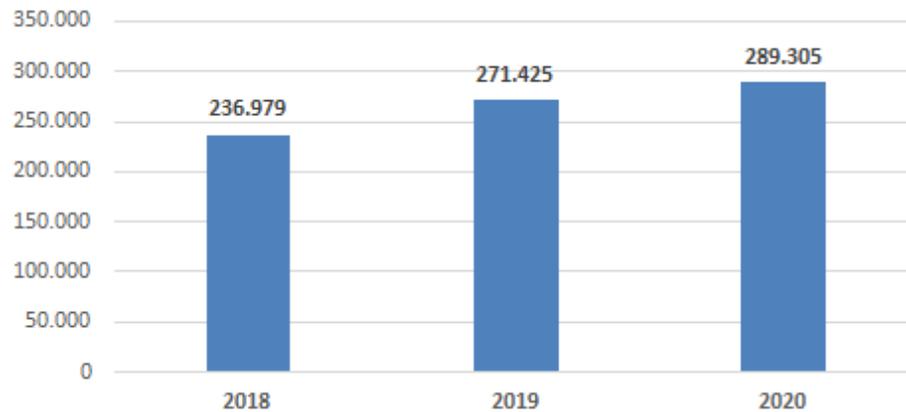
Para contextualizar o cenário econômico no Brasil e, principalmente, no DF de 2018 até 2020, foram feitos gráficos para representar o comportamento dos dados. Para a confecção destes gráficos, foi levado em consideração o número de empresas que realizavam serviços ou vendiam produtos relacionados à área de construção civil (IBGE, 2023). Além disso, também foi inserido um gráfico que mostra visualmente como o Distrito Federal possui um número elevado de engenheiros civis quando comparado com a média nacional (FÁBIO LIMA, 2013). Seguem algumas análises levantadas na pesquisa de Estatísticas do Cadastro Central de Empresas (CEMPRE) sobre o número de empresas relacionadas à construção civil no país como um todo e na capital.

Gráfico 2: Número de habitantes por engenheiro pelo Brasil



FONTE: FÁBIO LIMA, 2013.

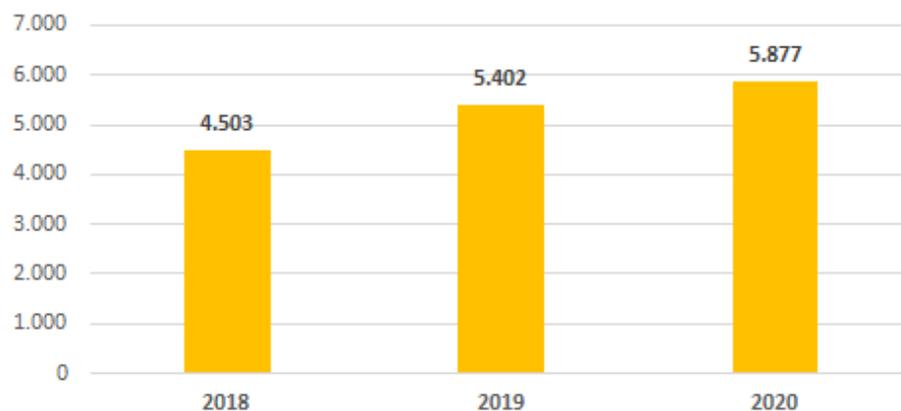
Gráfico 3: N° de Empresas no Brasil



FONTE: IBGE, 2023.

Os Gráficos 3 e 4 mostram que houve um aumento na quantidade de negócios que realizam serviços voltados para construção civil ao longo desses 3 anos. Esse aumento foi observado tanto na visão Brasil quanto na visão Distrito Federal e isso é um indício de aquecimento do setor no país.

Gráfico 4: N° de Empresas no DF

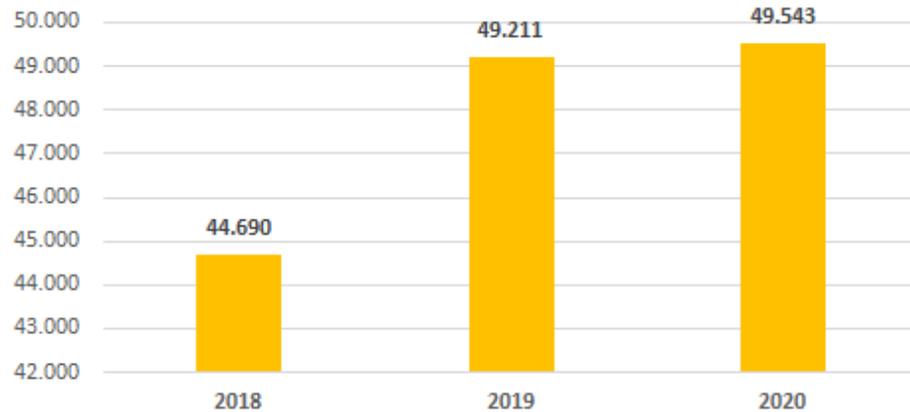


FONTE: IBGE, 2023.

O Gráfico 5 revela que o número de empregados no setor de construção civil no DF também aumentou nesse período de tempo, sendo 2019 o ano com o aumento mais significativo e uma

estabilização desse aumento em 2020. Uma possível explicação para este fenômeno é a necessidade de mais mão de obra que surgiu com a abertura de novas empresas desse setor nesses anos, como foi possível observar com o Gráfico 3 e Gráfico 4.

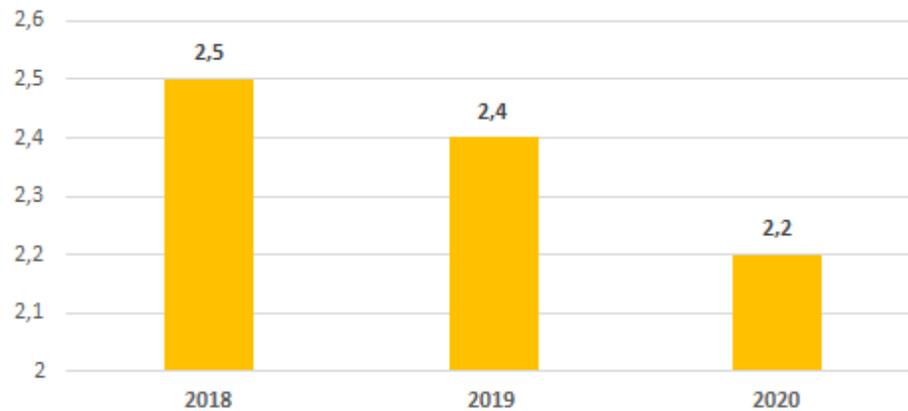
Gráfico 5: N° de Empregados na Construção Civil no DF



FONTE: IBGE, 2023.

O Gráfico 6 mostra que o salário médio dos funcionários deste ramo no DF diminuiu com o decorrer dos anos. Isso aparenta estar contradizendo as análises prévias, já que esse setor está em um crescimento econômico aparente nesse período. Uma hipótese que pode ser testada é essa diminuição ter ocorrido devido ao novo fluxo de funcionários que entraram neste setor.

Gráfico 6: Salário Médio de Construção Civil no DF



FONTE: IBGE, 2023.

Após essa a contextualização do cenário de Construção Civil no DF, também foi feito um estudo inicial sobre o coeficiente de variação (C.V) dessas variáveis referentes à unidade federativa. Como é possível verificar na Tabela 2, os C.V's das variáveis não ultrapassam de 11%, o que indica que os dados possuem baixa dispersão, sendo “Nº Empresas” a variável com maior variação nas observações.

Tabela 2: Resultados dos Coeficientes de Variação das variáveis referentes ao DF

	N° Empresas	N° Empregados	Salário Médio
Coeficiente de Variação	10,089%	3,667%	4,255%

FONTE: IBGE, 2023

Devido à quantidade baixa de observações para essas variáveis, o uso de Séries Temporais não é aconselhável. Então, para realizar as previsões pontuais do setor de Construção Civil para os seguintes anos, foram utilizadas outras variáveis para as quais tem-se mais observações. Neste caso, utilizou-se a variação percentual do PIB ao longo dos trimestres dos anos de 2017 até 2021.

4.4 Análise de Regressão Linear

Na análise de Regressão Linear Múltipla das variáveis referentes à Tabela 2, foi feito um estudo inicial das estatísticas da regressão para avaliar se o modelo se comporta bem para os dados. Os dados trazem informações das variáveis dos trimestres de 2018 até 2020, sendo feitas 12 observações. Esse período de tempo foi utilizado devido à disponibilidade dos dados. O modelo usa “Salário Médio” como variável resposta e “N° Empresas” e “N° Empregados” como as variáveis preditoras. É importante que foi levado em consideração as observações trimestrais de cada ano para a construção do teste.

A análise da Tabela 3 mostra que o modelo de Regressão Múltipla se ajusta relativamente bem para o conjunto de dados. Tanto o R-Quadrado quanto o R-Quadrado Ajustado apresentaram valores considerados “Fortes” para 12 observações feitas.

Tabela 3: Estatísticas observadas para a Regressão Linear Múltipla

Estatísticas	Resultados
R Múltiplo	0,9173
R-Quadrado	0,8414
R-Quadrado Ajustado	0,8062
Erro Padrão	0,0478
Observações	12

FONTE: IBGE, 2023

Entretanto, ainda é necessário analisar o resultado do teste de ANOVA e também verificar se os resíduos se ajustaram adequadamente. Como é possível verificar na Tabela 4, o valor F da estatística é muito maior que o valor F Crítico, com isso, rejeita-se a hipótese nula de que os coeficientes são, simultaneamente, iguais a zero, ou seja, que $\beta_0 = \beta_1 = 0$.

Tabela 4: Estatísticas observadas para o teste de ANOVA

Estatísticas	G.L	S.Q	M.Q	F	F Crítico
Regressão	2	0,10868	0,05434	23,87933	0,000252
Resíduo	9	0,02048	0,00228		
Total	11	0,12167			

FONTE: IBGE, 2023

Contudo, ainda é interessante estudar se os coeficientes são isoladamente iguais a zero, ou seja, se $\beta_0 = 0$ ou se $\beta_1 = 0$. Para tanto, foi aplicado o teste T em cada variável preditora e observar se há evidências estatísticas para rejeitar essa hipótese. De acordo com a Tabela 5, utilizando-se um nível de significância de 5%, foi verificado que “Nº Empresas” rejeita a hipótese nula, porém a variável “Nº Empregados” não rejeita a hipótese nula, ou seja, essa variável não contribui para o modelo da variável resposta “Salário Médio”, em outras palavras, essa variável preditora não ajuda a explicar a variável resposta.

Tabela 5: Estatísticas observadas para o teste T

Variáveis	Coefficientes	Erro-Padrão	Estatística T	P-valor
Interseção	2,3796	0,86010	2,76651	0,02187
Nº Empresas	-0,00035	0,00012	-2,92681	0,01684
Nº Empregados	0,00003	0,00003	1,27998	0,23255

FONTE: IBGE, 2023

Com isso, foi observado que o modelo de Regressão Múltipla de duas variáveis não consegue explicar razoavelmente o modelo. Portanto, será utilizado o modelo de Regressão Simples, já que apenas a variável “Nº Empresas” é razoável para explicar o modelo. Então, foram feitos os testes de Regressão Linear sem a variável “Nº Empregados”. Novamente, a variável resposta é “Salário Médio” e a variável preditora é “Nº Empresas” para as estatísticas das tabelas seguintes.

A Tabela 6 mostra que o modelo de Regressão Linear Simples se ajusta relativamente bem para o conjunto de dados. O R-Quadrado e o R-Quadrado Ajustado apresentaram valores considerados “Fortes” para 12 observações feitas, porém foi observado um leve aumento no Erro Padrão do modelo Simples em relação modelo Múltiplo.

Tabela 6: Estatísticas observadas para a Regressão Linear Simples

Estatísticas	Resultados
R Múltiplo	0,9014
R-Quadrado	0,8126
R-Quadrado Ajustado	0,7938
Erro Padrão	0,0492
Observações	12

FONTE: IBGE, 2023

O teste de ANOVA na Tabela 7 revelou que se rejeita a hipótese nula de que os coeficientes são, simultaneamente, iguais a zero devido ao valor F observado em relação ao F crítico.

Tabela 7: Estatísticas observadas para o teste de ANOVA

Estatísticas	G.L	S.Q	M.Q	F	F Crítico
Regressão	1	0,10496	0,10496	43,35278	0,0000062
Resíduo	10	0,02421	0,00242		
Total	11	0,12917			

FONTE: IBGE, 2023

O teste T da Tabela 8 mostrou que ambas variáveis foram consideradas significativas, ou seja, foi rejeitada a hipótese de os coeficientes serem iguais a zero.

Tabela 8: Estatísticas observadas para o teste T

Variáveis	Coefficientes	Erro-Padrão	Estatística T	P-valor
Interseção	3,46002	0,17045	20,2999	1,86E-09
Nº Empresas	-0,00020	0,00003	-6,5843	6,2E-05

FONTE: IBGE, 2023

Assim, é possível escrever o modelo como:

$$\widehat{\text{Salário Médio}} = 3,46002 - 0,0002 * \text{Nº Empresas}$$

Curiosamente, percebeu-se que o número de empresas impacta negativamente no salário médio dos funcionários do setor de Construção Civil, enquanto que o número de empregados não contribui para explicar o salário médio. Como o período de tempo contempla o período de pandemia da COVID-19, essa relação negativa entre as variáveis pode ser entendida como o surgimento de novas empresas do setor que, devido à competitividade do mercado, conseguiram menos projetos e, conseqüentemente, não obtinham uma margem de lucro satisfatória. Com essa suposição, o salário médio diminuiu nesse período.

Além disso, também foram comparados os valores de salário médio previstos pelo modelo proposto na Tabela 9. Como é possível ver, os resíduos do modelo final não apresentaram valores muito altos.

Tabela 9: Salário Médio Previsto pelo Modelo Final e Resíduos do Modelo

Observações	Salário Médio Observado	Salário Médio Previsto	Resíduos
1	2,5	2,53719	-0,03719
2	2,5	2,45583	0,04417
3	2,4	2,43534	-0,03534
4	2,4	2,39435	0,00565
5	2,4	2,35296	0,04704
6	2,4	2,33287	0,06713
7	2,3	2,33287	-0,03287
8	2,3	2,28164	0,01836
9	2,2	2,25561	-0,05561
10	2,2	2,23983	-0,03983
11	2,2	2,24823	-0,04823
12	2,3	2,23327	0,06673

FONTE: IBGE, 2023

Contudo, para a verificação do ajuste do modelo de Regressão Linear, também foi realizado os testes de homocedasticidade e normalidade dos resíduos. A Tabela 10 mostra o comportamento da normalidade dos resíduos sob o modelo final e, a um nível de significância de 5%, não foi rejeitada a hipótese nula, portanto, segue que os resíduos possuem distribuição Normal.

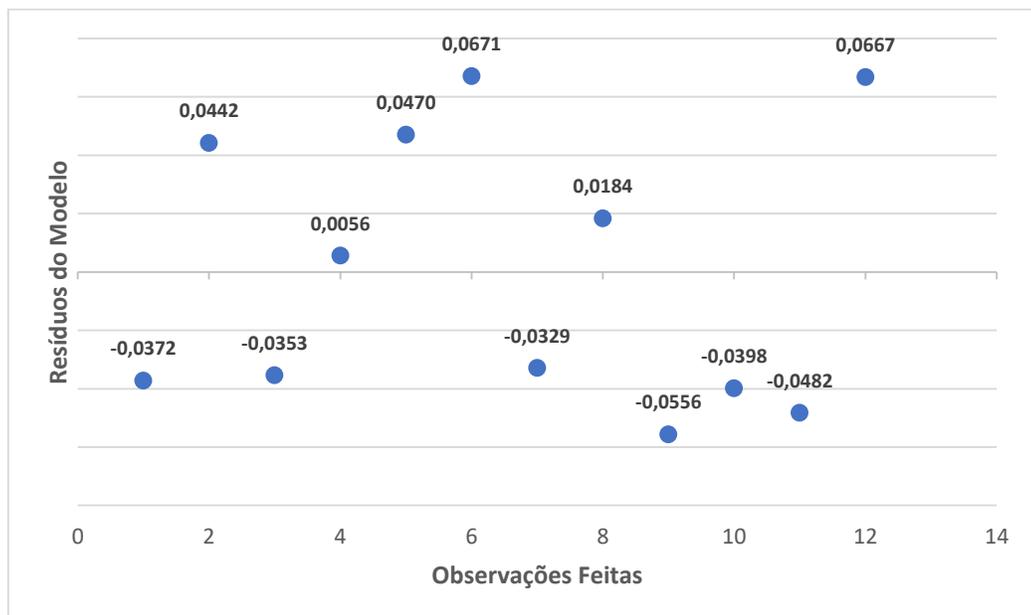
Tabela 9: Resultado da Estatística do teste de Shapiro-Wilk

Teste	Estatística (W)	P-valor
Shapiro-Wilk	0,86674	5,943%

FONTE: IBGE, 2023

Agora, para a questão da homocedasticidade, foi aplicado o teste de Breush-Pagan-Godfrey e a análise do Gráfico 7. Sob um nível de significância de 5% e de acordo com a Tabela 10, não foi rejeitada a hipótese nula, portanto, segue que os resíduos do modelo final são homocedásticos.

Gráfico 7: Homocedasticidade dos Resíduos do Modelo Final



FONTE: IBGE, 2023

Tabela 10: Resultado da Estatística do teste Breush-Pagan-Godfrey

Teste	Estatística (χ^2)	Graus de Liberdade (G.L)	P-valor
Breush-Pagan-Godfrey	1,6135	1	20,46%

FONTE: IBGE, 2023

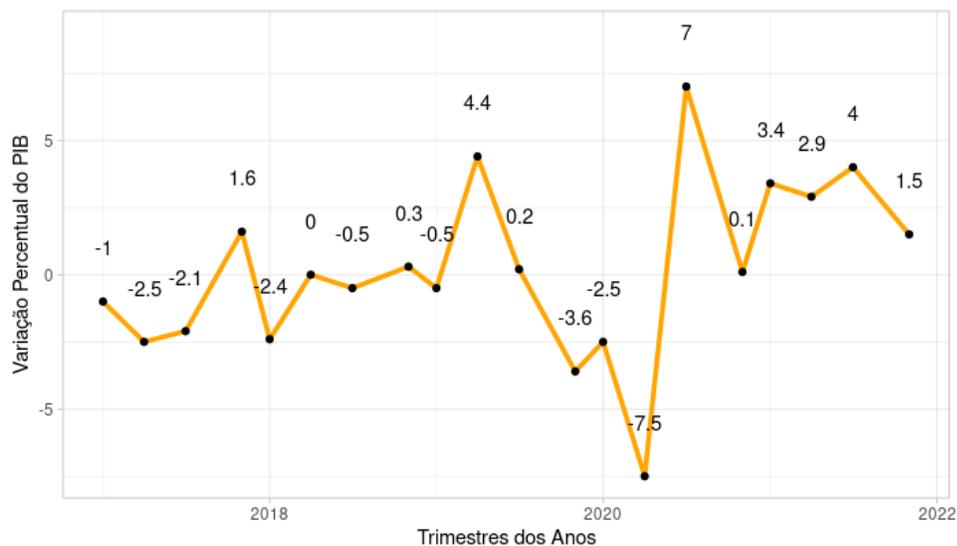
Após a verificação e confirmação da validade das premissas da Regressão Linear para o conjunto de dados, pode-se afirmar que a técnica estatística utilizada é adequada para o estudo.

Com isso, agora é possível prosseguir com o estudo de Séries Temporais para a variação do PIB e ajustar modelos para previsões pontuais.

4.5 Análise de Séries Temporais

O Gráfico 7 representa esse comportamento do PIB ao longo dos trimestres. Esse conjunto de dados do PIB será a série temporal utilizada para realizar previsões para os próximos anos. É importante destacar que a cada quatro observações dos dados representa um ciclo anual. Verifica-se que o setor de Construção Civil apresentava comportamento com poucas variações de 2017 a 2019. Observa-se a ocorrência de um decréscimo das atividades a partir do 3º trimestre de 2019, sendo que a maior variação sendo observada no 2º trimestre de 2020. Isso pode ser visto como uma evidência de como a COVID-19 teve um impacto inicial negativo nesse ramo de atividade, devido ao isolamento social que foi estabelecido na primeira metade de 2020.

Gráfico 8: Variação do PIB do DF de 2017 a 2021



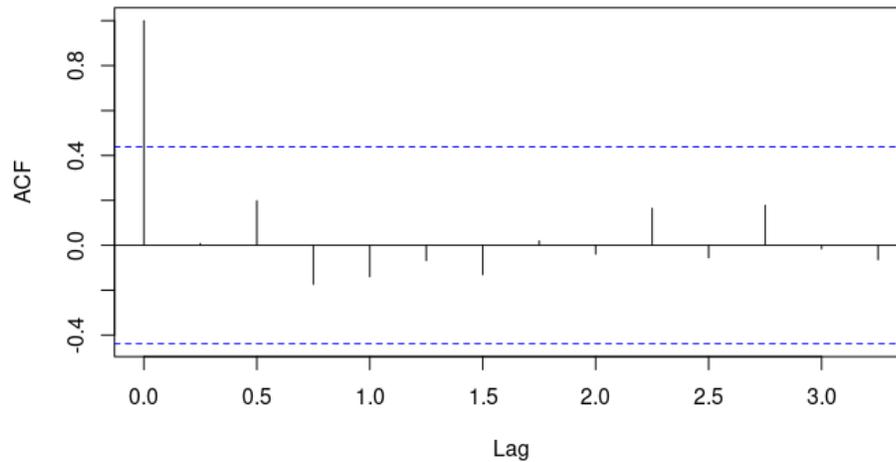
FONTE: CBIC, 2023

Contudo, após esse decréscimo substancial, houve um aumento considerável do PIB no 3º trimestre de 2020, sendo este a maior variação positiva observada no conjunto de dados. Isso se deveu à flexibilização do isolamento social nesse setor na segunda metade de 2020. Esse fato ainda pode ser reforçado com a análise das Gráficos 3 e 4, mostrando que houve um aquecimento relativo dessa área nesse ano. Algo curioso é que os trabalhadores de Construção Civil apresentaram a

menor proporção de imunização (81,5%) entre os funcionários de serviço e, mesmo assim, tiveram a menor taxa de infecção nesse grupo (21,8%) em um estudo feito no final de 2022 (Souza et.al, 2022).

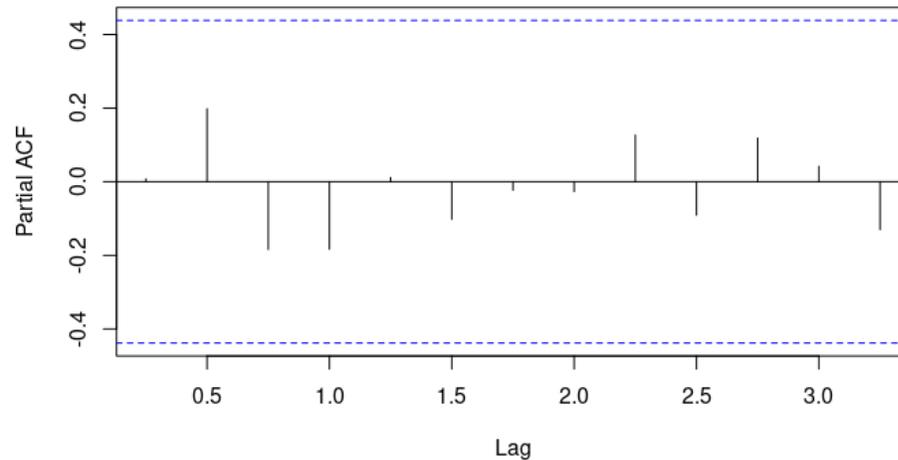
Na análise da série temporal do PIB, foram analisados primeiramente os gráficos gerados pela Função de Autocorrelação (FAC) e pela Função de Autocorrelação Parcial (PACF) para determinar o melhor modelo para o conjunto de dados. Para a modelagem da série, foi utilizado a função *ndiffs* no RStudio para verificar quantas diferenciações simples são necessárias para tornar a série estacionária. O resultado gerado mostrou que não é necessário realizar diferenciações simples para tornar a série estacionário, ou seja, tem-se que $m = 0$. Os gráficos 8 e 9 mostram que não há autocorrelações significativas, ou seja, tem-se que $p = 0$ $q = 0$.

Gráfico 9: Função de Autocorrelação do PIB



FONTE: CBIC, 2023

Gráfico 10: Função de Autocorrelação Parcial do PIB



FONTE: CBIC, 2023

Com essas análises feitas, agora serão testados alguns modelos a fim de escolher o mais adequado para o conjunto de dados do PIB. Para tanto, será analisado os valores de AIC, AICc, e BIC para os modelos testados.

Tabela 11: Valores de critério de observação dos modelos para a série do PIB

Modelo	AIC	AICc	BIC
ARIMA (1,0,0)	109,03	109,12	112,02
ARIMA (0,0,1)	111,34	110,57	113,35
ARIMA (0,0,0)	105,07	105,03	106,07
ARIMA (1,0,1)	110,42	110,84	114,40

FONTE: CBIC, 2023

Deseja-se selecionar o modelo que apresente os menores valores de AIC. Porém, percebe-se que os valores de AIC, AICc e BIC são muito próximos. Com isso, para selecionar o modelo que não esteja super parametrizado e também devido à estacionariedade da série, a modelagem

mais indicada será uma ARIMA (0,0,0) com média zero. A escolha desse modelo também foi indicado pela função *auto.arima* do Rstudio, que é um bom costume para verificar a adequabilidade do modelo, sendo que essa função selecionou o modelo ARIMA (0,0,0) com média zero. Adicionalmente, foi testado aplicar uma transformação Box-Cox para a série temporal, porém não foi observado uma melhora nos AIC's do modelo transformado, sendo que o valor de lambda obtido foi de 0,1054.

Tabela 12: Valores de critério de observação dos modelos para a série transformada do PIB

Modelo	AIC	AICc	BIC
ARIMA (1,0,0)	120,03	120,24	121,37
ARIMA (0,0,1)	122,40	122,67	124,45
ARIMA (0,0,0)	118,44	118,55	120,43
ARIMA (1,0,1)	121,78	122,13	125,76

FONTE: CBIC, 2023

Assim, foi escolhido o modelo sem a transformação Box-Cox. A adequabilidade dessa configuração pode ser verificada a partir do estudo do teste de Ljung-Box, com um *lag* de 5, para verificar se os resíduos são correlacionados, que foi aplicado na Tabela 8. A análise dessa tabela mostra que não há evidências estatísticas para rejeitar a hipótese de que os resíduos são não correlacionados ao nível de significância de 5%.

Tabela 13: Resultado do teste de Ljung-Box para os resíduos do modelo ARIMA (0,0,0)

Teste	Estatística	P-valor
Ljung-Box	0,001428	0,9693

FONTE: CBIC, 2023

Entretanto, também é comum observar se há normalidade nos resíduos. Para tanto, aplicou-se os testes de Shapiro-Wilk e Kolmogorov-Smirnov para os dados, onde os resultados podem ser

vistos na Tabela 13. A sua análise revela que a hipótese nula não é rejeitada para ambos os testes, ou seja, há normalidade nos resíduos.

Tabela 14: Resultado dos testes de Shapiro-Wilk e Kolmogorov-Smirnov para a normalidade dos resíduos do modelo ARIMA (0,0,0)

Teste	Estatística	P-valor
Shapiro-Wilk (W)	0,97711	0,8916
Kolmogorov-Smirnov (D)	0,3	0,1013

FONTE: CBIC, 2023

Contudo, o modelo selecionado foi um ARIMA (0,0,0) com média zero, isso significa que a série temporal segue um comportamento semelhante ao de um ruído branco. As características de um ruído branco são que a sua média é zero, sua variância é constante e não possui autocorrelação. Isso implica que realizar previsões pontuais para o conjunto de dados não é aconselhável em consequência da série temporal se aproximar de um ruído branco, pois as previsões pontuais serão iguais a média da série, que, neste caso, é zero e não contribuiria para a tomada de decisões.

5 Considerações Finais

Como foi possível ver ao longo deste trabalho, foi observado que o setor de Construção Civil no Distrito Federal é um ramo que vem se expandindo na década de 2020. O estudo dessa área pode trazer ganhos úteis para a sociedade, mesmo que esse estudo seja apenas uma projeção futura de seu comportamento ou meramente analisar a relação entre suas inúmeras variáveis.

Muitos desses comportamentos não foram explorados neste trabalho, como comparar o salário médio de funcionários de empresas localizadas em regiões administrativas diferentes no Distrito Federal, por exemplo. Além disso, também seria interessante efetuar a Regressão Múltipla com outras variáveis, como ano de inauguração da empresa, por exemplo.

Para o caso de previsões pontuais, também é aconselhável buscar bancos de dados que contêm informações mensais sobre as variáveis de interesse, ao invés de apenas informações trimestrais. Assim, é possível utilizar uma série temporal que não apresente características de um ruído branco e seja possível realizar previsões com acurácias satisfatórias.

Em suma, este trabalho explorou apenas um dos diversos nichos que podem ser estudados no setor de Construção Civil, ainda se limitando a somente uma unidade federativa no Brasil.

REFERÊNCIAS

ADMINISTRAÇÃO REGIONAL DE TAGUATINGA. **Terceira maior cidade do Distrito Federal – com mais de 220 mil habitantes – Taguatinga recebeu reforço na segurança do centro da cidade.** Administração Regional de Taguatinga, Brasília (DF), dez. 2020. Disponível em: < <https://www.taguatinga.df.gov.br/2020/12/04/terceira-maior-cidade-do-distrito-federal-com-mais-de-220-mil-habitantes-taguatinga-recebeu-reforco-na-seguranca-do-centro-da-cidade/>>. Acesso em: 11 jun. 2023.

BOX, G. E. P. & COX, D. R. **An analysis of transformations**, Journal of the Royal Statistical Society, Series B, 26, 211-252. 1964.

BURNS, Patrick. **Robustness of the Ljung-Box test and its rank equivalent.** Available at SSRN 443560, 2002.

CBIC. **Informativo Econômico:** Construção Civil superou todas as expectativas e cresceu 9,7% em 2021 [Banco de Dados 2022]. CBIC – Câmara Brasileira da Indústria da Construção, Brasília (DF), mar. 2023. Disponível em: <<https://cbic.org.br/wp-content/uploads/2022/03/informativo-economico-pib-04-marco-2022.pdf>>. Acesso em: 20 mar. 2023.

FERNANDEZ, Dinara Westphalen Xavier; RIBOLDI, João. **Análise de regressão e correlação.** Cadernos de matemática e estatística. Série B, Trabalho de apoio didático. Porto Alegre. N. 30 (jun. 1995), p. 1-74, 1995.

FIORUCCI, J. A. **Aula 10 – Modelos ARIMA Sazonais (SARIMA).** Youtube, 28 de setembro de 2020. Disponível em <<https://www.youtube.com/watch?v=Q8F15DLOmM&t=1095s>>. Acesso em: 20 mar. 2023.

GRÁCIO, Maria Cláudia Cabrini; DE OLIVEIRA, Ely Francina Tannuri. **Indicadores de proximidades em análise de cocitação de autores: um estudo comparativo entre coeficiente de Correlação de Pearson e Cosseno de Salton.** Informação & Sociedade: Estudos, v. 25, n. 2, p. 116-116, 2015.

GUERRA, A. C. **Sebrae: Pequenos Negócios Têm Maior taxa de mortalidade.** 2021. Disponível em: <<https://agenciabrasil.ebc.com.br/economia/noticia/2021-06/sebrae-pequenos-negocios-tem-maior-taxa-de-mortalidade>>.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estatísticas do**

Cadastro Central de Empresas – CEMPRE. Rio de Janeiro: IBGE, 2023. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/comercio/9016-estatisticas-do-cadastro-central-de-empresas.html?=&t=destaques>>. Acesso em: 20 mar. 2023.

LIMA, Fábio. **Piauí é o segundo estado com menos engenheiros do Brasil, revela IPEA.** Cidade Verde. Piauí (PI), 02 dez. 2013. Disponível em: <<https://cidadeverde.com/noticias/149829/piaui-e-o-segundo-estado-com-menos-engenheiros-do-brasil-revela-ipea>>. Acesso em 03 jun. 2023.

MEDRI, W. **Análise exploratória de dados.** Londrina: Universidade Estadual de Londrina, 2011.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis.** [S.l.]: John Wiley & Sons, 2021.

MORETTIN, P.; CASTRO, T. **Análise de séries temporais.** [S.l.]: Edgard Blucher, 2008.

PALMEIRA, E. et al. **Gestão financeira das microempresas.** Revista de Economia, n. 153, 2011.

PERUFO, L. D.; GODOY, L. P. **Mortalidade de microempresas: Um estudo de campo realizado com microempresários da região centro do estado do rio grande do sul.** Revista Pretexto, p. 11–27, 2019.

RAZALI, Nornadiah Mohd et al. **Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests.** Journal of statistical modeling and analytics, v. 2, n. 1, p. 21-33, 2011.

SOUZA, J. U. de et al. **Covid-19 E as repercussões no mercado de trabalho da área metropolitana ... 2022.** Disponível em: < https://www.ipe.df.gov.br/wp-content/uploads/2019/05/SE-Covid-19_e_as_repercussoes_no_mercado_de_trabalho_da_Area_Metropolitana_de_Brasilia.pdf >. Acesso em 15 mar. 2023.