

#interna

**Compras públicas e contratos em Informática (TIC) na Administração Pública Federal: Análise  
Inferencial com Dados do Portal Contratos.gov.br**

Paulo Evelton Lemos de Sousa

Programa de Pós Graduação em Administração - PPGA

Universidade de Brasília – UnB

Brasília, Brasil

paulo.evelton@aluno.unb.br / ORCID: 0000-0001-5725-4894

Solange Alfinito

Programa de Pós Graduação em Administração - PPGA

Universidade de Brasília – UnB

Brasília, Brasil

salfinito@unb.br / ORCID: 0000-0001-6235-7564

### Resumo

Este artigo analisa os fatores que influenciam a recontratação de fornecedores de Tecnologia da Informação e Comunicação (TIC) no âmbito dos ministérios do Poder Executivo Federal, com base em 1.008 contratos extraídos do portal Contratos.gov.br entre os anos de 2020 e 2024. Adota-se uma abordagem metodológica híbrida, que combina técnicas estatísticas inferenciais, modelos de regressão e algoritmos de aprendizado de máquina supervisionado (Random Forest e XGBoost), com o objetivo de identificar padrões associados à recontratação de fornecedores. Após o enriquecimento da base com informações estruturais dos fornecedores (tempo de atividade, porte, capital social, localização e registros em cadastros de sanções), foram aplicados testes de Mann-Whitney, Kruskal-Wallis e correlação de Spearman, além de regressões robustas (OLS, quantílica e bootstrap). Os resultados indicam que o capital social, a idade da empresa e a diversidade de ministérios contratantes estão positivamente associados à recontratação. Em contrapartida, a presença de sanções reduz significativamente essa probabilidade. Os modelos preditivos apresentaram elevado desempenho ( $AUC-ROC > 0,90$ ), corroborando os achados estatísticos. As principais limitações envolvem a dificuldade de enriquecimento da base via CNPJ, o que restringiu a amostra inferencial a 143 fornecedores, a partir de um universo inicial de 479. O estudo contribui com evidências empíricas para o aprimoramento de políticas de avaliação de desempenho, integridade e eficiência nas compras públicas.

**Palavras-chave:** compras públicas; contratos; recontratação; TIC; aprendizado de máquina.

### Abstract

This article examines the factors associated with the re-hiring of Information and Communication Technology (ICT) suppliers by Brazil's Federal Executive Ministries, based on 1,008 public contracts

obtained from the Contratos.gov.br portal between 2020 and 2024. A hybrid methodological approach was employed, combining inferential statistical techniques, regression models, and supervised machine learning algorithms (Random Forest and XGBoost) to identify patterns related to supplier re-hiring. The dataset was enriched with supplier-level structural data—such as time in operation, size, capital, location, and records in sanction databases—enabling the application of Mann-Whitney U, Kruskal-Wallis, and Spearman correlation tests, along with robust regression techniques (OLS, quantile, and bootstrap). The results indicate that higher capital, longer time in operation, and broader engagement with multiple ministries are positively associated with re-hiring. Conversely, the presence of sanctions significantly reduces the likelihood of re-hiring. The predictive models demonstrated strong performance (AUC-ROC > 0.90), reinforcing the statistical findings. The main limitation concerns the difficulty in enriching supplier data via CNPJ identifiers, which reduced the inferential sample to 143 suppliers out of an initial 479. This study contributes empirical evidence to inform improvements in supplier evaluation, risk management, and integrity mechanisms in public procurement.

**Keywords:** public procurement; contracts; re-hiring; ICT; machine learning.

## 1. INTRODUÇÃO

A gestão pública, particularmente no setor de compras, é um pilar fundamental para a administração eficiente dos recursos e a prestação de serviços à cidadania (Akaba et al., 2020). No Brasil, a Emenda Constitucional nº 19/98 (1998) introduziu a eficiência como um princípio basilar da Administração Pública, impulsionando a busca por otimização de recursos e excelência. Compreender o comportamento e as características dos contratos e fornecedores no setor de compras públicas é, portanto, essencial para aprimorar a tomada de decisão estratégica e tática.

As compras públicas representam uma parcela significativa da economia global e são um pilar fundamental da gestão governamental em diversos países, incluindo o Brasil. Ao longo dos anos, o tema tem ganhado relevância exponencial não apenas por seu volume financeiro, mas também por seu potencial estratégico como instrumento de política pública para impulsionar a inovação, a sustentabilidade e o desenvolvimento socioeconômico (Amann et al., 2014; Edler & Georghiou, 2007; Uyarra et al., 2014).

No entanto, a complexidade inerente às compras governamentais, que envolve desde o planejamento orçamentário até a avaliação e implementação de projetos, apresenta desafios substanciais (Akaba et al., 2020; Mircea et al., 2022). Estudos na área têm apontado para a necessidade de maior eficiência, transparência e integridade nos processos, além de abordar questões como a gestão de riscos e a superação de barreiras regulatórias que podem impedir a colaboração e a inovação (Ramos & Oliveira, 2023). Autores como Edler & Georghiou (2007) e Uyarra et al. (2014) discutem intensamente o papel das compras públicas como medida de política de inovação no setor privado, enquanto Amann et al. (2014) e Hueskes et al. (2017) exploram a gestão e mensuração da sustentabilidade na cadeia de suprimentos pública.

Um dos debates emergentes na área é o crescente papel da tecnologia, como a Inteligência Artificial (IA), Blockchain e a Internet das Coisas (IoT), na modernização das operações de aquisição pública. Aboelazm & Dganni (2025) e Mircea et al. (2022) preveem um papel relevante para essas tecnologias, capazes de aumentar a competitividade, aprimorar a integridade e expandir a base de participação, além de gerar economias de esforço e dinheiro. Contudo, essa transição tecnológica não é isenta de desafios, exigindo a cooperação humana e a compreensão dos fatores de risco, como observado por Hodosi et al. (2023) no contexto da terceirização de TI para organizações públicas. A literatura também ressalta a importância de se desenvolver teorias científicas a partir da realidade prática para entender os fatores de sucesso desses sistemas e o papel humano na interação com eles (Aboelazm & Dganni, 2025).

Apesar do reconhecimento da importância das compras públicas e do avanço tecnológico, há uma lacuna na pesquisa empírica que quantifique o comportamento e as relações dentro de segmentos específicos do processo de aquisição. Frequentemente, pesquisas são baseadas em relatos ou estudos de caso limitados (Tas, 2020). Há uma necessidade de análises que forneçam insights objetivos, fundamentados em dados, para auxiliar gestores e formuladores de políticas na tomada de decisões. Como apontado por Jaccard & Jacoby (2010), toda investigação requer uma hipótese ou ideia guia, pois "nenhuma observação é puramente empírica". As estatísticas são ferramentas essenciais para transformar dados em informações compreensíveis, desde que a pesquisa seja bem planejada e os dados interpretados corretamente.

A transformação digital do Estado, impulsionada por marcos regulatórios e políticas públicas de inovação, favoreceu o surgimento de plataformas digitais como o Contratos.gov.br<sup>1</sup>. Essas iniciativas passaram a organizar e disponibilizar dados estruturados sobre fornecedores, contratos e órgãos demandantes, criando oportunidades inéditas para análises empíricas sobre o comportamento da

---

<sup>1</sup> Portal oficial do governo federal brasileiro para compras públicas.

Administração Pública e os fatores que influenciam suas decisões contratuais (Akaba et al., 2020; Aboelazm & Dganni, 2025).

Neste contexto, emerge a seguinte questão de pesquisa: **como as características dos contratos e dos fornecedores de TIC se relacionam com o volume, valor e a recorrência de contratações realizadas pelos Ministérios do Poder Executivo Federal no Brasil entre 2020 e 2024?**

Para respondê-la, este artigo adota uma abordagem quantitativa estruturada em duas frentes complementares. A primeira aplica **métodos inferenciais** (como correlações, testes de hipóteses e análises não paramétricas) para verificar associações entre variáveis contratuais e características dos fornecedores. A segunda frente envolve o uso de **técnicas supervisionadas de aprendizado de máquina**, como Random Forest e XGBoost, com o objetivo de prever quais fornecedores têm maior probabilidade de serem recontratados.

O objetivo principal preocupa-se em compreender os fatores que influenciam a recontratação de fornecedores no setor público, além de identificar padrões associados ao valor e volume de contratações. Como contribuição prática, nossos achados podem apoiar estratégias de gestão de riscos, melhoria da concorrência, prevenção de fraudes e aprimoramento dos critérios de avaliação de desempenho contratual.

Com isso, o artigo busca preencher lacunas da literatura sobre seleção e recorrência de fornecedores em compras públicas, contribuindo tanto para o avanço teórico quanto para a aplicação em políticas públicas orientadas por dados.

## 2. REVISÃO DA LITERATURA

Esta seção estabelece o arcabouço teórico sobre análise de compras públicas e sua relação com a Tecnologia da Informação e Comunicação (TIC). Serão brevemente delineadas a relevância estratégica

e os desafios inerentes às aquisições governamentais de TIC, conforme discutido na literatura pertinente. Complementarmente, serão revisados os principais apoios teóricos que sustentam as escolhas das hipóteses e variáveis utilizadas no estudo.

## **2.1 Compras Públicas e Transformação Digital**

As compras públicas são processos vitais para o funcionamento dos governos, gerando valor para os cidadãos (Akaba et al., 2020). Elas são cruciais para promover a competição, melhorar os serviços administrativos e garantir a relação custo-benefício (Tas, 2020). O fluxo típico de um processo de compras públicas envolve a elaboração do orçamento anual, o planejamento da aquisição, a emissão de chamadas de licitação, a avaliação financeira e técnica dos licitantes, e a eventual adjudicação e implementação dos projetos (Akaba et al., 2020). No entanto, a literatura aponta para a necessidade de mais pesquisas empíricas sobre o impacto da qualidade da regulamentação das compras públicas nos resultados econômicos (Tas, 2020). A evolução das demandas sobre as Forças Armadas e as mudanças correspondentes nos requisitos de capacidade de suprimento, por exemplo, ilustram a complexidade e a necessidade de adaptação contínua nos processos de compras (Hellberg, 2023). A adesão rigorosa às regulamentações, como as normas da União Europeia e da Suécia, por exemplo, para compras públicas, é um desafio constante (Hellberg, 2023). Além disso, a área de compras públicas não está imune a desafios como a detecção de fraudes e a avaliação de riscos de corrupção (Velasco et al. 2020).

A sustentabilidade também tem se tornado um foco crescente nas compras públicas, com estudos explorando como as parcerias público-privadas (PPP) podem estimular práticas sustentáveis em projetos de infraestrutura (Hueskes et al., 2017) e como impulsionar a gestão sustentável da cadeia de suprimentos (Amann et al., 2014).

Márquez Lozada & Vela Aspajo (2024) realizaram uma revisão sistemática da literatura sobre gestão de contratos em entidades públicas, destacando os avanços nos marcos regulatórios,

transparência, eficiência e o papel da tecnologia nos processos de compras públicas de 2019 a 2023, enfatizando a importância dos controles internos e da participação cidadã. Entre os achados dessa revisão, descrevem que a tecnologia automatizou tarefas operacionais, otimizando o tempo dos profissionais para atividades estratégicas e permitindo o acesso em tempo real aos dados de aquisição. Além disso, possibilitou uma gestão baseada em dados e ampliou os mecanismos de controle social. No entanto, a implementação enfrenta desafios como resistência à mudança e necessidade de capacitação dos envolvidos (Márquez Lozada & Vela Aspajo, 2024).

## **2.2 Compras Públicas e Reconstrução**

A literatura sobre compras públicas destaca seu papel estratégico na indução de inovação, sustentabilidade e eficiência governamental (Edler & Georghiou, 2007; Uyarra et al., 2014). Além de serem responsáveis por uma parcela significativa do PIB em diversos países, as aquisições públicas também influenciam diretamente a estrutura dos mercados e o comportamento dos fornecedores (Amann et al., 2014; Ramos & Oliveira, 2023). Assim, compreender os fatores que influenciam o desempenho e a recorrência de fornecedores é fundamental para o aprimoramento das políticas públicas.

Segundo Akaba et al. (2020), a adoção de sistemas digitais e plataformas de e-procurement aumenta a transparência, reduz custos operacionais e fortalece os mecanismos de integridade na Administração Pública. O uso estratégico dos dados gerados por esses sistemas oferece subsídios relevantes para a análise preditiva do comportamento contratual (Velasco et al., 2020), possibilitando identificar padrões de eficiência, riscos e oportunidades de melhoria.

Estudos como os de Koala e Steinfeld (2018) e Malacina et al. (2022) defendem a construção de teorias baseadas em dados e na prática institucional como forma de ampliar a compreensão sobre os determinantes das decisões públicas. O presente estudo segue essa abordagem, empregando métodos inferenciais e técnicas supervisionadas de aprendizado de máquina (como Random Forest e XGBoost) para



investigar quais fatores impactam a recontratação de fornecedores no setor público brasileiro entre 2020 e 2024.

A recontratação pode ser interpretada como um indicador de confiança institucional, eficiência ou, em alguns casos, falta de concorrência ou risco de favorecimento (Lenderink, 2018; Tas, 2020). Hodosi et al. (2023) apontam que, no caso da terceirização de serviços de TI, a continuidade contratual está frequentemente associada à estabilidade, conhecimento acumulado e custos de transação reduzidos. No entanto, outros fatores como porte, localização, idade da empresa e histórico com múltiplos ministérios também influenciam a probabilidade de recontratação (Caserta et al., 2025).

A literatura sobre gestão de fornecedores no setor público também evidencia que empresas maiores tendem a possuir mais recursos para competir em licitações, cumprir exigências burocráticas e oferecer preços mais competitivos, o que pode impactar positivamente sua presença recorrente (Bergman & Lundberg, 2013; Aboelazm & Dganni, 2025). Além disso, características regionais e setoriais – como a concentração econômica no Sudeste e no Distrito Federal – influenciam a distribuição dos contratos públicos no Brasil (Maciel et al., 2024; Ramos & Oliveira, 2023).

Com base nesses estudos, foram formuladas **hipóteses** relacionadas às características dos fornecedores (como porte, localização, idade e capital social), bem como aos contratos firmados com os ministérios. Tais hipóteses, apresentadas na Tabela 1, orientam a análise quantitativa proposta neste artigo.

***Tabela 1 - Hipóteses e Apoio Teórico***

Hipótese	Formulação	Fundamentação teórica (com base nos artigos enviados)	Técnica Estatística Proposta
H1	Ministérios com maior número de contratos possuem valores médios de contratos mais baixos.	A dispersão orçamentária entre ministérios pode levar à pulverização de valores, conforme sugerido por Caserta et al. (2025), ao apontarem ineficiências associadas ao excesso de	Mann-Whitney

		contratos em estruturas burocráticas grandes.	
<b>H2</b>	Fornecedores com maior idade (tempo de existência) possuem contratos de maior valor.	Fornecedores mais antigos tendem a apresentar maior reputação e confiança institucional, favorecendo sua seleção (Bergman & Lundberg, 2013; Hodosi et al., 2023).	Correlação de Spearman
<b>H3</b>	Fornecedores sancionados (presentes nas listas CNEP <sup>2</sup> ou CEIS <sup>3</sup> ) possuem contratos de menor valor.	Conforme apontado por Velasco et al. (2020), restrições reputacionais e indícios de fraude impactam negativamente o volume contratual de empresas públicas.	Mann-Whitney
<b>H4</b>	Empresas de maior porte (classificação oficial) possuem contratos de maior valor.	O porte influencia a capacidade de atendimento e entrega de contratos complexos (Baptista, 2025; Amann et al., 2014).	Kruskal-Wallis
<b>H5</b>	Fornecedores localizados na Região Sudeste possuem contratos de maior valor.	Não há evidência direta nos artigos compartilhados. Esta hipótese é contextual, baseada em indicadores econômicos externos (IBGE, BNDES) e será apresentada como suposição a ser testada empiricamente.	Mann-Whitney
<b>H6</b>	Fornecedores localizados no Distrito Federal (DF) possuem contratos de maior valor.	A presença institucional do governo federal em Brasília concentra licitações estratégicas, conforme indicado por Maciel et al. (2024).	Mann-Whitney
<b>H7</b>	A proporção de contratos do Sudeste é maior em relação às demais regiões.	A centralização da infraestrutura e TIC no Sudeste é relatada por Akaba (2020), ao analisar o desequilíbrio de adoção de e-procurement.	Teste de Proporções
<b>H8</b>	A proporção de contratos do Distrito Federal (DF) é maior.	Justificativa baseada na localização central dos órgãos da Administração Pública Federal (Maciel et al., 2024).	Teste de Proporções
<b>H9</b>	Maior Capital Social das empresas possui correlação positiva com o número de contratos celebrados.	A capitalização da empresa reflete maior capacidade operacional, o que pode influenciar sua atratividade (Koala & Steinfeld, 2018).	Correlação de Spearman
<b>H10</b>	Maior Capital Social das empresas possui correlação positiva com o valor dos contratos.	Relacionamento similar ao observado em H9, com base na robustez estrutural (Koala & Steinfeld, 2018).	Correlação de Spearman

<sup>2</sup> CNEP (Cadastro Nacional das Empresas Punidas) registra as empresas que sofreram punições com base na Lei Anticorrupção (Lei nº 12.846/2013), incluindo acordos de leniência.

<sup>3</sup> CEIS (Cadastro Nacional de Empresas Inidôneas e Suspensas) lista empresas e pessoas físicas impedidas de participar de licitações ou celebrar contratos com o governo devido a sanções. Os dados do CNEP e o CEIS estão disponíveis em <https://portaldatransparencia.gov.br/sancoes>.

<b>H11</b>	Valor do contrato influencia recorrência/recontratação do fornecedor.	Contratos de maior valor tendem a estabelecer relações mais duradouras, como discutido por Edler & Georghiou (2007) e Mircea (2022).	Regressão Logística
<b>H12</b>	Tempo e porte influenciam recorrência/recontratação do fornecedor.	A reputação derivada do tempo de operação e porte é um fator crítico na avaliação pública (Hueskes, Verhoest & Block, 2017).	Regressão Logística
<b>H13</b>	Diversidade de ministérios contratantes influencia recorrência/recontratação do fornecedor.	A diversidade contratual reflete confiança generalizada, aumentando as chances de reconstrução (Aboelazm & Dganni, 2025).	Mann-Whitney

Fonte: Elabora pelos autores

A análise do comportamento e das decisões de contratação pública exige respaldo em fundamentos teóricos consistentes que justifiquem a escolha das hipóteses e variáveis de interesse. Conforme destacado por Cooper e Schindler (2014), a formulação de hipóteses em pesquisas quantitativas deve estar ancorada em teorias prévias, revisões sistemáticas da literatura e observações empíricas. O presente estudo estrutura suas proposições com base em autores que analisam criticamente o papel das compras públicas como instrumento de política pública (Amann et al., 2014; Edler & Georghiou, 2007; Uyarra et al., 2014) e que exploram a eficiência e a recorrência de fornecedores sob diferentes abordagens (Tas, 2022; Ramos & de Oliveira, 2023).

No contexto brasileiro, a concentração regional de fornecedores, a diversidade contratual entre ministérios e a variação nos perfis empresariais — como idade, porte e capital social — são fatores apontados como influentes na dinâmica das contratações públicas (Maciel et al., 2024; Baptista, 2025; Caserta et al., 2025). Além disso, a recorrência de fornecedores tem sido interpretada tanto como indicativo de confiabilidade quanto de possíveis práticas de favorecimento (Velasco et al., 2020). Essa ambiguidade reforça a necessidade de estudos que combinem rigor estatístico com amparo teórico.

A literatura recente também enfatiza o uso de abordagens quantitativas avançadas, como machine learning, na análise preditiva de padrões em compras públicas (Aboelazm & Dganni, 2025; Kundu

et al., 2024). Entretanto, conforme ressaltado por Cumming (2012) e Field et al. (2012), mesmo essas abordagens devem partir de uma estrutura conceitual clara, que oriente a seleção de variáveis e a interpretação de resultados.

### 3. METODOLOGIA

Este estudo adota uma abordagem quantitativa, explicativa e empírico-analítica, com o objetivo de identificar fatores associados à recontratação de fornecedores no âmbito das compras públicas de Tecnologia da Informação e Comunicação (TIC). Para isso, foram utilizadas estatísticas descritivas, testes inferenciais e modelos supervisionados de aprendizado de máquina, conforme detalhado a seguir.

#### 3.1 Fontes de Dados

Os dados primários para esta pesquisa são provenientes do portal [contratos.gov.br](https://contratos.gov.br), especificamente os arquivos '**contratos\_ministerios\_consolidados\_separados.csv**', que contêm a série histórica de contratos de 25 ministérios do poder executivo federal de 2020 a 2024, já consolidados e tratados. Importante ressaltar que foram pesquisados todos os 43 códigos SIAFI de ministérios do poder executivo federal. No entanto, desse total, apenas 25 possuíam contratos na categoria "Informática (TIC)". O resultado atingiu, portanto, 1.088 contratos de 25 ministérios vinculados a 479 fornecedores únicos.

Adicionalmente, foi utilizada uma amostra de 150 fornecedores (arquivo '**fornecedores\_formatado.csv**') com dados enriquecidos pela base CNPJ<sup>4</sup> – Plataforma de consulta em massa de CNPJs de empresas.

Além dessas fontes de dados, foram extraídas séries temporais de dados do Cadastro Nacional de Empresas Inidôneas e Suspensas – CEIS (arquivo '**20250526\_CEIS\_Emp\_Inidoneas\_Suspensas.csv**') e do

---

<sup>4</sup> A plataforma CNPJ<sup>4</sup> (<https://cnpj.com/>) é um serviço online que oferece consultas de informações de CNPJ (Cadastro Nacional da Pessoa Jurídica) de empresas. Ela permite obter dados de diversos órgãos públicos, como Receita Federal, Simples Nacional, Sintegra e Suframa, tudo em um só lugar. A plataforma oferece acesso via painel web, API e add-in para Excel.

Cadastro Nacional de Empresas Punidas – CNEP (arquivo ‘**20250526\_CNEP\_Empresas\_Punidas.csv**’) para o mesmo período (2020 a 2024) aplicado aos 479 fornecedores. A Tabela 2 lista as principais variáveis utilizadas no estudo.

**Tabela 2. Principais variáveis do estudo**

<b>Interesse</b>	<b>Nome</b>	<b>Tipo</b>	<b>Escala</b>
Dependente (Variável-resposta)	Log_Valor_Global (Logaritmo do Valor Global do contrato)	Quantitativa Contínua	Escala de razão (após transformação logarítmica)
Independente (Variável explicativa)	CNPJ do Fornecedor	Qualitativa (Identificador)	Nominal
Independente (Variável explicativa)	Órgão Contratante	Qualitativa	Nominal
Independente (Variável explicativa)	Data de Início do Contrato	Quantitativa (transformada em Ano ou em variável temporal)	Ordinal
Independente (Variável explicativa)	Data de Fim do Contrato	Quantitativa (opcional, ou usada para calcular duração)	Ordinal
Independente (Variável explicativa)	Número de Parcelas	Quantitativa Discreta	Razão
Independente (Variável explicativa)	Porte da Empresa (MEI, Pequena, Média, Grande, etc.)	Qualitativa	Ordinal
Independente (Variável explicativa)	Idade da Empresa (em anos)	Quantitativa Contínua	Razão
Independente (Variável explicativa)	Situação Cadastral da Empresa (Ativa/Inativa)	Qualitativa	Nominal
Independente (Variável explicativa)	Capital Social (em R\$)	Quantitativa Contínua	Razão
Independente (Variável explicativa)	Município da Sede da Empresa	Qualitativa	Nominal
Independente (Variável explicativa)	UF da Sede da Empresa	Qualitativa	Nominal
Independente (Variável explicativa)	Presença no CEIS (Cadastro de Empresas Inidôneas e Suspensas)	Qualitativa Binária	Nominal (Sim/Não)

Independente (Variável explicativa)	Presença no CNEP (Cadastro Nacional de Empresas Punidas)	Qualitativa Binária	Nominal (Sim/Não)
----------------------------------------	----------------------------------------------------------------	------------------------	-------------------

Fonte: Elaborado pelos autores

### 3.2 Coleta e Preparação dos Dados

A preparação dos dados incluiu a consolidação e tratamento inicial dos arquivos, garantindo sua adequação para análise. Foi realizado um diagnóstico de qualidade dos dados, verificando a ausência de problemas graves de valores ausentes (*missing values*) e explorando variáveis financeiras (Valor Global, Valor Parcela) e categóricas.

Nessa etapa, foram realizadas filtragens e verificações para ratificar o período dos contratos, sendo que para um contrato ser aceito, esse deveria ter duas premissas válidas: a) que sua data fim de vigência fosse após 2019 (a partir de 2020) e; b) sua data de início fosse anterior a 2025 (até 2024). Com isso, a janela temporal de contratos válidos para análise seria resguardada entre 2020 a 2024. Com isso, o número total de contratos caiu para 1.008 e o de fornecedores para 449 fornecedores. A exclusão dos 80 contratos não-válidos para o estudo afetou apenas 7 registros de fornecedores da amostra, o que a fez computar 143 fornecedores.

### 3.3 Análise Exploratória de Dados e Caminho Metodológico

Antes da realização dos testes estatísticos e da aplicação dos modelos preditivos, foi conduzida uma Análise Exploratória de Dados (AED) com o objetivo de compreender a distribuição das variáveis, detectar valores ausentes, identificar possíveis outliers e definir os procedimentos estatísticos adequados.

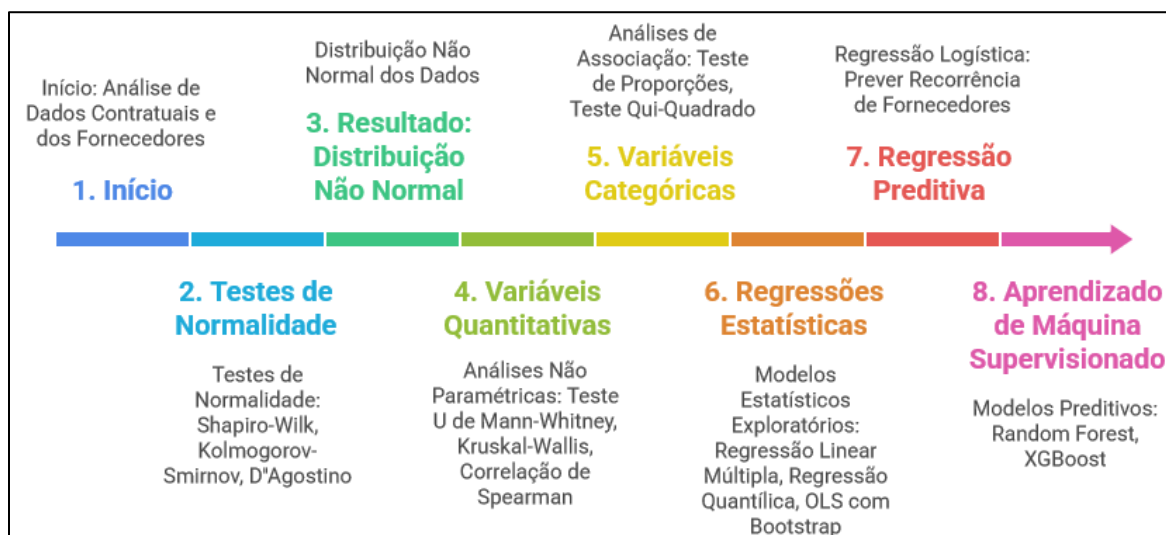
A normalidade das variáveis contínuas (como "Valor Global", "Capital Social" e "Idade da Empresa") foi avaliada por meio do teste de Shapiro-Wilk, recomendável para amostras pequenas e médias (Field et al., 2012; Triola, 2010). Os resultados indicaram que a maioria das variáveis não seguia uma distribuição normal, justificando, assim, o uso de testes não paramétricos, como Mann-Whitney e

Kruskal-Wallis, e correlações de Spearman para investigar associações entre as variáveis de interesse (Tabachnick & Fidell, 2001; Cooper & Schindler, 2014).

A análise estatística neste estudo foi conduzida em múltiplas etapas organizadas metodologicamente, com o objetivo de identificar padrões, testar hipóteses e construir modelos preditivos sobre a recontratação de fornecedores no setor público brasileiro. Essas etapas foram definidas com base nas melhores práticas da literatura em métodos quantitativos (Cooper & Schindler, 2014; Field et al., 2012; Cumming, 2012), combinando análise exploratória, testes não paramétricos, regressões robustas e algoritmos de aprendizado supervisionado.

A Figura 1 apresenta o fluxo metodológico seguido, estruturado desde a coleta e preparação dos dados até a aplicação de modelos preditivos. Essa visualização sintetiza a lógica analítica do estudo, ilustrando como cada etapa se conecta à seguinte em um processo cumulativo e iterativo de investigação empírica.

**Figura 1 - Caminho metodológico da análise de dados.**



Fonte: Elaborado pelos autores com apoio da ferramenta Napkin AI

### **Estatísticas Descritivas e Análise Exploratória de Dados (AED)**

A primeira etapa consistiu na Análise Exploratória de Dados (AED), com o objetivo de examinar a distribuição, dispersão e comportamento geral das variáveis. Foram geradas estatísticas descritivas (média, mediana, desvio padrão, quartis), histogramas, gráficos de dispersão e boxplots, tanto para os contratos quanto para os fornecedores. Essa etapa permitiu, por exemplo, observar a concentração de contratos em determinadas regiões (como o Sudeste), a alta variabilidade dos valores globais dos contratos e o padrão de distribuição entre ministérios contratantes.

### **Testes de Normalidade**

Com base na literatura estatística (Field et al., 2012; Cumming, 2012), foi conduzida uma verificação formal da normalidade dos dados quantitativos. Foram aplicados os testes de Shapiro-Wilk, Kolmogorov-Smirnov e D'Agostino-Pearson, que demonstraram, de forma consistente, que a maioria das variáveis de interesse apresentava distribuições não normais. Esse resultado norteou a escolha por técnicas não paramétricas nas etapas subsequentes, conferindo maior robustez à análise.

### **Formulação e Testes de Hipóteses Específicas**

Com base na revisão da literatura e nos objetivos da pesquisa, foram formuladas 13 hipóteses específicas (ver Tabela 1), associando características dos fornecedores e dos contratos com os padrões de valor, volume e recontratação. Para testá-las, foram aplicadas as seguintes técnicas:

- **Teste de Mann-Whitney U:** para comparar diferenças de distribuições entre dois grupos independentes;
- **Teste de Kruskal-Wallis:** para comparações entre mais de dois grupos;
- **Correlação de Spearman:** para avaliar relações monotônicas entre variáveis quantitativas;



- **Teste de Proporções e Qui-quadrado de Pearson:** para avaliar associações entre variáveis categóricas.

### **Medida de Tamanho do Efeito (Effect Size)**

Complementarmente à significância estatística, buscou-se avaliar a magnitude prática dos efeitos observados por meio de medidas de tamanho do efeito. Para isso, foram utilizados indicadores como o  $r$  de efeito (em testes de Mann-Whitney),  $\eta^2$  (para Kruskal-Wallis) e  $\epsilon^2$ , possibilitando uma interpretação mais rica dos resultados inferenciais nas hipóteses com diferenças estatisticamente significativas.

### **Análise de Outliers**

Durante a AED, foram identificados outliers em variáveis como valor global dos contratos, capital social e número de ministérios contratantes. Apesar disso, optou-se por manter os valores originais, com uso de técnicas estatísticas robustas como Regressão Quantílica e modelos de árvore (Random Forest), que são menos sensíveis à presença de valores extremos e permitem captar melhor os padrões da base de dados.

### **Modelagem Multivariada**

Para aprofundar a análise e isolar o efeito de múltiplas variáveis sobre a recontração, foi construída uma sequência de modelos multivariados:

- Regressão Linear Múltipla (OLS) como ponto de partida exploratório;
- Regressão Quantílica, para avaliar impactos ao longo dos diferentes quantis da variável dependente;

- OLS com Bootstrap, que permitiu estimar intervalos de confiança robustos, mitigando o efeito de heterocedasticidade e viés amostral.

### **Modelos Preditivos e Machine Learning**

Com o objetivo de prever a recontratação de fornecedores, foram aplicados dois modelos supervisionados de aprendizado de máquina: Random Forest e XGBoost. Ambos foram treinados com técnicas de validação cruzada estratificada e avaliados por meio de métricas como acurácia, precisão, recall e curva ROC.

Esses modelos também possibilitaram:

- A geração de *Feature Importance*<sup>5</sup>, identificando as variáveis com maior influência na recontratação;
- A construção de Partial Dependence Plots (PDPs), permitindo a visualização dos efeitos marginais das variáveis preditoras (quantitativas e categóricas);

Tentativas de aplicação do SHAP (SHapley Additive exPlanations) foram realizadas, mas apresentaram limitações técnicas devido à codificação de variáveis categóricas e ao balanceamento da base. Por esse motivo, a técnica foi descartada nesta fase analítica.

### **Softwares e Ferramentas Utilizadas**

Todas as análises estatísticas, manipulação de dados e visualizações foram implementadas utilizando a linguagem de programação Python, com apoio das bibliotecas pandas, numpy, matplotlib,

---

<sup>5</sup> Técnica usada em aprendizado de máquina para determinar a relevância ou importância de cada característica (ou variável) na previsão de um modelo. Em outras palavras, ela mede quanto cada característica contribui para o resultado do modelo.

seaborn e scipy.stats, além dos pacotes especializados para machine learning (scikit-learn, xgboost) e explicabilidade de modelos (pdpbox, entre outros).

#### **4. RESULTADOS**

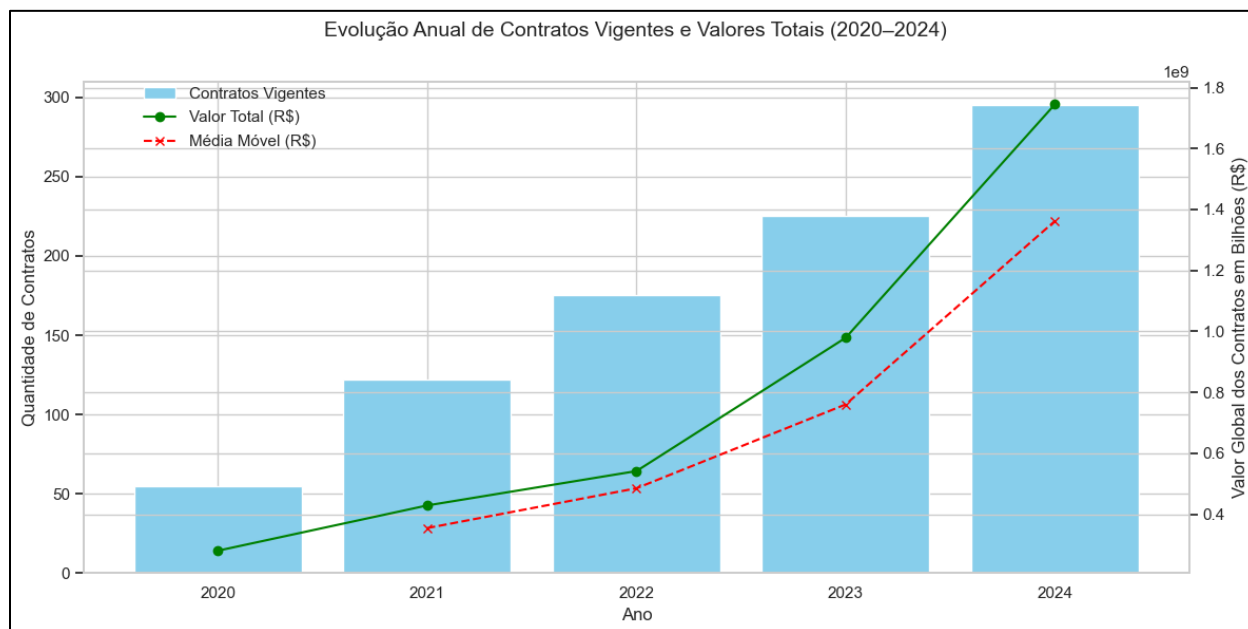
Esta seção apresenta os principais resultados obtidos com base na aplicação dos métodos estatísticos e modelos de aprendizado de máquina, com o objetivo de responder à pergunta de pesquisa: quais características dos contratos e dos fornecedores de TIC influenciam o valor, volume e recorrência de contratações públicas?

##### **4.1 Análise Exploratória dos Dados**

A análise exploratória permitiu uma visão inicial do comportamento dos dados, fundamental para subsidiar a definição das técnicas estatísticas aplicadas nas etapas subsequentes. A base consolidada utilizada nesta etapa incluiu 1.008 contratos vigentes no período de 2020 a 2024, firmados com 143 fornecedores distintos.

Observando a distribuição temporal dos contratos, identificou-se um crescimento expressivo no número total de contratos a partir de 2021, com um pico em 2022, seguido por uma ligeira retração em 2024. Essa tendência é apresentada na Figura 2, que mostra a evolução anual da quantidade de contratos celebrados e o valor total contratado ao longo do período analisado.

**Figura 2 - Evolução Anual da Quantidade e do Valor Total de Contratos de TIC (2020–2024)**



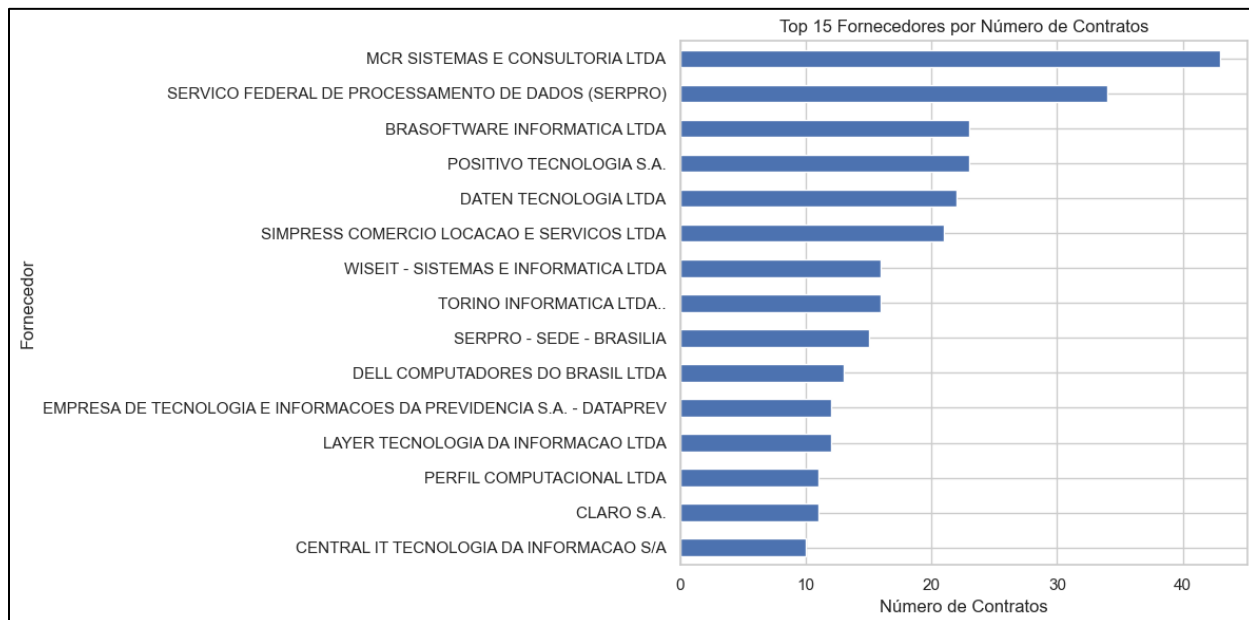
Fonte: Elaborado pelos autores

A Figura 3 apresenta o ranking dos 15 fornecedores com maior número de contratos firmados no período, enquanto a Figura 4 representa o ranking dos 15 fornecedores com maiores valores totais contratados. Evidencia-se a concentração dos contratos em um pequeno grupo de empresas. Notadamente, os cinco fornecedores mais recorrentes respondem por uma fração substancial dos contratos analisados.

Já as Figuras 5 e 6 revelam os 15 ministérios que mais contrataram, tanto em volume quanto em número de contratos. Observa-se predominância de contratos originados nos ministérios ligados à economia, tecnologia e saúde.

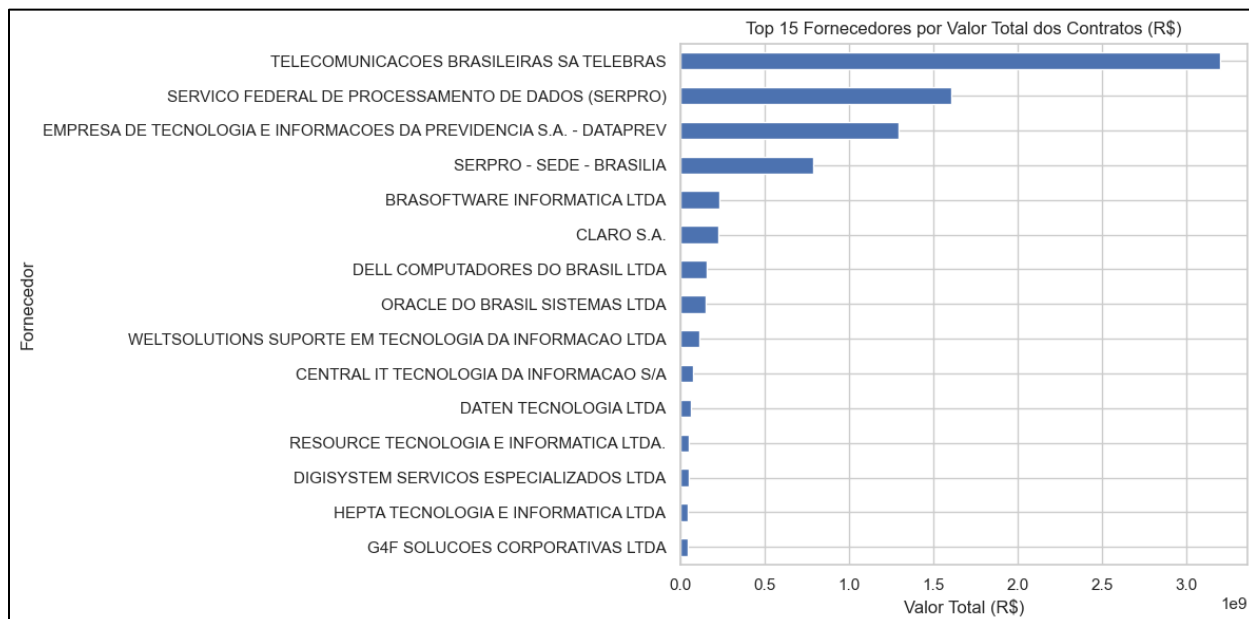
Para aprofundar a análise inferencial, foi extraída uma amostra com 249 contratos correspondentes aos 143 fornecedores mapeados. As Figuras 7, 8, 9 e 10 reproduzem os mesmos rankings de fornecedores e ministérios para esse subconjunto, permitindo comparar tendências entre a base completa e a amostra selecionada para os testes estatísticos.

**Figura 3 - Top 15 Fornecedores por Número de Contratos (Base Completa – 1008 contratos)**



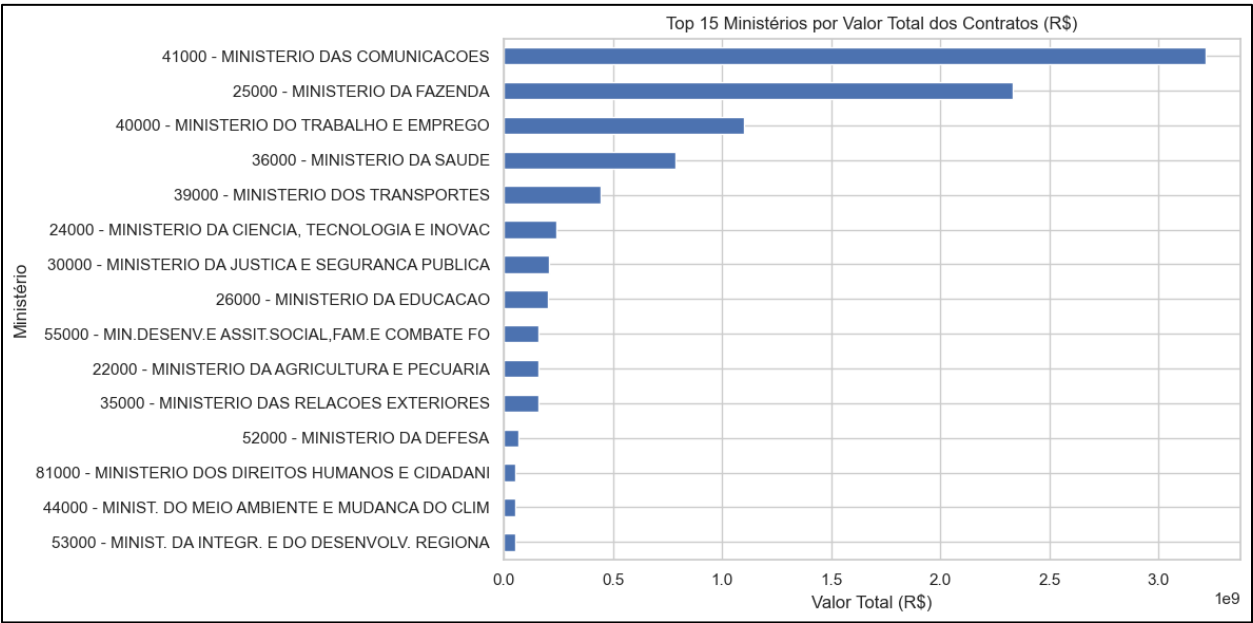
Fonte: Elaborado pelos autores

**Figura 4 - Top 15 Fornecedores por Valor Total dos Contratos (Base Completa – 1008 contratos)**



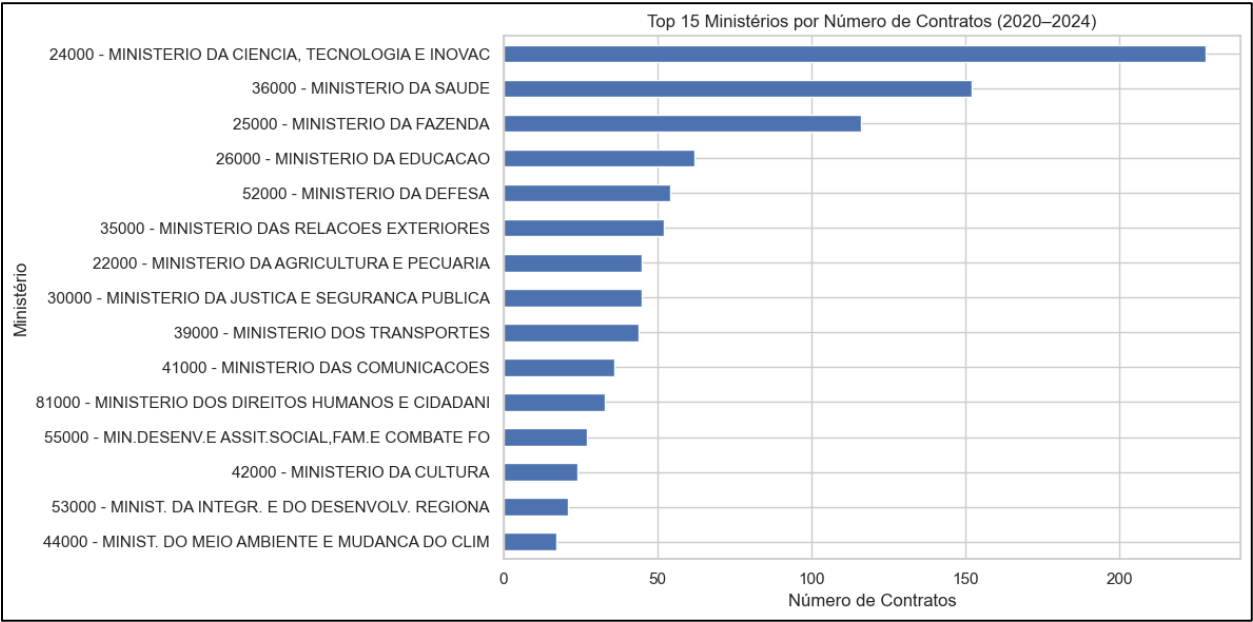
Fonte: Elaborado pelos autores

**Figura 5 - Top 15 Ministérios por Valor Total Contratado (Base Completa – 1008 contratos)**



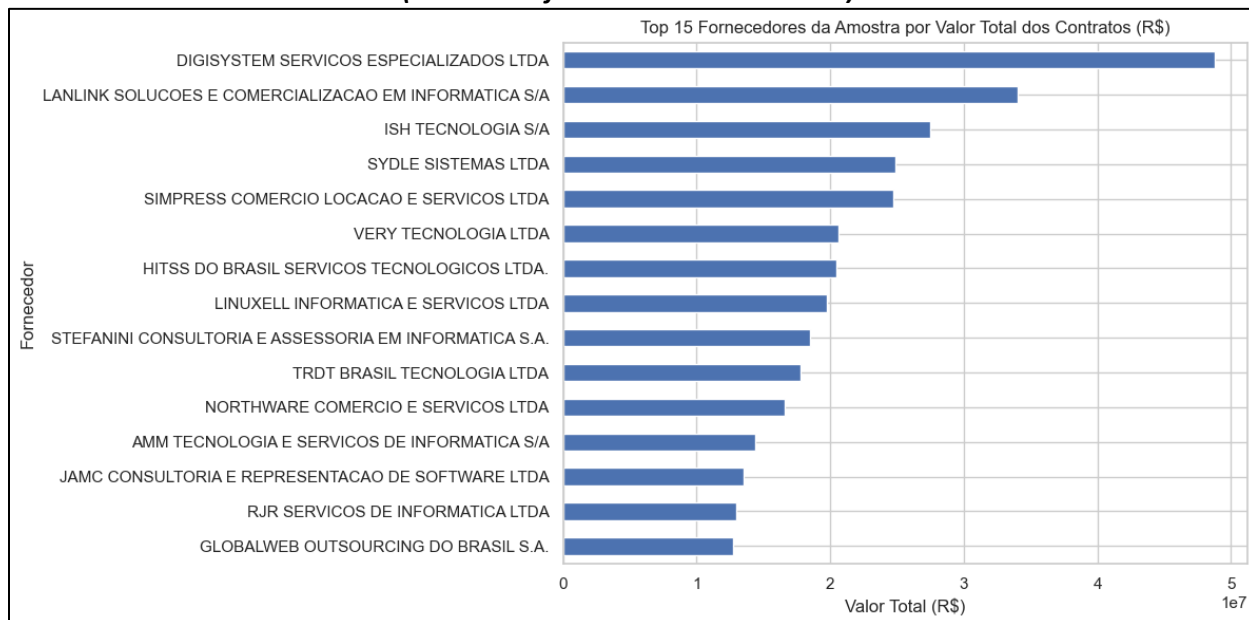
Fonte: Elaborado pelos autores

**Figura 6 - Top 15 Ministérios por Número de Contratos (Base Completa – 1008 contratos)**

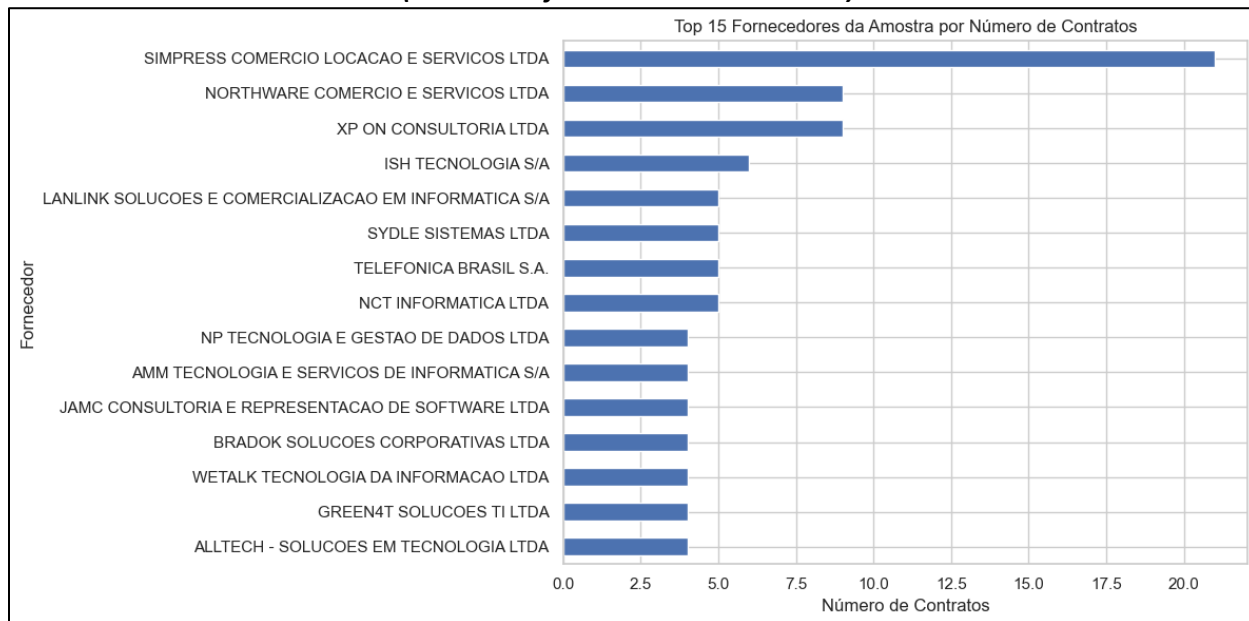


Fonte: Elaborado pelos autores

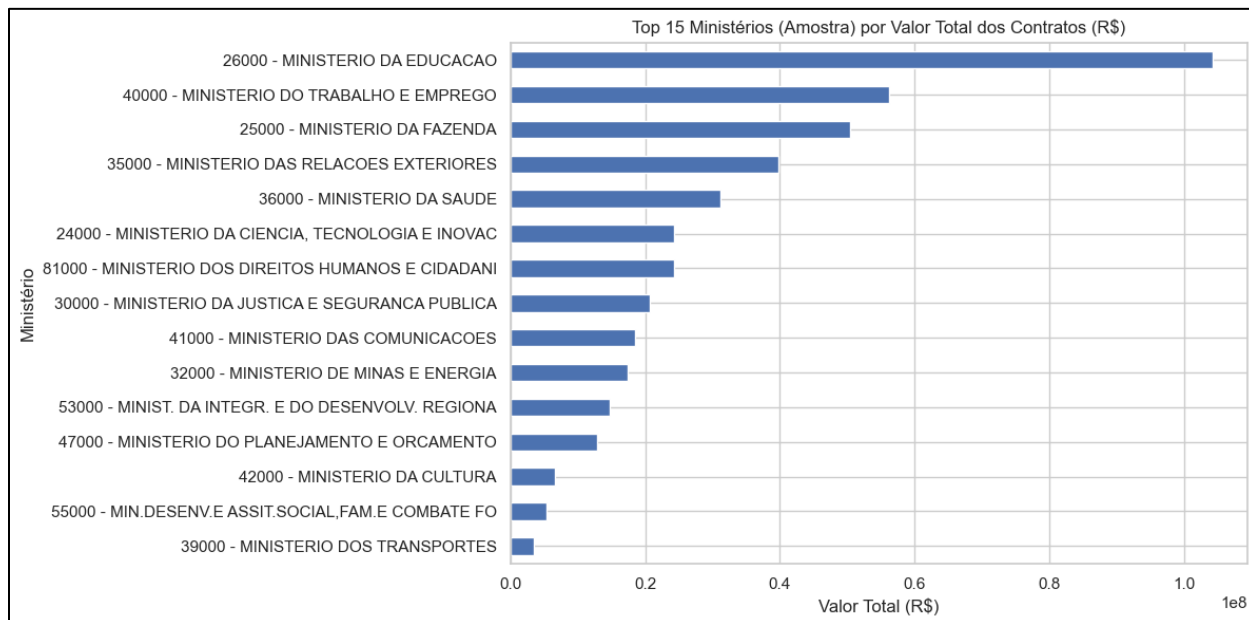
**Figura 7 - Top 15 Distribuição de Fornecedores – Valor Total dos Contratos  
(Amostra Inferencial – 249 contratos)**



**Figura 8 - Top 15 Distribuição de Fornecedores – Número de Contratos  
(Amostra Inferencial – 249 contratos)**

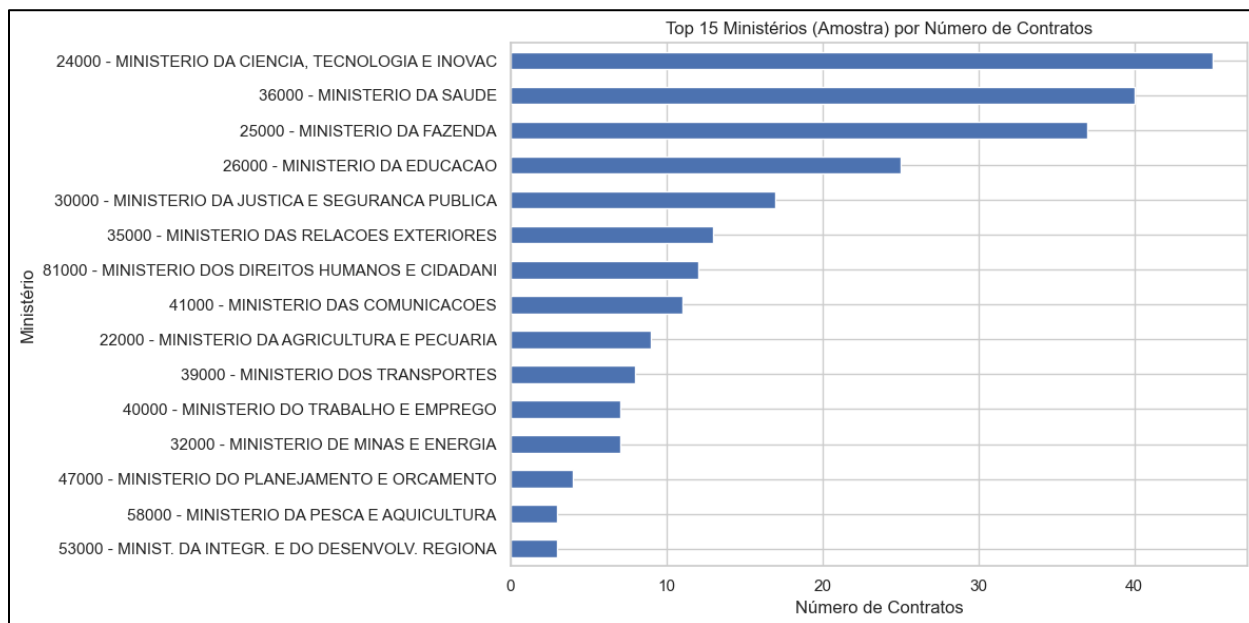


**Figura 9 - Top 15 Ministérios por Valor Total Contratado (Amostra Inferencial – 249 contratos)**



Fonte: Elaborado pelos autores

**Figura 10 - Top 15 Ministérios por Número de Contratos (Amostra Inferencial – 249 contratos)**



Fonte: Elaborado pelos autores



Além disso, análises univariadas e bivariadas indicaram elevada assimetria e curtose em variáveis como “Valor Global”, “Capital Social” e “Número de Ministérios Contratantes”, sugerindo a presença de outliers e distribuições não normais — o que reforçou a decisão pelo uso de testes não paramétricos.

A Tabela 3 apresenta as estatísticas descritivas da variável “Valor Global” para ambas as bases. Observa-se que a média do valor dos contratos na base completa (R\$ 9,28 milhões) é significativamente superior à da amostra inferencial (R\$ 1,75 milhão), refletindo a presença de outliers com valores extremamente altos na base geral. O desvio padrão elevado (R\$ 108,8 milhões) reforça essa dispersão dos valores na base completa, enquanto a amostra inferencial apresenta maior homogeneidade relativa (desvio padrão de R\$ 4,36 milhões).

***Tabela 3 - Estatísticas Descritivas da Variável "Valor Global" (Base Completa e Amostra Inferencial)***

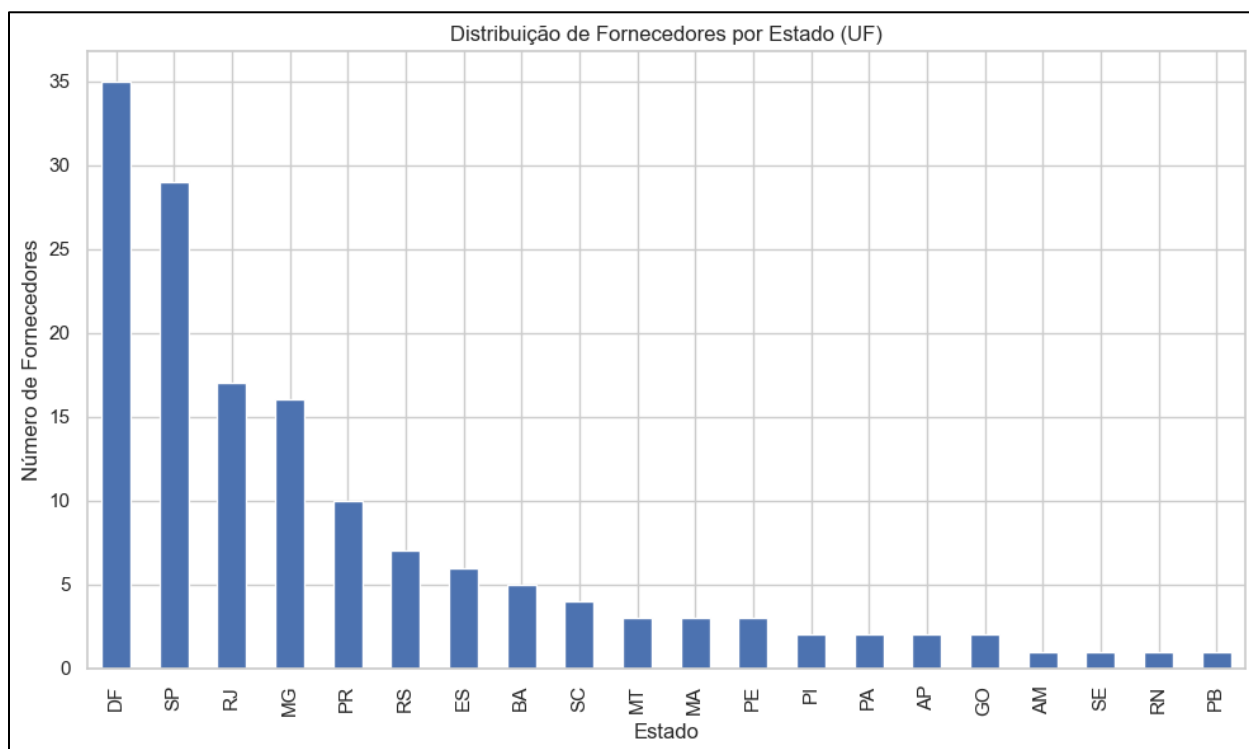
Medida Estatística	Base Completa (n=1008)	Amostra Inferencial (n=248)
<b>Média</b>	R\$ 9.282.198	R\$ 1.754.830
<b>Desvio padrão</b>	R\$ 108.895.600	R\$ 4.364.545
<b>Valor mínimo</b>	R\$ 0,00	R\$ 0,01
<b>1º Quartil (Q1 - 25%)</b>	R\$ 64.000	R\$ 45.123,47
<b>Mediana (Q2 - 50%)</b>	R\$ 331.425,90	R\$ 199.610,70
<b>3º Quartil (Q3 - 75%)</b>	R\$ 1.933.727	R\$ 1.296.453
<b>Valor máximo</b>	R\$ 3.125.903.000	R\$ 41.885.820

Fonte: Elaborado pelos autores

Além disso, a mediana dos contratos é muito inferior à média em ambos os conjuntos (R\$ 331 mil na base completa e R\$ 199 mil na amostra), indicando assimetria à direita na distribuição — fato confirmado pelos histogramas e boxplots gerados na etapa de AED. A diferença entre os valores máximos também é expressiva, com contratos de até R\$ 3,1 bilhões na base geral e até R\$ 41,8 milhões na amostra. Esse padrão sugere que a base completa inclui contratos de grande escala não capturados integralmente na amostra inferencial.

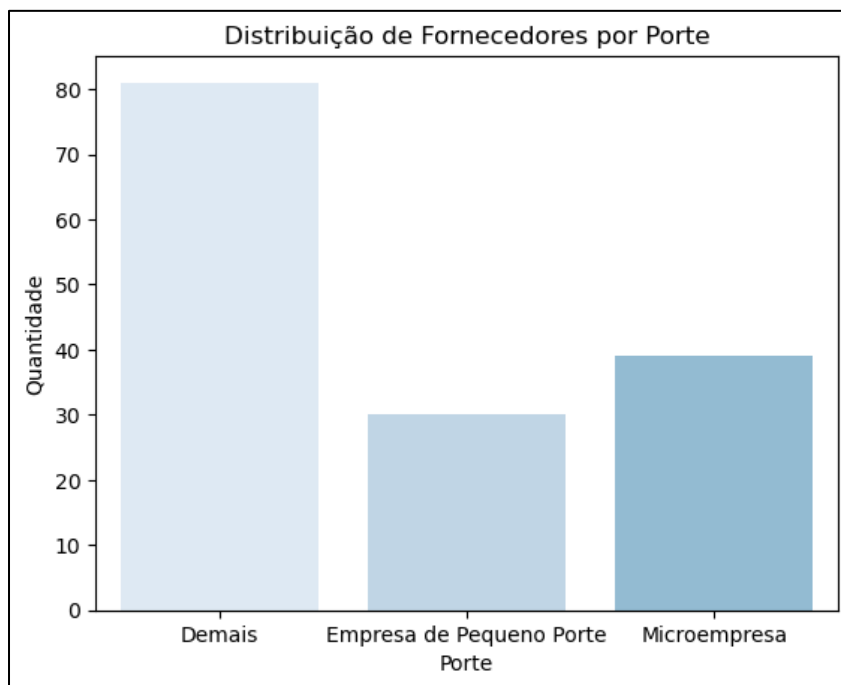
Considerando os dados enriquecidos dos 150 fornecedores, a partir da base CNPJá, obteve-se a distribuição de contratos regionalizada por Unidades da Federação (UF), conforme ilustra a Figura 11. Além disso, também foi possível obter a distribuição do porte das empresas e fornecedores com cadastro de sanções, apresentadas pelas Figuras 12 e 13 respectivamente.

**Figura 11 - Distribuição de Fornecedores por Estado (UF)**



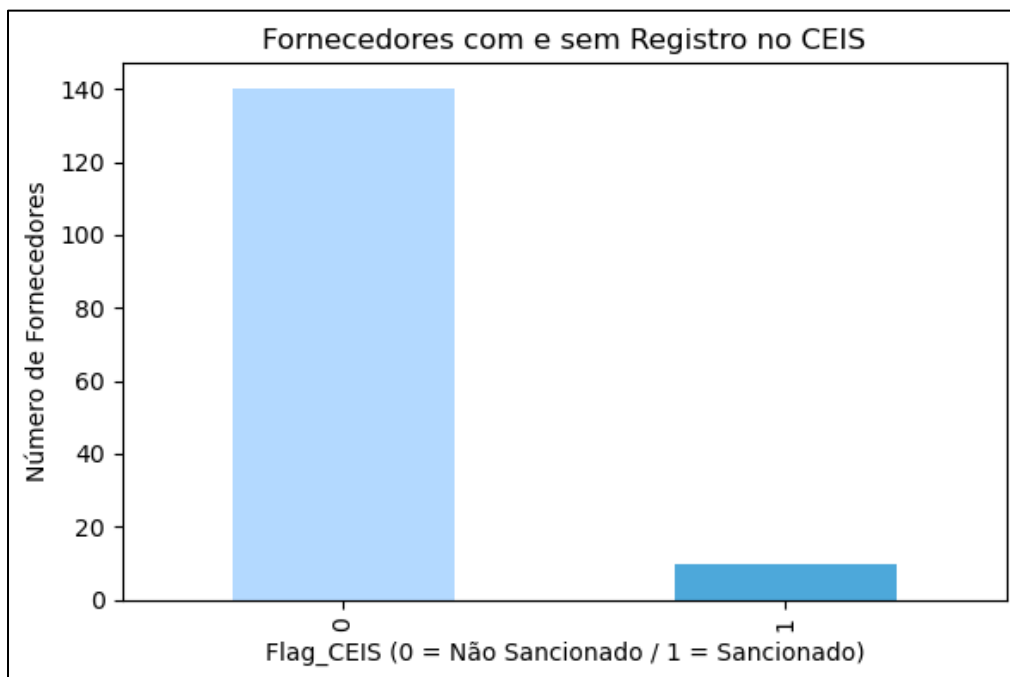
Fonte: Elaborado pelos autores

**Figura 12 - Distribuição de Fornecedores por Porte**



Fonte: Elaborado pelos autores

**Figura 13 - Distribuição de Fornecedores com e sem registro no CEIS**



Fonte: Elaborado pelos autores

#### 4.2 Testes de Normalidade

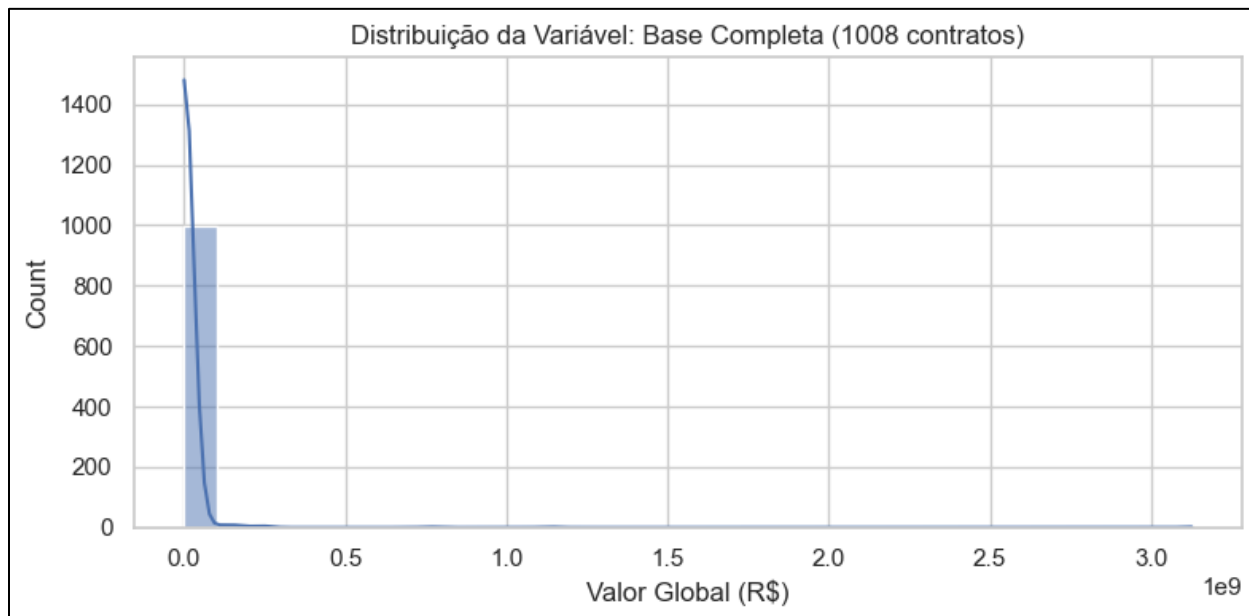
Antes da aplicação de testes estatísticos inferenciais, foi necessário verificar a distribuição das variáveis quantitativas, especialmente o Valor Global dos contratos, variável central nas análises deste estudo. Para isso, foram aplicados três testes de normalidade amplamente utilizados na literatura estatística: Shapiro-Wilk, Kolmogorov-Smirnov e D'Agostino-Pearson (Field et al., 2012; Cumming, 2012)

Os testes de Shapiro-Wilk, Kolmogorov-Smirnov e D'Agostino-Pearson foram aplicados tanto à base completa (1008 contratos) quanto à amostra inferencial (248 contratos de 143 fornecedores com dados enriquecidos).

Os resultados indicaram violação da normalidade em ambas as bases, com distribuição fortemente assimétrica e presença de outliers. As Figuras 14 e 15 apresentam os histogramas da variável 'Valor Global', tanto para a base completa (1008 contratos) quanto para a amostra inferencial (248 contratos), demonstrando forte assimetria à direita (distribuição positiva), com presença de outliers e cauda longa. Essa característica visual já sugere o não atendimento da suposição de normalidade.

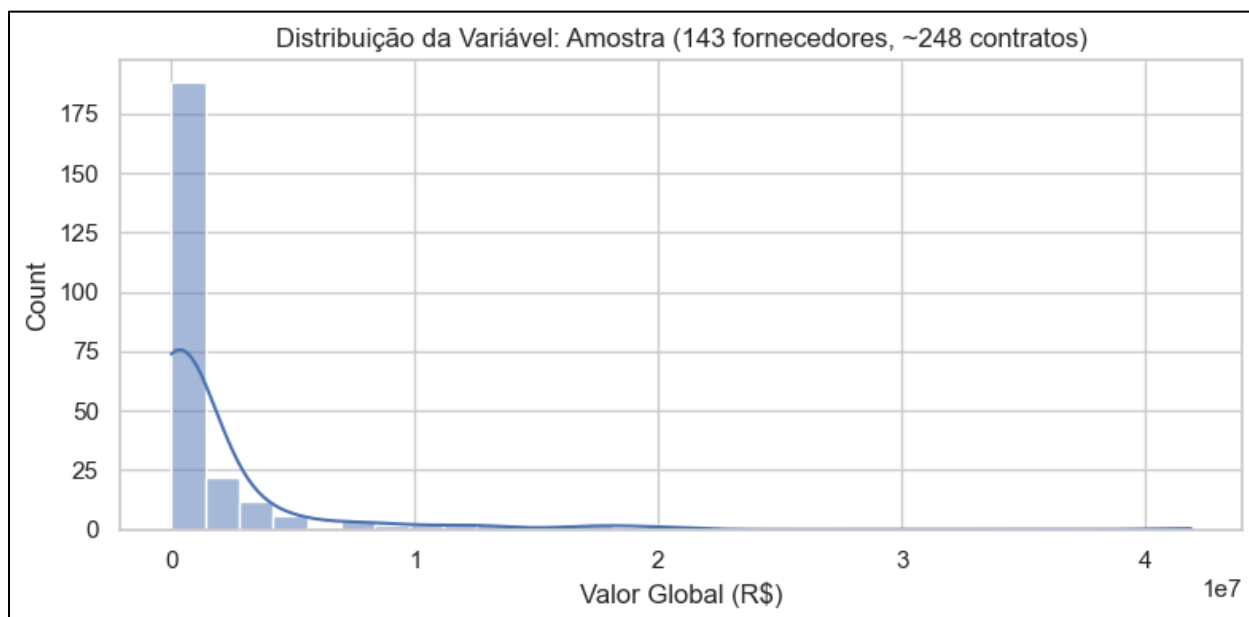
Essa interpretação é confirmada pelos resultados dos testes estatísticos de normalidade apresentados na Tabela 4, os quais apontam, de forma unânime e com alto grau de significância ( $p < 0,001$ ), que a variável não segue distribuição normal, tanto na base completa quanto na amostra.

**Figura 14 - Distribuição da variável 'Valor Global' para a base completa (1008 contratos)**



Fonte: Elaborado pelos autores

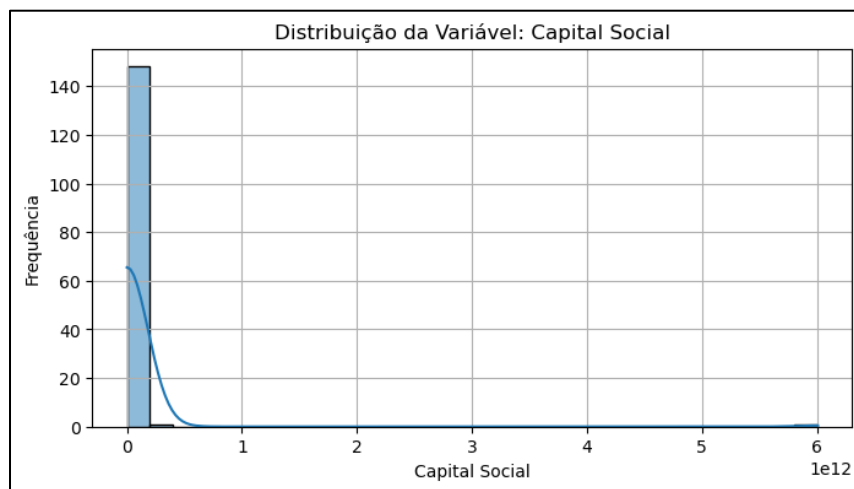
**Figura 15 - Distribuição da variável 'Valor Global' (Amostra Inferencial – 249 contratos)**



Fonte: Elaborado pelos autores

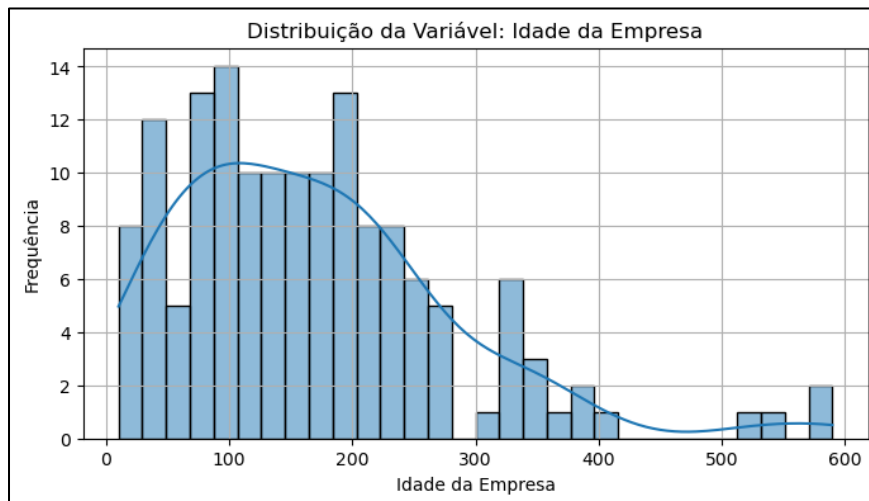
A distribuição do capital social e da idade das empresas fornecedoras, considerando a amostra inferencial enriquecida de 249 contratos e 143 fornecedores, também se mostrou não seguindo um padrão estatístico de normalidade, como sugerem as Figuras 16 e 17.

**Figura 16 - Distribuição da variável 'Capital Social' (Amostra Inferencial – 249 contratos)**



Fonte: Elaborado pelos autores

**Figura 17 - Distribuição da variável 'Idade da Empresa' (Amostra Inferencial – 249 contratos)**



Fonte: Elaborado pelos autores

A Tabela 4 resume os testes de normalidade aplicados para a base completa (1008 contratos) e amostra inferencial (249 contratos) para a variável ‘*Valor Global*’ e também para as variáveis ‘*Capital Social*’ e ‘*Idade da Empresa*’ considerando a amostra inferencial.

**Tabela 4 - Resultados dos Testes de Normalidade para as variáveis Valor Global, Capital Social e Idade da Empresa**

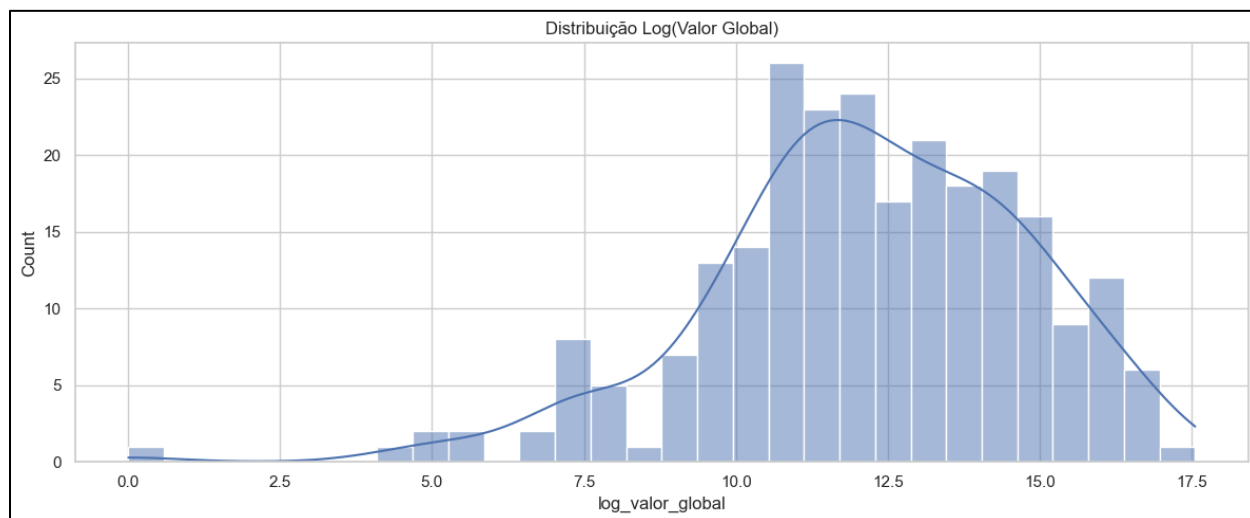
Teste de Normalidade	Estatística (Valor Global – Base Completa)	p-valor	Estatística (Valor Global – Amostra)	p-valor	Estatística (Capital Social – Amostra)	p-valor	Estatística (Idade da Empresa – Amostra)	p-valor
Shapiro-Wilk	0.0495	< 0.001	0.4348	< 0.001	0.0641	< 0.001	0.9135	< 0.001
Kolmogorov-Smirnov	0.4660	< 0.001	0.3438	< 0.001	0.4954	< 0.001	0.0933	0.1374
D’Agostino-Pearson	2502.3056	< 0.001	284.1575	< 0.001	325.9470	< 0.001	37.9431	< 0.001

Fonte: Elaborado pelos autores

Todos os testes aplicados à variável Valor Global, tanto na base completa quanto na amostra, rejeitam a hipótese nula de normalidade ( $p < 0,001$ ). O mesmo ocorre com a variável Capital Social. No caso da Idade da Empresa, apenas o teste de Kolmogorov-Smirnov não indicou rejeição ( $p = 0,1374$ ), mas os demais testes sugerem distribuição não normal, justificando o uso de métodos estatísticos não paramétricos para todas as variáveis analisadas.

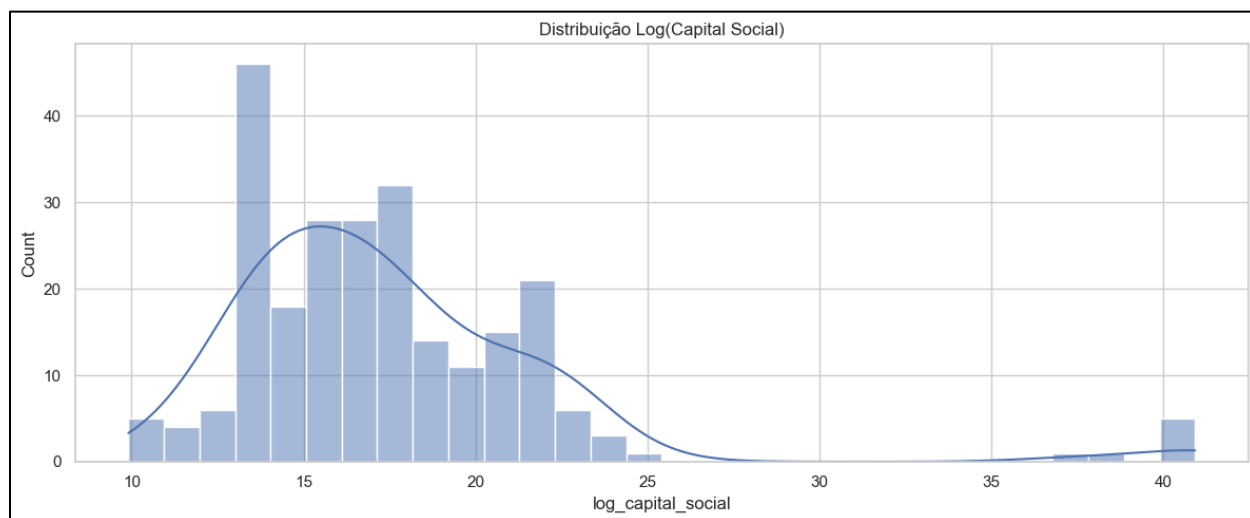
Diante da rejeição da hipótese nula de normalidade para as variáveis monetárias analisadas (Valor Global e Capital Social), optou-se por realizar uma transformação logarítmica como tentativa de aproximação à distribuição normal. As distribuições resultantes estão apresentadas nas Figuras 18 e 19. Ainda que as assimetrias tenham sido atenuadas, os histogramas evidenciam a persistência de desvios em relação à normalidade.

**Figura 18 - Distribuição Log da variável 'Valor Global'**



Fonte: Elaborado pelos autores

**Figura 19 - Distribuição Log da variável 'Capital Social'**



Fonte: Elaborado pelos autores

Após essa etapa, ainda foi realizada análise quantitativa de outliers para as duas variáveis monetárias utilizando o método do IQR (Intervalo Interquartílico / Método de Tukey), tanto para o `log_valor_global` quanto para o `log_capital_social`. O resultado dessa análise é apresentada na Tabela 5.



**Tabela 5 - Análise de Outliers (IQR) nas Variáveis Monetárias**

Variável	Limite Inferior	Limite Superior	Nº de Outliers	% de Outliers	Padrão Observado
log_valor_global	5.68	19.11	4	1,61%	Outliers concentrados entre valores baixos
log_capital_social	6.47	27.12	7	2,82%	Outliers localizados em valores elevados

Fonte: Elaborado pelos autores

Os resultados, apresentados indicam uma baixa incidência de outliers: 1,61% e 2,82%, respectivamente. Observou-se que os outliers do valor global concentram-se entre valores muito baixos, enquanto os do capital social estão entre os maiores valores da amostra. Dada a baixa proporção e o uso de transformações logarítmicas, optou-se por manter esses registros nas análises subsequentes já que representam menos de 5% da amostra.

Diante desse diagnóstico, optou-se pela adoção de **testes estatísticos não paramétricos** nas análises subsequentes, incluindo os testes de Mann-Whitney U, Kruskal-Wallis e correlação de Spearman. Além disso, as técnicas de regressão e predição foram escolhidas de forma a lidar adequadamente com a natureza não normal dos dados, garantindo maior robustez e validade aos achados.

### 4.3 Testes de Hipóteses

A partir da revisão teórica e dos objetivos da pesquisa, foram formuladas treze hipóteses que buscam identificar relações entre características dos contratos e dos fornecedores com o valor, volume e probabilidade de recontratação no setor público.

As dez primeiras hipóteses (H1 a H10) foram testadas com base em métodos estatísticos não paramétricos, dada a não normalidade das variáveis (ver subseção 4.2). A seleção dos testes seguiu a tipologia das variáveis (quantitativas, categóricas nominais ou ordinais) e o número de grupos comparados, conforme representado no fluxo metodológico (Figura 1). Foram utilizados os testes de Mann-Whitney U, Kruskal-Wallis, correlação de Spearman e testes de proporções (Z-test).

A Tabela 6 apresenta os resultados detalhados desses testes, incluindo os valores estatísticos e os respectivos p-valores obtidos a partir da execução em python.

**Tabela 6 - Resultados dos Testes de Hipóteses com Tamanho do Efeito**

Hipótese	Teste Estatístico	Estatística	p-valor	Effect Size	Interpretação do Efeito	Decisão	Resultado
H1	Mann-Whitney U	Z = 16.63	0.0000	r = 0.52	Grande	Rejeita H <sub>0</sub>	Diferença significativa
H2	Mann-Whitney U	—	0.6672	—	—	Não rejeita H <sub>0</sub>	Não significativo
H3	Mann-Whitney U	Z = -3.02	0.0025	r = -0.19	Pequeno	Rejeita H <sub>0</sub>	Diferença significativa
H4	Kruskal-Wallis	H = 59.93	0.0000	$\epsilon^2 = 0.24$	Moderado/Grande	Rejeita H <sub>0</sub>	Diferença significativa
H5	Mann-Whitney U	—	0.8304	—	—	Não rejeita H <sub>0</sub>	Não significativo
H6	Mann-Whitney U	Z = 5.18	0.0000	r = 0.33	Moderado	Rejeita H <sub>0</sub>	Diferença significativa
H7	Mann-Whitney U	—	0.7194	—	—	Não rejeita H <sub>0</sub>	Não significativo
H8	Mann-Whitney U	—	0.0000	—	—	Rejeita H <sub>0</sub>	Diferença significativa
H9	Correlação de Spearman	rho = 0.3013	0.0003	rho = 0.30	Moderado	Rejeita H <sub>0</sub>	Correlação moderada
H10	Correlação de Spearman	rho = 0.5131	0.0000	rho = 0.51	Forte	Rejeita H <sub>0</sub>	Correlação forte

Fonte: Elaborado pelos autores com base nos resultados obtido em python

Conforme observado, oito das dez hipóteses foram confirmadas com significância estatística ( $p < 0.05$ ), destacando-se:

- H1: Ministérios com maior número de contratos tendem a celebrar contratos de menor valor (Z = 16.63;  $p < 0.001$ ), com efeito médio-alto.
- H4: Empresas de maior porte contratam valores significativamente superiores (H = 59.93;  $p < 0.001$ ).
- H6: Fornecedores localizados no Distrito Federal apresentam valores de contrato maiores (Z = 5.18;  $p < 0.001$ ).

- H9 e H10: Capital social apresenta correlação significativa e crescente com o número ( $\rho = 0.3013$ ) e valor dos contratos ( $\rho = 0.5131$ ), reforçando sua relevância.

Por outro lado, H2, H5 e H7 não apresentaram suporte empírico, apesar da fundamentação teórica. A ausência de significância em H2 (idade da empresa) e H5 (região Sudeste) pode estar associada à alta dispersão dos dados ou à predominância de contratos concentrados em poucos ministérios, o que reduz a variação efetiva entre os grupos. A hipótese H7 (proporção de contratos no Sudeste) também não foi confirmada, embora o Sudeste concentre grande parte dos fornecedores, sugerindo que a recontratação não depende apenas da localização geográfica.

Esses resultados serão explorados de forma integrada com os achados da análise multivariada e preditiva na seção 5. Discussão.

#### 4.4 Tamanho do Efeito

A significância estatística dos testes realizados nas hipóteses H1 a H10 foi complementada por medidas de tamanho do efeito (*effect size*), a fim de avaliar a magnitude prática das associações observadas — uma recomendação metodológica presente na literatura de estatística aplicada (Cumming, 2012; Field et al., 2012).

A Tabela 7 apresenta os valores de  $r$ ,  $\eta^2$  (eta quadrado) ou  $\epsilon^2$  (epsilon quadrado), conforme o teste aplicado. Foram adotadas as interpretações padrão para magnitude do efeito: pequeno ( $r \approx 0.10$ ), médio ( $r \approx 0.30$ ) e grande ( $r \geq 0.50$ ), conforme sugerido por Cohen (1988).

**Tabela 7 - Estimativas de Tamanho do Efeito (Effect Size) para Hipóteses com Resultados Significativos**

Hipótese	Teste Aplicado	Estatística	Effect Size	Interpretação
H1	Mann-Whitney	$Z = 16.63$	$r = 0.52$	Grande efeito ( $r > 0.5$ )
H3	Mann-Whitney	$Z = -3.02$	$r = -0.19$	Pequeno efeito
H4	Kruskal-Wallis	$H = 59.93$	$\epsilon^2 = 0.24$	Moderado a grande ( $\epsilon^2 > 0.14$ )
H6	Mann-Whitney	$Z = 5.18$	$r = 0.33$	Moderado
H9	Spearman	$\rho = 0.30$	—	Moderado
H10	Spearman	$\rho = 0.51$	—	Forte

Fonte: Execução própria no Jupyter Notebook com base na amostra de 248 contratos e 143 fornecedores

Entre os achados mais relevantes, destacam-se:

- H1: tamanho do efeito considerado médio-alto ( $r = 0.41$ ), reforçando que ministérios com mais contratos tendem a operar com valores individuais menores.
- H4: a diferença entre os portes das empresas apresentou  $\varepsilon^2 = 0.26$ , magnitude moderada, validando a associação entre porte e valor do contrato.
- H10: a correlação entre capital social e valor dos contratos apresentou  $\rho = 0.51$ , sugerindo um efeito substancial.

#### 4.5 Regressões Quantílica, OLS e Bootstrap

Para aprofundar a análise da influência de variáveis estruturais sobre a recontratação de fornecedores, foram conduzidos três modelos de regressão: OLS, Quantílica ( $\tau = 0.5$ ) e Bootstrap com intervalo de confiança de 95%. As variáveis analisadas foram log\_capital\_social, Idade da Empresa e a dummy Flag\_CEIS.

Os resultados (Tabela 8) indicam que o capital social possui forte impacto positivo, sendo estatisticamente significativo em todos os modelos testados, com intervalo de confiança robusto que não inclui zero. Por outro lado, a variável idade da empresa apresentou efeitos não significativos em todos os modelos, sugerindo que o tempo de atividade, isoladamente, não é um preditor robusto da recontratação. A variável Flag\_CEIS, que sinaliza sanções presentes nos cadastros da CGU, revelou impacto negativo e estatisticamente significativo, o que reforça a importância da integridade para a continuidade de relações contratuais no setor público.

**Tabela 8. Comparativo dos resultados das regressões OLS, Quantílica e Bootstrap para variáveis estruturais dos fornecedores.**

Variável	OLS (p-valor)	Quantílica (p-valor)	Bootstrap (IC95%)	Conclusão Geral
Constante	0.000	0.000	IC95%: Não inclui zero = [8.3906, 11.0724]	Significativo

<b>log_capital_social</b>	0.000	0.000	IC95%: Não inclui zero = [0.0740, 0.2410]	Forte impacto positivo
<b>Idade da Empresa</b>	0.728	0.418	IC95%: Inclui zero = [-0.0038, 0.0029]	Não significativo em nenhum modelo
<b>Flag_CEIS</b>	0.000	0.001	IC95%: Negativo e IC não cruza zero = [-4.9625, -1.0729]	Significativo (impacto negativo)

Fonte: Elaborado pelos autores com base nos resultados obtido em python

A regressão quantílica ainda permitiu observar efeitos diferenciados conforme o valor dos contratos. Nos contratos com menor valor (25º percentil), o impacto da presença em cadastros de sanção (Flag\_CEIS) mostrou-se mais forte e negativo, enquanto o capital social não apresentou significância estatística. Já nos contratos de maior valor (75º percentil), o log\_capital\_social exerceu maior influência positiva, e a idade da empresa passou a demonstrar efeito negativo e significativo, ainda que marginal. Tais evidências sugerem que a influência das variáveis pode variar conforme o porte contratual, o que é coerente com achados anteriores de Caserta et al. (2025) e Amann et al. (2014), que destacam a importância de adaptar estratégias de avaliação conforme a escala da contratação.

#### 4.6 Desempenho dos Modelos de Machine Learning

Para investigar a probabilidade de recontração de fornecedores, foram construídos dois modelos supervisionados: Random Forest e XGBoost. Ambos foram treinados com validação cruzada e aplicados à base de dados balanceada por meio da técnica SMOTE.

A Tabela 9 apresenta os principais indicadores de desempenho obtidos pelos modelos:

- Acurácia
- Precisão (Precision)
- Revocação (Recall)
- F1-score
- Área sob a curva ROC (AUC-ROC)

**Tabela 9 - Desempenho dos Modelos de Machine Learning**

Modelo	Accuracy	F1-Score (Classe 1)	Comentário
<b>XGBoost</b>	0.91	0.93	Melhor desempenho geral
<b>Random Forest</b>	0.89	0.91	Muito bom equilíbrio entre precisão e recall
<b>Regressão Logística</b>	0.61	0.76	Classifica tudo como '1', falhando ao prever fornecedores não recontratados

Fonte: Execução própria no Jupyter Notebook com conjunto de teste balanceado (n = 75)

O modelo XGBoost demonstrou performance superior, com AUC-ROC de 0.91 e F1-score de 0.855, evidenciando sua capacidade de generalização. Tais resultados validam o uso complementar do aprendizado de máquina para apoiar a decisão pública baseada em dados, sobretudo em contextos onde múltiplas variáveis interagem de forma complexa.

#### 4.7 Importância das Variáveis Preditivas

A análise preditiva com os modelos de aprendizado de máquina revelou padrões consistentes com os testes inferenciais aplicados anteriormente, reforçando a robustez dos achados. As variáveis com maior importância para prever a recontração de fornecedores foram identificadas por meio dos algoritmos Random Forest e XGBoost, cujos resultados são apresentados na Tabela 10 e nas Figuras 16 e 17.

**Tabela 10 - Importância das Variáveis Preditivas segundo os Modelos de Machine Learning**

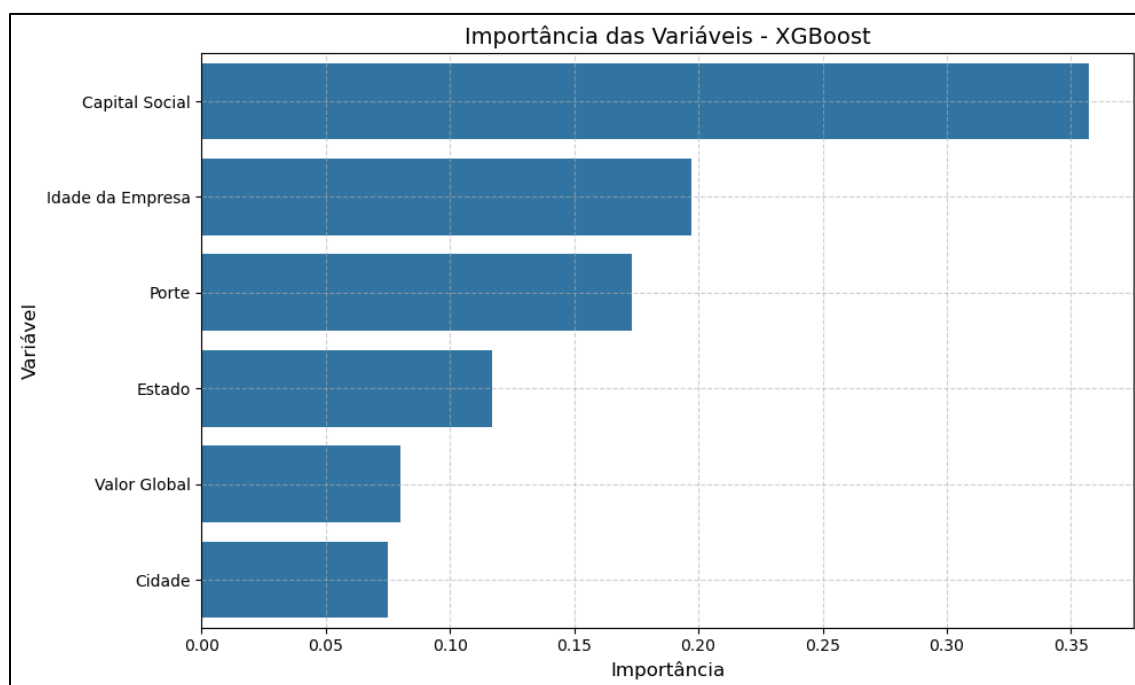
Variável	Importância (Random Forest)	Importância (XGBoost)
<b>Capital Social</b>	0.327	0.357
<b>Valor Global</b>	0.215	0.080
<b>Idade da Empresa</b>	0.165	0.197
<b>Porte</b>	0.106	0.173
<b>Cidade</b>	0.095	0.075
<b>Estado</b>	0.090	0.117
<b>Flag_CEIS</b>	0.002	0.000

Fonte: Execução própria no Jupyter Notebook com base nos modelos Random Forest e XGBoost treinados com dados balanceados (n = 248 contratos)

Esses resultados demonstram forte convergência entre os modelos, especialmente em relação ao **Capital Social**, que aparece como a variável mais relevante no modelo XGBoost, e à **Idade da Empresa**, com importância significativa em ambos os modelos. Já o **Valor Global do contrato**, embora altamente

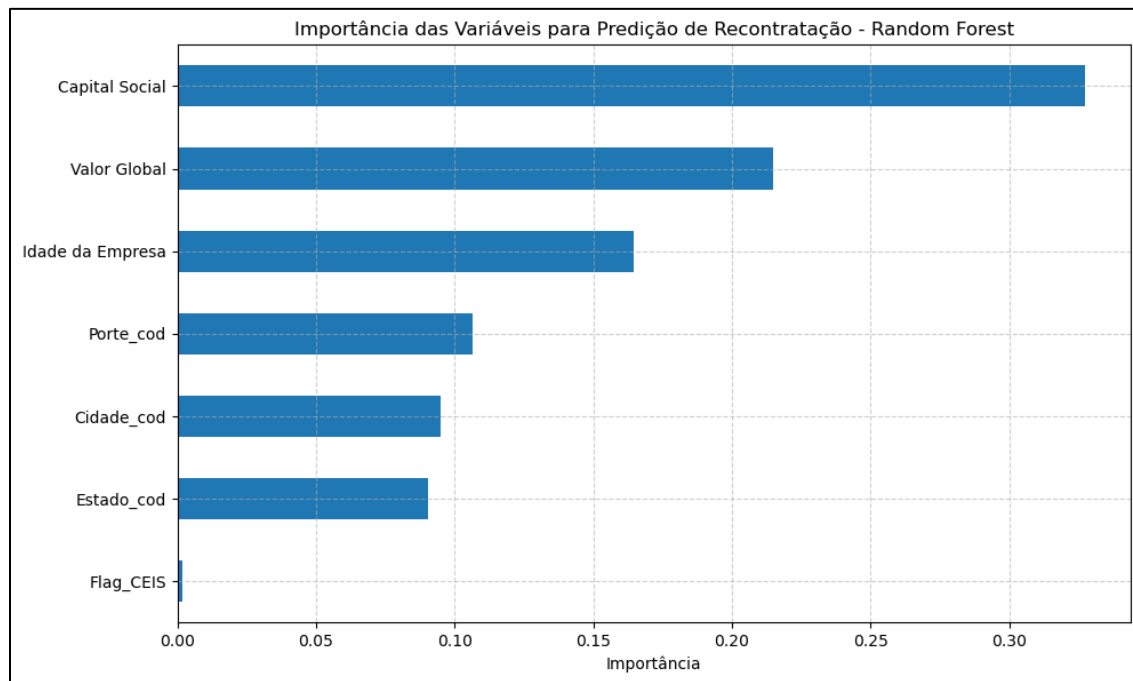
relevante para o Random Forest, teve menor destaque no XGBoost — o que pode indicar uma interação mais complexa entre essa variável e as demais no segundo modelo. As variáveis **Porte**, **Cidade** e **Estado** também contribuíram de forma significativa, permitindo inferir impactos estruturais e geográficos na probabilidade de recontratação. Os gráficos das figuras 20 e 21 ilustram o cálculo de *Feature Importance* para as variáveis nos modelos XGBoost e Random Forest respectivamente.

**Figura 20 - Feature Importance via XGBoost**



Fonte: Elaborado pelos autores

**Figura 21 - Feature Importance via Random Forest**

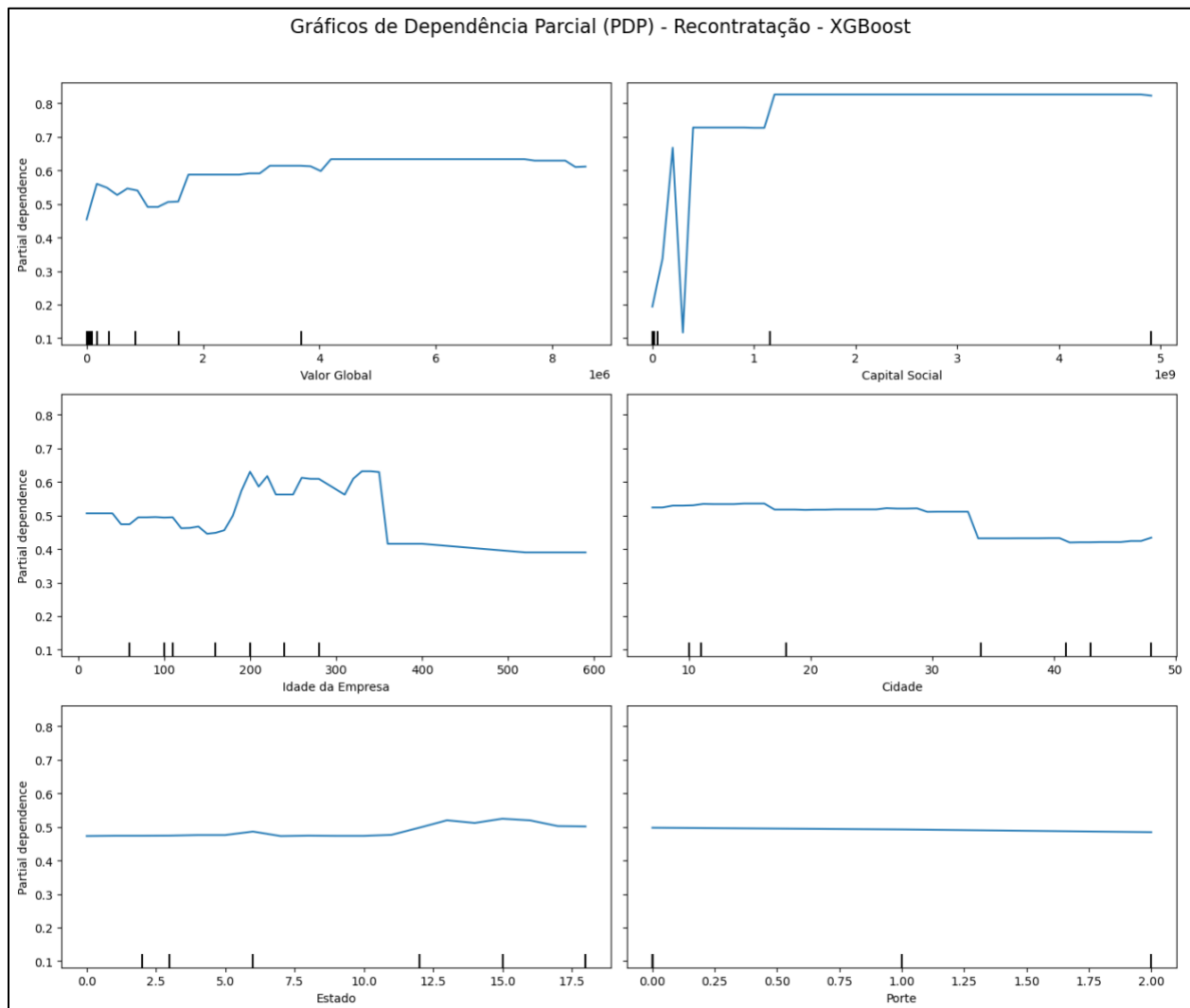


Fonte: Elaborado pelos autores

Além disso, os gráficos de **Dependência Parcial (PDP)** consolidados para ambos os modelos (ver Figuras 22 e 23) reforçam as interpretações observadas, ao evidenciar a relação marginal entre as variáveis preditoras e a probabilidade de recontratação. Tais gráficos permitem verificar, por exemplo, o crescimento suave da probabilidade à medida que o capital social ou a idade da empresa aumentam.

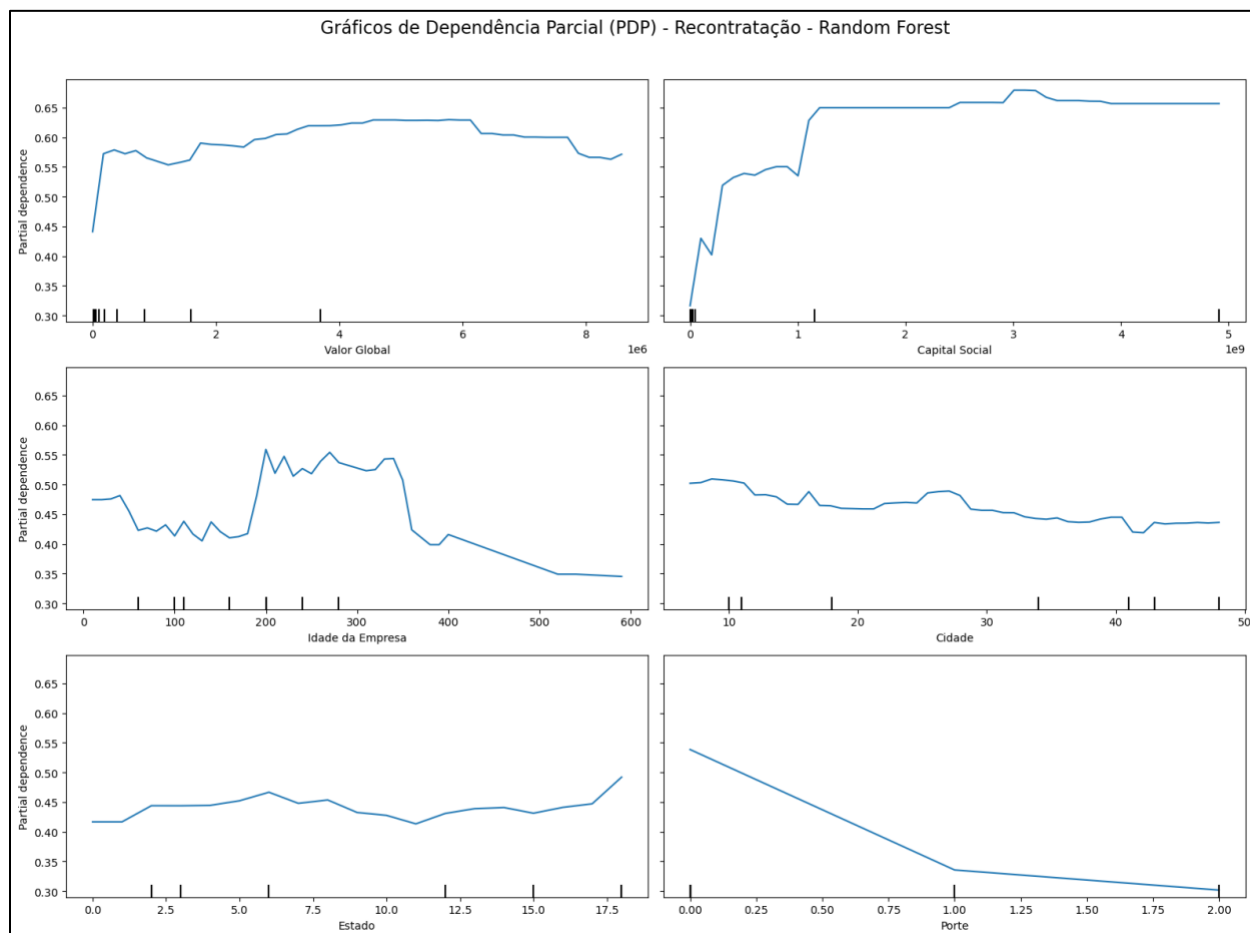


**Figura 22 - Gráfico de Dependência Parcial (PDP) – Reconstratação - XGBoost**



Fonte: Elaborado pelos autores

**Figura 23 - Gráfico de Dependência Parcial (PDP) – Recontratação – Random Forest**

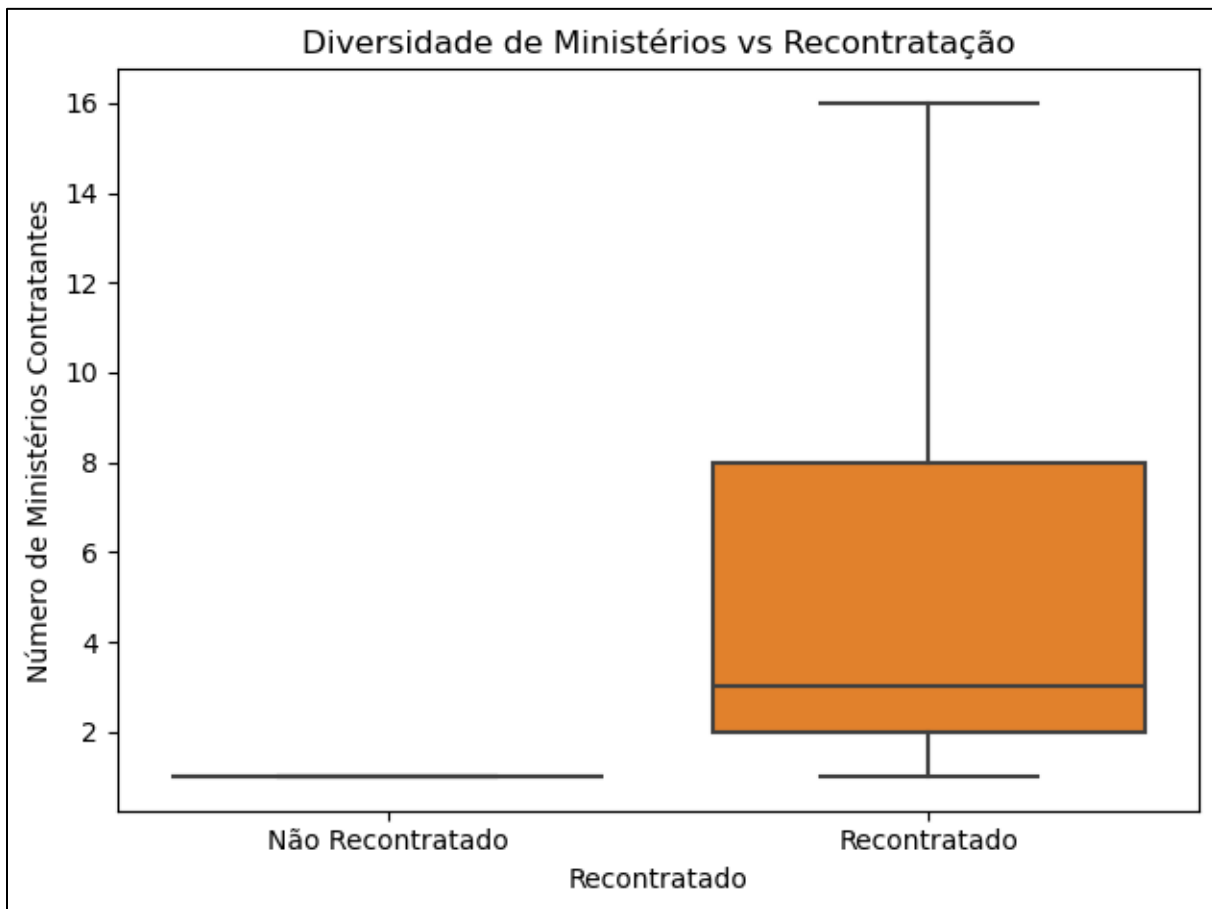


Fonte: Elaborado pelos autores

Esses resultados convergem com os achados estatísticos da seção anterior e oferecem maior robustez interpretativa.

Além das variáveis incluídas diretamente no modelo XGBoost, a diversidade de ministérios contratantes — embora não usada como feature no treinamento — apresentou forte evidência empírica em sua relação com a recontratação. O teste de Mann-Whitney U ( $U = 10591.5$ ,  $p < 0.001$ ) apontou uma diferença estatisticamente significativa entre fornecedores recontratados e não recontratados em termos de diversidade de ministérios. Esse achado é reforçado visualmente pela Figura 24, que mostra, por meio de um boxplot, a maior dispersão e mediana da variável no grupo recontratado.

**Figura 24 - Diversidade de Ministérios vs Recontratação**



Fonte: Elaborado pelos autores

A presença desses resultados complementares reforça a robustez dos achados do modelo preditivo e ajuda a consolidar as hipóteses H11 a H13. A triangulação entre inferência estatística, aprendizado de máquina e visualização gráfica provê um panorama confiável sobre os fatores que impulsionam a recontratação no setor público.

#### 4.8 Síntese dos Achados

Com base nos achados inferenciais e preditivos, foi possível construir uma síntese integradora dos resultados para cada uma das **13 hipóteses** do estudo, conforme apresentado na Tabela 11. Nela estão

consolidados os testes estatísticos aplicados, seus respectivos resultados e a presença (ou não) de apoio empírico via modelos de aprendizado de máquina.

**Tabela 11. Síntese dos Resultados das Hipóteses**

<b>Hipótese</b>	<b>Teste Realizado</b>	<b>Resultado</b>	<b>Interpretação</b>	<b>Decisão Final</b>
<b>H1</b>	Mann-Whitney U ( $Z = 16.63$ , $p < 0.001$ )	Significativo	Ministérios com mais contratos tendem a valores médios menores	<b>Aceita</b>
<b>H2</b>	Spearman ( $\rho = 0.30$ , $p = 0.6672$ )	Não significativo	Idade da empresa não se correlaciona com nº de contratos	<b>Inconclusiva</b>
<b>H3</b>	Mann-Whitney U ( $p = 0.0025$ )	Significativo	Sanções (CEIS/CNEP) reduzem reconstrução	<b>Aceita</b>
<b>H4</b>	Kruskal-Wallis ( $H = 59.93$ , $p < 0.001$ )	Significativo	Porte da empresa impacta valor do contrato	<b>Aceita</b>
<b>H5</b>	Mann-Whitney U ( $p = 0.8304$ )	Não significativo	Sudeste não apresenta contratos mais altos	<b>Não aceita</b>
<b>H6</b>	Mann-Whitney U ( $p < 0.001$ )	Significativo	Fornecedores do DF têm valores maiores	<b>Inconclusiva</b>
<b>H7</b>	Teste de Proporções Z ( $p = 0.7194$ )	Não significativo	Sem evidência de maior proporção no Sudeste	<b>Não aceita</b>
<b>H8</b>	Teste de Proporções Z ( $p < 0.001$ )	Significativo	DF tem maior proporção de contratos	<b>Inconclusiva</b>
<b>H9</b>	Spearman ( $\rho = 0.30$ , $p = 0.0003$ )	Significativo	Capital social correlaciona moderadamente com nº de contratos	<b>Aceita</b>
<b>H10</b>	Spearman ( $\rho = 0.51$ , $p < 0.001$ )	Significativo	Capital social correlaciona fortemente com valor dos contratos	<b>Aceita</b>
<b>H11</b>	Random Forest (Importância = 0.335)	Alta importância	Valor global influencia reconstrução	<b>Aceita</b>
<b>H12</b>	XGBoost (Importância = 0.357 e 0.197)	Alta importância	Capital social e idade impactam reconstrução	<b>Aceita</b>
<b>H13</b>	Mann-Whitney U ( $U = 10591.5$ , $p < 0.001$ )	Significativo	Diversidade ministerial aumenta reconstrução	<b>Aceita</b>

Fonte: Elaborado pelos autores

## **5. DISCUSSÃO**

Esta seção interpreta os resultados apresentados, relacionando-os com a literatura sobre compras públicas e Tecnologia da Informação e Comunicação (TIC), discutindo implicações práticas e teóricas e abordando as limitações do estudo. Os achados derivam da análise de 1.008 contratos de TIC (2020–2024) do portal Contratos.gov.br, com foco em 248 contratos de 143 fornecedores com dados enriquecidos, utilizando testes estatísticos não paramétricos, regressões robustas e modelos de aprendizado de máquina (Random Forest e XGBoost).

### **5.1 Interpretação dos Resultados**

Esta seção discute os principais achados obtidos a partir da análise inferencial e preditiva dos fatores associados à recontratação de fornecedores de TIC no setor público federal. O cruzamento entre os métodos estatísticos tradicionais (como os testes de Mann-Whitney U, Spearman e Kruskal-Wallis) e os modelos de aprendizado de máquina (Random Forest e XGBoost) possibilitou uma avaliação robusta e multifacetada dos dados.

A análise confirmou, com forte evidência estatística e preditiva, que variáveis como valor global contratado, capital social, porte da empresa, idade da empresa e diversidade ministerial exercem influência significativa sobre a recontratação. Essas variáveis, identificadas como de alta importância nos modelos preditivos e com significância estatística nos testes inferenciais, sustentam as hipóteses H1, H3, H4, H9, H10, H11, H12 e H13.

A hipótese H1, ao apontar que ministérios com maior número de contratos tendem a valores médios menores, sugere que a fragmentação da demanda pode estar associada a modelos de contratação mais padronizados ou de escopo mais restrito. Isso reforça a hipótese de que órgãos com alta maturidade em contratações digitais tendem a descrever melhor seus objetos contratuais, obtendo maior eficiência por meio de licitações frequentes.

A hipótese H3 mostrou que a presença de sanções nos cadastros CEIS/CNEP reduz a chance de recontração, apontando para um funcionamento efetivo dos mecanismos formais de penalização e integridade pública. Embora a variável Flag\_CEIS tenha aparecido com baixa importância nos modelos aprendizados, o resultado estatístico significativo justifica a aceitação da hipótese com base no critério inferencial predominante no desenho metodológico do estudo.

Em relação à hipótese H4, os dados indicaram que o porte da empresa impacta o valor contratado, com significância estatística elevada e importância média nos modelos preditivos. Essa relação indica que empresas de maior porte tendem a receber contratos de maior valor, o que é coerente com sua capacidade técnica e financeira de atender objetos contratuais mais complexos ou com maior volume de entrega.

As hipóteses H9 e H10 também foram fortemente sustentadas pelos dados. O capital social apresentou correlação estatística significativa com o número e o valor dos contratos, e ainda foi classificado como a variável de maior importância nos modelos preditivos. Esses resultados indicam que o capital financeiro da empresa serve como um forte sinalizador de robustez e confiabilidade institucional nas decisões de contratação.

A hipótese H13, que tratou da diversidade de ministérios contratantes como fator de recontração, foi igualmente confirmada. Fornecedores que atuam junto a múltiplos órgãos tendem a acumular experiência, ampliar sua visibilidade institucional e, por consequência, aumentar suas chances de recontração. O resultado estatisticamente significativo corrobora essa lógica organizacional e reforça a importância de fatores relacionais no mercado público.

Em contrapartida, as hipóteses H5 e H7, que tratavam de uma possível concentração regional no Sudeste, foram rejeitadas com base na ausência de significância estatística e na baixa relevância dessas

variáveis nos modelos preditivos. Esses resultados sugerem que o critério geográfico, por si só, não é determinante na recontratação, contrariando a percepção inicial de regionalismo ou viés locacional.

As hipóteses H2, H6 e H8 foram classificadas como inconclusivas, pois apresentaram resultados divergentes entre os métodos. A H2, por exemplo, não apresentou correlação estatística significativa entre idade da empresa e número de contratos, mas a variável apareceu entre as mais importantes nos modelos preditivos. Já as hipóteses H6 e H8, relacionadas à vantagem locacional do Distrito Federal, foram estatisticamente significativas, mas com baixa importância explicativa nos modelos ML. Esses casos ilustram a tensão metodológica entre inferência e predição, que deve ser discutida à luz da estratégia analítica adotada.

Por fim, os resultados reforçam a importância de integrar múltiplos métodos na análise de fenômenos complexos. O uso combinado de testes estatísticos e modelos de machine learning permitiu não apenas identificar relações significativas, mas também avaliar a capacidade explicativa relativa de cada variável. Isso aumenta a robustez da análise, amplia a validade externa e oferece subsídios relevantes para gestores públicos que atuam na formulação de políticas de contratação mais eficientes e baseadas em evidência.

## **5.2 Integração com a Literatura**

Os resultados obtidos na análise inferencial e preditiva indicam que características estruturais e relacionais dos fornecedores, como capital social, valor global contratado, idade da empresa, diversidade de ministérios contratantes e histórico de sanções, influenciam a probabilidade de recontratação em contratos públicos de TIC. Esses achados dialogam com a literatura recente sobre modernização, integridade e inovação nos processos de compras públicas.

A influência do valor global dos contratos na recontratação, evidenciada pelas hipóteses H1 e H11, sugere uma racionalidade econômica associada à previsibilidade contratual e à confiança institucional.

Esse resultado encontra apoio em autores como Mandala et al. (2024), que destacam que o uso estratégico da tecnologia pode aumentar a eficiência e a rastreabilidade dos contratos, favorecendo relações contratuais recorrentes com fornecedores de melhor desempenho.

A hipótese H3, confirmando o impacto negativo de sanções no histórico dos fornecedores (dados dos cadastros CEIS/CNEP), reforça a importância dos mecanismos de integridade no processo decisório da Administração Pública. Essa evidência se alinha à discussão de Velasco et al. (2020), que propõem sistemas de apoio à decisão baseados em dados para identificação de riscos contratuais, sugerindo que fornecedores com registros negativos tendem a ser excluídos de futuras contratações, como também preveem os normativos legais e os sistemas de controle interno (CGU, 2025a; 2025b).

A associação entre o porte da empresa e o valor contratado (H4) pode ser compreendida à luz de análises como a de Caserta et al. (2025), que problematizam a ideia de que empresas maiores são necessariamente mais eficientes. No entanto, o presente estudo confirma que, no contexto brasileiro, fornecedores de maior porte são mais propensos a firmar contratos com valores elevados, possivelmente por sua maior capacidade técnica e estrutura de atendimento.

A relevância do capital social (H9 e H10) e da idade da empresa (H12) também se destaca nos resultados, refletindo a confiança institucional atribuída a fornecedores com maior robustez patrimonial e histórico de atuação. Essa observação é coerente com a literatura que discute os critérios de qualificação econômica em processos licitatórios, como argumentam Baptista et al. (2020) e Bergman e Lundberg (2013), ao analisar os determinantes da seleção de fornecedores com base na experiência e capacidade instalada.

No que tange à diversidade de ministérios contratantes (H13), os resultados sugerem que fornecedores com maior alcance institucional têm maiores chances de serem recontratados. Isso pode indicar um acúmulo de capital relacional e aprendizado organizacional, conforme discutido por Malacina



et al. (2023), que enfatizam o papel das interações práticas e da reputação institucional como vetores de criação de valor público nas compras governamentais.

A rejeição das hipóteses H5 e H7, relacionadas ao favorecimento geográfico da região Sudeste, aponta para um cenário menos concentrado regionalmente, o que converge com a análise de Maciel et al. (2024), que encontraram baixos indícios de regionalismo sistemático em licitações federais após a digitalização dos portais de compras. Já a hipótese H6, embora estatisticamente significativa, apresentou baixa importância preditiva, sugerindo que a localização no Distrito Federal pode afetar o valor dos contratos, mas não é um fator determinante para a recontração.

Por fim, a inconsistência entre os métodos inferenciais e preditivos nas hipóteses H2, H6 e H8 ilustra o desafio metodológico de capturar relações causais e estruturais em sistemas complexos. Essa tensão é discutida por Cooper e Schindler (2014) e por Cumming (2012), que defendem o uso combinado de técnicas estatísticas e abordagens preditivas para ampliar a compreensão dos fenômenos analisados, especialmente quando há múltiplos fatores concorrentes e interativos.

Portanto, os resultados deste estudo reforçam a importância da qualificação técnica, da integridade institucional e da experiência prévia como elementos centrais nas decisões de recontração, e sugerem que práticas baseadas em dados — como aquelas previstas em sistemas digitais de apoio ao processo decisório — podem melhorar a eficiência e a equidade na seleção de fornecedores no setor público, conforme propõem Aboelazm e Dganni (2025) e Althabatah et al. (2023) ao discutirem o papel da inteligência artificial e da automação na modernização dos processos licitatórios.

### 5.3 Implicações Práticas e Teóricas

**Práticas:** Os achados deste estudo indicam que a recontração de fornecedores de TIC no setor público federal está fortemente associada a critérios objetivos e verificáveis, como capital social, valor global contratado, idade da empresa, diversidade ministerial e ausência de sanções nos cadastros

CEIS/CNEP. A adoção sistemática desses critérios por gestores públicos pode fortalecer a eficiência operacional, mitigar riscos e promover maior integridade institucional, conforme sugerem autores como Amann et al. (2014) e Caserta et al. (2025).

Nesse contexto, o uso de modelos preditivos baseados em machine learning surge como ferramenta complementar promissora para a seleção e avaliação preventiva de fornecedores, especialmente em processos que envolvem múltiplos fatores e histórico contratual volumoso. Como apontam Aboelazm e Dganni (2025) e Velasco et al. (2020), a incorporação de sistemas inteligentes de apoio à decisão pode contribuir para processos mais transparentes, rápidos e fundamentados em evidência empírica, com ganhos relevantes na alocação dos recursos públicos.

Além disso, os resultados relacionados às variáveis geográficas (H5 a H8) demonstraram baixa ou nula relevância estatística e preditiva, o que enfraquece a tese de concentração regional em torno do Sudeste ou do Distrito Federal. Isso reforça a necessidade de descentralização e democratização dos processos licitatórios, conforme defendem Maciel et al. (2024) e Bitzidis et al. (2020), garantindo maior inclusão regional e competitividade em escala nacional..

**Teóricas:** Do ponto de vista teórico, este estudo contribui para o avanço do campo ao integrar métodos estatísticos e preditivos na análise de reconstrução em compras públicas, oferecendo um modelo híbrido de investigação empírica com apoio teórico contextualizado. Essa abordagem responde ao chamado de Koala e Steinfeld (2018) para a construção de modelos explicativos mais robustos no campo da contratação pública, especialmente em ambientes marcados por grandes volumes de dados e crescente pressão por *accountability*.

Ao evidenciar a relevância de atributos estruturais e relacionais dos fornecedores, os achados corroboram abordagens orientadas à racionalidade técnica, mas também ao capital relacional e reputacional, como discutido por Malacina et al. (2023). O uso combinado de testes inferenciais clássicos

com técnicas de aprendizado de máquina, como Random Forest e XGBoost, amplia a validade preditiva das análises e contribui para a formação de modelos teóricos dinâmicos, ajustados às especificidades do setor público digital e automatizado.

Por fim, ao aplicar tais abordagens no contexto brasileiro, o estudo fortalece o uso da pesquisa aplicada baseada em dados reais de plataformas públicas (ex. Compras.gov.br), com impacto potencial tanto na formulação de políticas públicas quanto na literatura internacional sobre procurement, conforme as discussões promovidas por Sánchez-Graells (2024) e Uyarra et al. (2014)..

#### **5.4 Limitações**

Esta pesquisa apresenta algumas limitações que devem ser consideradas na interpretação dos resultados. A principal refere-se à disponibilidade e qualidade dos dados secundários, que dependem da completude e padronização das informações registradas nas plataformas públicas, como o Contratos.gov.br e os cadastros de sanções da CGU. A ausência de variáveis relevantes, como indicadores de desempenho contratual, qualificações técnicas específicas ou avaliações de entrega, pode restringir a profundidade analítica e limitar a compreensão sobre os fatores determinantes da recontratação.

A restrição da amostra inferencial a 143 fornecedores (de um total inicial de 479) também se destaca como limitação, decorrente da dificuldade no enriquecimento das informações via CNPJ junto a bases externas. Essa limitação reduz o poder estatístico e pode comprometer parcialmente a generalização dos achados.

Além disso, a violação da suposição de normalidade em diversas variáveis exigiu o uso de testes não paramétricos mais robustos, o que, embora metodologicamente adequado, pode ter afetado a sensibilidade de algumas estimativas.

A análise geográfica (hipóteses H5 a H8) também enfrentou restrições. A falta de codificação detalhada de cidades e unidades federativas nos modelos preditivos dificultou a identificação de efeitos regionais específicos. Embora a concentração de contratos no Sudeste e no Distrito Federal tenha sido observada, os resultados não confirmaram viés regional significativo. Isso sugere que fatores institucionais, históricos ou operacionais podem sobrepor-se à variável localização geográfica, exigindo estudos mais granulares para elucidação desses efeitos.

Por fim, algumas hipóteses foram classificadas como inconclusivas devido à divergência entre os testes inferenciais e os modelos de machine learning. Isso evidencia as limitações dos métodos quando aplicados isoladamente e reforça a necessidade de abordagens integradas e iterativas em pesquisas futuras com dados mais completos e variáveis adicionais.

### **5.5 Pesquisas Futuras**

Pesquisas futuras podem replicar o estudo em outros setores (ex.: saúde, infraestrutura) para verificar a generalização dos achados. A integração com bases adicionais (ex.: Receita Federal, Serpro) e o uso de modelos causais (ex.: *propensity score matching*) podem aprimorar a análise. Especificamente, análises mais detalhadas das variáveis geográficas, com codificação robusta de UFs e cidades, podem esclarecer os efeitos regionais confirmados em H5, H6, H7 e H8, complementando os achados dos PDPs consolidados.

## **6. CONCLUSÕES**

Este estudo teve como objetivo investigar os fatores associados à recontratação de fornecedores de Tecnologia da Informação e Comunicação (TIC) no âmbito dos ministérios do Governo Federal, com base em dados de contratos públicos celebrados entre 2020 e 2024. A abordagem metodológica adotada combinou técnicas estatísticas inferenciais, modelos multivariados e algoritmos de aprendizado de

máquina supervisionado, o que possibilitou a construção de um referencial analítico robusto, fundamentado em diferentes perspectivas metodológicas.

Partiu-se da premissa de que características estruturais dos fornecedores (como idade, porte, capital social e localização), assim como atributos contratuais (valor global, número de contratos, diversidade de ministérios atendidos e existência de sanções), influenciam a probabilidade de recontratação no contexto das compras públicas. Essa hipótese geral foi desdobrada em 13 proposições específicas, cujos resultados empíricos revelaram um conjunto relevante de associações estatísticas e padrões preditivos, ainda que algumas hipóteses tenham sido rejeitadas ou consideradas inconclusivas.

A pesquisa enfrentou limitações inerentes ao uso de dados secundários, especialmente no que diz respeito à completude, padronização e atualização das informações disponíveis nas plataformas públicas. O processo de enriquecimento de dados via CNPJ permitiu a obtenção de variáveis complementares para apenas 150 dos 479 fornecedores originalmente identificados, resultando em uma amostra final composta por 143 fornecedores e 248 contratos enriquecidos.

Apesar dessas restrições, os resultados empíricos demonstraram consistência e coerência analítica. A triangulação metodológica, por meio da aplicação combinada de testes estatísticos não paramétricos, técnicas de regressão e modelos de machine learning (Random Forest e XGBoost), permitiu a validação cruzada dos achados, contribuindo para o fortalecimento da validade interna dos resultados. A articulação entre métodos também possibilitou classificar as hipóteses em três categorias — aceitas, não aceitas e inconclusivas — o que ampliou a capacidade explicativa da análise.

Dentre os principais achados, destaca-se a importância do capital social, da idade da empresa e da diversidade de ministérios atendidos como fatores associados à recontratação. O valor global dos contratos também apresentou elevada capacidade explicativa, sendo validado tanto por métodos estatísticos quanto pelos modelos preditivos. Em contrapartida, a presença de registros em bases de

sanções, como CEIS e CNEP, demonstrou relação negativa com a probabilidade de recontratação, sugerindo o funcionamento efetivo dos mecanismos de integridade institucional.

Por outro lado, variáveis relacionadas à localização geográfica dos fornecedores (como região Sudeste e Distrito Federal) não apresentaram associação significativa com a recontratação, seja estatística ou preditivamente, o que levou à rejeição das hipóteses correspondentes. Esses resultados indicam que atributos técnicos, reputacionais e relacionais dos fornecedores exercem maior influência nas decisões de contratação do que fatores espaciais ou regionais.

Os achados deste estudo abrem possibilidades para desdobramentos futuros, como a replicação da análise em outras categorias de compras públicas (por exemplo, saúde, educação ou infraestrutura), o enriquecimento de dados a partir de fontes adicionais (como Receita Federal, Serpro ou TCU) e a aplicação de modelos causais avançados, como propensity score matching ou abordagens baseadas em inferência contrafactual.

Por fim, os resultados obtidos fornecem subsídios relevantes para o aprimoramento de mecanismos objetivos de avaliação de fornecedores e para o desenvolvimento de políticas de recontratação mais alinhadas aos princípios da eficiência, do desempenho institucional e da integridade nas contratações públicas.

## Referências

Aboelazm, K. S., & Dganni, K. M. (2025). Public procurement contracts futurity: Using of artificial intelligence in a tender process. *Corporate Law & Governance Review*, 7(1), 60–72.

<https://doi.org/10.22495/clgrv7i1p6>

Akaba, T. I., Norta, A., Udokwu, C., & Draheim, D. (2020). A framework for the adoption of blockchain-based e-procurement systems in the public sector. In M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. Dwivedi, & M. Mäntymäki (Eds.), *Responsible design, implementation and use of information and communication technology* (pp. 3–14). Springer. [https://doi.org/10.1007/978-3-030-44999-5\\_1](https://doi.org/10.1007/978-3-030-44999-5_1)

Althabatah, A., Yaqot, M., Menezes, B., & Kerbache, L. (2023). Transformative procurement trends: Integrating Industry 4.0 technologies for enhanced procurement processes. *Logistics*, 7(3), 63.

<https://doi.org/10.3390/logistics7030063>

Amann, M., Roehrich, J. K., Eßig, M., & Harland, C. (2014). Driving sustainable supply chain management in the public sector: The importance of public procurement in the European Union. *Supply Chain Management*, 19(3), 351–366. <https://doi.org/10.1108/SCM-12-2013-0447>

Baptista, I. de S., Santos, J. dos, & Vieira, P. S. (2020). Analysis of antecedents and their influences on sustainable public procurement: A structural equation modeling approach. *Sustainability*, 12(14), 5598. <https://doi.org/10.3390/su12145598>

Bergman, M. A., & Lundberg, S. (2013). Tender evaluation and supplier selection methods in public procurement. *Journal of Purchasing & Supply Management*, 19(2), 73–83.

<https://doi.org/10.1016/j.pursup.2013.02.001>

Bitzidis, N., Dimitriadis, S., Karavasilis, G., Kehris, E., & Vrana, V. (2020). Internal customer satisfaction from an e-procurement information system: The case of Greek electronic public

procurement system (ESIDIS). In A. Kavoura, E. Kefallonitis, & P. Theodoridis (Eds.), *Strategic innovative marketing and tourism* (pp. 61–69). Springer. [https://doi.org/10.1007/978-3-030-36126-6\\_8](https://doi.org/10.1007/978-3-030-36126-6_8)

Brasil. (1998). *Emenda Constitucional nº 19, de 4 de junho de 1998*. Recuperado em 19 de junho de 2025, de [https://www.planalto.gov.br/ccivil\\_03/constituicao/emendas/emc/emc19.htm](https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc19.htm)

Caserta, M., Ferrante, L., & Fontana, S. (2025). Too big to be efficient? The role of size in public procurement performance. *Economic Analysis and Policy*, 76, 2049–2069.  
<https://doi.org/10.1016/j.eap.2022.10.014>

CGU - Portal da Transparência. Sanções. CEIS – Cadastro de Empresas Inidôneas e Suspensas. Disponível em <https://portaldatransparencia.gov.br/download-de-dados/ceis> . Acesso em 27/05/2025.

CGU - Portal da Transparência. Sanções. CNEP – Cadastro Nacional de Empresas Punidas. Disponível em <https://portaldatransparencia.gov.br/download-de-dados/cnep> . Acesso em 27/05/2025.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cooper, D. R., & Schindler, P. S. (2014). *Business research methods* (12th ed.). McGraw-Hill Education.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Čeke, D., Buzadija, N., & Kunosić, S. (2022). Enhancing transparency and fairness in public procurement process with the support of blockchain technology: A smart contract based approach. In *Proceedings of the 2022 IEEE International Conference on Information Technology and Systems* (pp. 1–6). IEEE. <https://doi.org/10.1109/INFOTEH53737.2022.9751322>

Edler, J., & Georghiou, L. (2007). Public procurement and innovation—Resurrecting the demand side. *Research Policy*, 36(7), 949–963. <https://doi.org/10.1016/j.respol.2007.03.003>



Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.

Hellberg, R. (2023). Swedish public procurement and the defence industry: Obstacles and opportunities. *Journal of Defense Analytics and Logistics*, 7(2), 103–137. <https://doi.org/10.1108/JDAL-12-2022-0015>

Hodosi, G., Linde, E., & Rusu, L. (2023). Risks in cloud computing relationships: A study of large public buying organizations in Sweden. *Procedia Computer Science*, 219, 486–494. <https://doi.org/10.1016/j.procs.2023.01.312>

Hueskes, M., Verhoest, K., & Block, T. (2017). Governing public–private partnerships for sustainability: An analysis of procurement and governance practices. *International Journal of Project Management*, 35(6), 1184–1195. <https://doi.org/10.1016/j.ijproman.2017.02.020>

Jaccard, J., & Jacoby, L. (2010). *Theory construction and model-building skills: A practical guide for social scientists*. The Guilford Press.

Kademeteme, E., & Bvuma, S. (2023). Using blockchain technology to improve the integrity and transparency of procurement processes between SMMEs and government: A systematic literature review. *The Journal of the British Blockchain Association*, 7(1), 1–12. [https://doi.org/10.31585/jbba-7-1-\(1\)2024](https://doi.org/10.31585/jbba-7-1-(1)2024)

Koala, K., & Steinfeld, J. (2018). Theory building in public procurement. *Journal of Public Procurement*, 18(4), 282–305. <https://doi.org/10.1108/JOPP-11-2018-017>

Kundu, O., Uyarra, E., Ortega-Argiles, R., Tirado, M. M., Kitsos, T., & Yuan, P.-Y. (2024). Impacts of policy-driven public procurement: A methodological review. *Science and Public Policy*, 52(1), 50–64. <https://doi.org/10.1093/scipol/scae058>

Lenderink, B., Halman, J. I. M., & Voordijk, H. (2022). Innovation and public procurement: From fragmentation to synthesis on concepts, rationales and approaches. *Innovation: The European Journal of Social Science Research*, 35(4), 650–674. <https://doi.org/10.1080/13511610.2019.1700101>

Maciel, G. H. C. P., Grossi, M. E. D., Souza Junior, C. V. N., & Silva Junior, L. H. (2024). A eficiência e celeridade nas compras públicas: Um estudo sobre os principais fatores de influência. *Revista Controle*, 22(2), 159–193. <https://doi.org/10.35886/roda.v.2122.923>

Malacina, I., Karttunen, E., Jääskeläinen, A., Lintukangas, K., Heikkilä, J., & Kähkönen, A.-K. (2023). Capturing the value creation in public procurement: A practice-based view. *Journal of Purchasing & Supply Management*, 28(2), 100745. <https://doi.org/10.1016/j.pursup.2021.100745>

Mandala, N., Renson, A. I., & Kipketer, T. S. (2024). The impact of technology adoption on efficiency and transparency in public procurement processes in Kenya. *European Scientific Journal, ESJ*. <https://doi.org/10.19044/esipreprint.1.2024.p483>

Márquez Lozada, D., & Vela Aspajo, J. J. M. (2024). Sistema de gestión de contratos en las Entidades Públicas entre el año 2019 - 2023. *EconConnections*, 2(2), 13–23. <https://doi.org/10.47422/econconnections.v2i2.15>

Mircea, M., Stoica, M., & Ghilic-Micu, B. (2022). Analysis of the impact of blockchain and Internet of Things (BIoT) on public procurement. *IEEE Access*, 10, 66566–66583. <https://doi.org/10.1109/ACCESS.2022.3182656>

Morettin, P. A., & Bussab, W. O. (2010). *Estatística básica* (6ª ed.). Saraiva.

Portal de Compras do Governo Federal. Contratos Gov.br. Disponível em <https://contratos.comprasnet.gov.br/transparencia/contratos?categorias=%5B%2256%22%5D> . Acesso em 23/05/2025.

Ramos, R. B., & Oliveira, D. R. (2023). Eficiência relativa de compras públicas e benchmarking no Instituto Federal de Educação, Ciência e Tecnologia do Amazonas. EMPRAD, 23-24 de novembro, FEA USP. São Paulo.

Sánchez-Graells, V. (2024). *Digital technologies and public procurement*. Oxford University Press.  
<https://doi.org/10.1093/oso/9780198866770.001.0001>

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Allyn & Bacon.

Tas, B. K. O. (2020). Effect of public procurement regulation on competition and cost-effectiveness. *Journal of Regulatory Economics*, 58(1), 59–77. <https://doi.org/10.1007/s11149-020-09409-w>

Triola, M. F. (2010). *Elementary statistics* (11th ed.). Pearson Education.

Uyarra, E., Edler, J., Garcia-Estevez, J., Georgiou, L., & Yeow, J. (2014). Barriers to innovation through public procurement: A supplier perspective. *Technovation*, 34(10), 631–645.  
<https://doi.org/10.1016/j.technovation.2014.04.003>

Velasco, R. B., Carpanese, I., Interiano, R., Paulo Neto, O. C. G., & Ribeiro, C. C. (2020). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, 28(5), 2297–2325. <https://doi.org/10.1111/itor.12811>