Universidade de Brasília
Departamento de Estatística

Estimação do valor da hora-aluno para cursos ofertados pelo Pronatec.

Gilson Ricardo Daniel

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

Brasília
2021

Gilson Ricardo Daniel

Estimação do valor da hora-aluno para cursos ofertados pelo Pronatec.

Orientador(a): Profª Drª Ana Maria Nogales Vasconcelos

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

Brasília
2021

Sumário

1 Introdução 5

2 Contextualização 7

2.1 Educação Profissional e Tecnológica 7

2.2 Regulamentação 8

2.3 Execução Orçamentária 10

2.4 Pronatec na Rede Federal 12

2.5 Problema Identificado 13

3 Objetivos 14

3.1 Objetivo Geral 14

3.2 Objetivos Específicos 14

4 Metodologia 15

4.1 Análise de Variância 15

4.1.1 Suposições do Modelo ANOVA 16

4.1.2 Hipótese Clássica da ANOVA 17

4.2 Base de Dados 19

5 Análise e Resultados 24

5.1 Análise Descritiva dos Dados 25

5.1.1 Análise Descritiva por Eixo Tecnológico 26

5.1.2 Análise Descritiva por Região 29

5.1.3 Análise Descritiva por Tipo de Instituição 31

5.1.4 Transformação da Variável Resposta 33

5.1.5 Valores Discrepantes da Variável Resposta 34

5.2 Análise de Variância 1: com Dados Discrepantes 35

5.2.1 Análise dos Resíduos ANOVA 1 36

5.3 Análise de Variância 2: sem Dados Discrepantes 38

5.3.1 Análise dos Resíduos ANOVA 2 38

5.4 Estimação dos Coeficientes do Modelo 2 40

5.5 Aplicação do Modelo 2 nos Dados de Teste 40

5.6 Valores das Horas-Aluno. 41

6 Conclusão 42

7 Referências 44

8 Apêndices 45

8.1 Apêndice 1: Estimativas Coeficientes do Modelo 45

8.2 Apêndice 2: Valores Estimados e Sugeridos para Hora-Aluno 50

8.3 Apêndice 3: Código do R-Studio 52

RESUMO

O Programa Nacional de Acesso ao Ensino Técnico e Emprego - Pronatec, foi instituído pela Lei nº 12.513/2011, com a finalidade de ampliar a oferta de Educação Profissional e Tecnológica - EPT no país. Após nove anos de ofertas regulares, em 2020, devido ao cenário de pandemia pela Covid-19, foi necessário adaptar a oferta de cursos, focando em cursos de qualificação profissional ou formação inicial e continuada - FIC, na modalidade de Educação a Distância - EaD, ofertados pelas instituições da Rede Federal de Educação Profissional, Científica e Tecnológica - RFEPCT. Ao fim do ano, verificou-se que o valor da hora-aluno executado pelas instituições de ensino ficaram abaixo do valor por elas proposto e aprovado pelo Ministério da Educação - MEC. Assim, há evidências de que o valor efetivo da hora-aluno é menor do que o valor de referência, atualmente estipulado em R$ 4,50 (quatro reais e cinquenta centavos). Por meio da técnica de Análise de Variância - ANOVA, evidenciou-se que três fatores influenciam no valor da hora-aluno: (i) o eixo tecnológico do curso, (ii) a região geográfica onde se localiza a instituição ofertante, e (iii) o tipo de instituição. Adotando-se o modelo sem os valores discrepantes para a variável resposta, que indica o valor da hora-aluno efetivamente executado, são sugeridos valores para a hora-aluno de diferentes combinações entre eixos tecnológicos, regiões geográficas e tipos de instiuições, de maneira a auxiliar a tomada de decisões dos gestores públicos do Pronatec.

Palavras-chave: Pronatec; Educação Profissional; Rede Federal; políticas educacionais.

1 Introdução

O Programa Nacional de Acesso ao Ensino Técnico e Emprego – Pronatec foi instituído por meio da Lei nº 12.513, de 26 de outubro de 2011, com a finalidade de ampliar a oferta de educação profissional e tecnológica no país. Dentre os seus objetivos, destacam-se: (i) expandir, interiorizar e democratizar a oferta de cursos de educação profissional técnica de nível médio e de cursos de qualificação profissional, ou formação inicial e continuada – FIC, nas modalidades de educação presencial e de educação a distância – EaD; (ii) contribuir para a melhoria da qualidade do ensino médio público, por meio da articulação com a educação profissional; e (iii) ampliar as oportunidades educacionais dos trabalhadores, por meio do incremento da formação e qualificação profissional.

Para o atingimento dos seus objetivos, o Programa conta com ações de incentivo e fomento à ampliação de vagas e à expansão das redes públicas de ensino, incluindo a Rede Federal de Educação Profissional, Científica e Tecnológica – RFEPCT, as redes es-

taduais, distrital e municipais de educação, além das instituições de ensino que compõem os Sistemas Nacionais de Aprendizagem – SNA. Dentre tais ações, foi criada a Bolsa-Formação, em duas modalidades: (i) Bolsa-Formação Estudante, para o financiamento de cursos de educação profissional técnica de nível médio, os denominados cursos técnicos; e (ii) Bolsa-Formação Trabalhador, para o financiamento dos cursos de qualificação profissional. Ressalta-se que, além de fomentar a oferta de cursos de educação profissional e tecnológica – EPT nas redes públicas e nas instituições dos SNA, o Pronatec, via Bolsa-Formação, permite o financiamento da União para a oferta de vagas em cursos de EPT ofertados por instituições privadas de ensino, tanto por instituições privadas de ensino superior – IPES quanto por escolas técnicas de nível médio – ETNM.

O Pronatec surgiu como uma agenda governamental de alta prioridade, sendo utilizado, inclusive, como plataforma política durante período eleitoral. Ao todo, de 2011 a 2020, foram investidos mais de R$ 14 bilhões nas suas ações, as quais resultaram em cerca de 5 milhões de matrículas em cursos de educação profissional e tecnológica em todas as regiões do Brasil. A partir de 2016, os investimentos no programa diminuíram substancialmente, passando de R$ 3,5 bilhões em 2015 para cerca de R$ 1,6 bilhão em 2016, sendo que a cada exercício posterior a diminuição dos recursos ficou mais evidente, chegando a R$ 45 milhões em 2019.

Considerando a diminuição dos recursos orçamentários disponíveis e a situação de pandemia pela Covid-19 instaurada a partir do início de 2020, a destinação dos recursos do programa precisou ser replanejada, de acordo com as restrições das atividades presenciais impostas pelas medidas de distanciamento social. Nesse sentido, o foco do programa, em 2020, foi o fomento de vagas em cursos de qualificação profissional, com duração média de 200 horas, ofertados totalmente na modalidade de EaD, em parceria com as instituições da RFEPCT.

Tais instituições apresentaram propostas para a oferta de vagas, propondo, inclusive, o valor da hora-aluno a ser financiado pelo Governo Federal. Ao longo da execução dos cursos, verificou-se que os valores totais aprovados foram superestimados pelas instituições ofertantes, pois essas não necessitaram de todo o recurso previsto para a execução dos cursos.

Portanto, faz-se necessário um estudo aprofundado para a estimação do valor efetivo da hora-aluno praticado pelas instituições da RFEPCT na oferta de cursos de qualificação profissional na modalidade de EaD, com o objetivo de evitar as sobras orçamentárias e aumentar a eficiência na aplicação dos recursos do programa, que a cada ano estão nominalmente menores.

2 Contextualização

2.1 Educação Profissional e Tecnológica

Conhecer a trajetória da Educação Profissional e Tecnológica – EPT no Brasil é importante para que o contexto atual dessa modalidade educacional seja entendida, inclusive para o estudo proposto, o qual foca na oferta de cursos de EPT por instituições federais de ensino, em todas as regiões brasileiras, em um momento singular, de pandemia mundial. Assim, a seguir é apresentado um resumo do levantamento histórico da EPT realizado por Santos e Marchesan (2017).

A EPT tem como marco inicial o ano de 1909, quando Nilo Peçanha, ao assumir a Presidência da República, criou dezenove Escolas de Aprendizes Artífices em cada uma das então capitais dos Estados brasileiros, por meio do Decreto nº 7.566, de 23 de setembro de 1909. Essas escolas surgiram com o objetivo de ofertar ensino profissional gratuito para qualificação de mão-de-obra para o atendimento das demandas dos setores produtivos e das economias locais, pois cada unidade, independentes entre si, ofereciam cursos distintos, de acordo com a necessidade da região geográfica onde estava instalada.

Com a promulgação da Constituição de 1937, por Getúlio Vargas, a educação técnica passou a ser vista como um instrumento estratégico para o desenvolvimento social e econômico do país, ocorrendo, por meio da nova Constituição, a transformação das Escolas de Aprendizes Artífices em Liceus Industriais, os quais passaram a trabalhar em sintonia com a expansão da indústria.

Já em 1942, os Liceus Industriais passaram a ser denominados por Escolas Industriais e Técnicas, quando o ensino profissionalizante e técnico foi equiparado ao nível médio. Tal ação foi uma das consequências da alteração da estrutura educacional brasileira, ocorrida no mesmo ano. Porém, em 1959, as Escolas Industriais e Técnicas foram transformadas em Escolas Técnicas Federais, detentoras de autonomia administrativa e pedagógica. Esse momento foi muito importante para a EPT, pois a partir de então, essa modalidade de educação foi equiparada ao ensino acadêmico, por meio da Lei de Diretrizes e Bases da Educação Nacional de 1961.

Com o regime militar, mais especificamente na década de 1970, houve uma grande reforma na educação brasileira, tendo como uma das ações a instituição da profissionalização como obrigatória, fazendo com que todos os cursos de nível médio tivessem caráter profissionalizante. Assim, e com a grande demanda por educação profissional, em 1978 foram criados os três primeiros Centros Federais de Educação Tecnológica – Cefet, nos Estados de Minas Gerais, Paraná e Rio de Janeiro, os quais foram consolidados como unidades padrão para o ensino técnico e profissional.

A nova, e atual, Lei de Diretrizes e Bases da Educação Nacional, Lei nº 9.394, de 20 de dezembro de 1996, transformou, mais uma vez, a EPT no Brasil, favorecendo a publicação do Decreto Lei nº 2.208/1996, o qual separou as disciplinas da educação geral com aquelas da educação profissional. Porém, com o Decreto nº 5.154/2004, a educação profissional passou a integrar, novamente, a educação básica, em articulação com o ensino médio.

No ano de 2008, foi instituída a Rede Federal de Educação Profissional, Científica e Tecnológica – RFEPCT, por meio da Lei nº 11.892/2008, criando 38 Institutos Federais de Educação, Ciência e Tecnologia – IF, mantendo 2 Cefet, o Colégio Pedro II, a Universidade Tecnológica Federal do Paraná – UTFPR e 22 Escolas Técnicas Vinculadas às Universidades Federais. Portanto, a RFEPCT é composta, atualmente, por 64 instituições federais de ensino, as quais somam mais de 600 unidades em todos o território nacional.

Por fim, em 2011, foi criado o Programa Nacional de Acesso ao Ensino Técnico e Emprego – Pronatec, pela Lei nº 12.513/2011, com os seguintes objetivos:

I - expandir, interiorizar e democratizar a oferta de cursos de educação profissional técnica de nível médio presencial e a distância e de cursos e programas de formação inicial e continuada ou qualificação profissional;

II - fomentar e apoiar a expansão da rede física de atendimento da educação profissional e tecnológica;

III - contribuir para a melhoria da qualidade do ensino médio público, por meio da articulação com a educação profissional;

IV - ampliar as oportunidades educacionais dos trabalhadores, por meio do incremento da formação e qualificação profissional;

V - estimular a difusão de recursos pedagógicos para apoiar a oferta de cursos de educação profissional e tecnológica;

VI - estimular a articulação entre a política de educação profissional e tecnológica e as políticas de geração de trabalho, emprego e renda. (Incluído pela Lei nº 12.816, de 2013)

2.2 Regulamentação

A regulamentação da assistência da União via Bolsa-Formação do Pronatec é dada, atualmente, pela Portaria MEC nº 817, de 13 de agosto de 2015, e suas alterações, a qual define os atores do programa e suas respectivas competências. O Ministério da Educação – MEC, por intermédio de sua Secretaria de Educação Profissional e Tecnológica – SETEC/MEC e de sua atual Subsecretaria de Tecnologia da Informação e Comunicação

– STIC/SE/MEC, atua como gestor nacional da política e como mantenedor e operacionalizador dos sistemas eletrônicos para a oferta e a execução da Bolsa-Formação, sendo que atualmente é utilizado o Sistema Nacional de Informações da Educação Profissional e Tecnológica – Sistec. O Fundo Nacional de Desenvolvimento da Educação – FNDE, autarquia vinculada ao MEC, é responsável pela execução orçamentária e financeira das ações do Pronatec, via ação orçamentária específica prevista na Lei Orçamentária Anual – LOA da União. Os Ministérios e outros órgãos e entidades da Administração Pública Federal atuam como parceiros demandantes, indicando as áreas econômicas nas quais há demanda por qualificação profissional. Por fim, os parceiros ofertantes são as redes de ensino que ofertam, efetivamente, os cursos em todos o país, mediante autorização da SETEC/MEC.

O repasse, pela União, aos parceiros ofertantes, é baseado nas vagas aprovadas e devidamente homologadas no módulo "Pronatec" do Sistec. O procedimento de homologação garante ao parceiro ofertante que as vagas foram efetivamente aprovadas e estão liberadas para execução, pois somente a homologação no sistema é possibilidade ao ofertante registrar as matrículas dos alunos beneficiários do programa.

O montante a ser repassado é calculado de acordo com a hora-aluno executada, ou seja, a unidade de mensuração da execução do Pronatec é a hora-aluno. Ao longo doa anos, o valor de referência da hora-aluno foi modificado, sendo que, a partir de 2013, foi fixado em R$ 10,00 (dez reais) a hora-aluno para cursos ofertados na modalidade de educação presencial e em R$ 4,50 (quatro reais e cinquenta centavos) a hora-aluno para cursos ofertados na modalidade de EaD. Portanto, tomando como exemplo um curso técnico com carga horária de 1.000 horas ofertado na modalidade de educação presencial, cada matrícula registrada no sistema corresponde a R$ 10.000,00 (dez mil reais), ou seja, no caso hipotético do preenchimento de uma turma com 30 matrículas, o parceiro ofertante terá o direito a receber da União o montante de R$ 300.000,00 (trezentos mil reais) para a execução do objeto pactuado. Caso o mesmo curso fosse ofertado na modalidade de EaD, o valor por aluno seria de R$ 4.500,00 (quatro mil e quinhentos reais).

O instrumento utilizado para a efetivação dos repasses aos parceiros ofertantes varia de acordo com a rede ofertante. As instituições da RFEPCT, por sua característica federal, recebem os recursos do Pronatec via Termo de Execução Descentralizada – TED. As redes públicas estaduais, distrital e municipais de educação e os SNA recebem os recursos via Transferência Direta – TD, autorizada na Lei nº 12.513/2011. Já as IPES e as ETNM também recebem os recursos via TD, porém não se fazendo necessária a apresentação de prestação de contas ao fim das ofertas dos cursos, pois os pagamentos são realizados mensalmente, mediante validação do aluno, no Sistec, nas suas frequências mensais, procedimento que funciona como um controle social da execução do objeto previsto.

2.3 Execução Orçamentária

Para o financiamento dos cursos do Pronatec são utilizadas dotações disponíveis na Lei Orçamentária Anual – LOA da União. Como o programa foi instituído em outubro de 2011, não havia, então, ação orçamentária específica para o Pronatec, sendo, portanto, necessária a criação de uma ação orçamentária específica, a 20RW – Apoio à Formação Profissional e Tecnológica. Os recursos de tal ação foram destinados exclusivamente às ações do Pronatec de 2011 a 2019, porém, a partir de 2013, o nome sofreu uma pequena alteração, passando a ser denominada como 20RW – Apoio à Formação Profissional, Científica e Tecnológica.

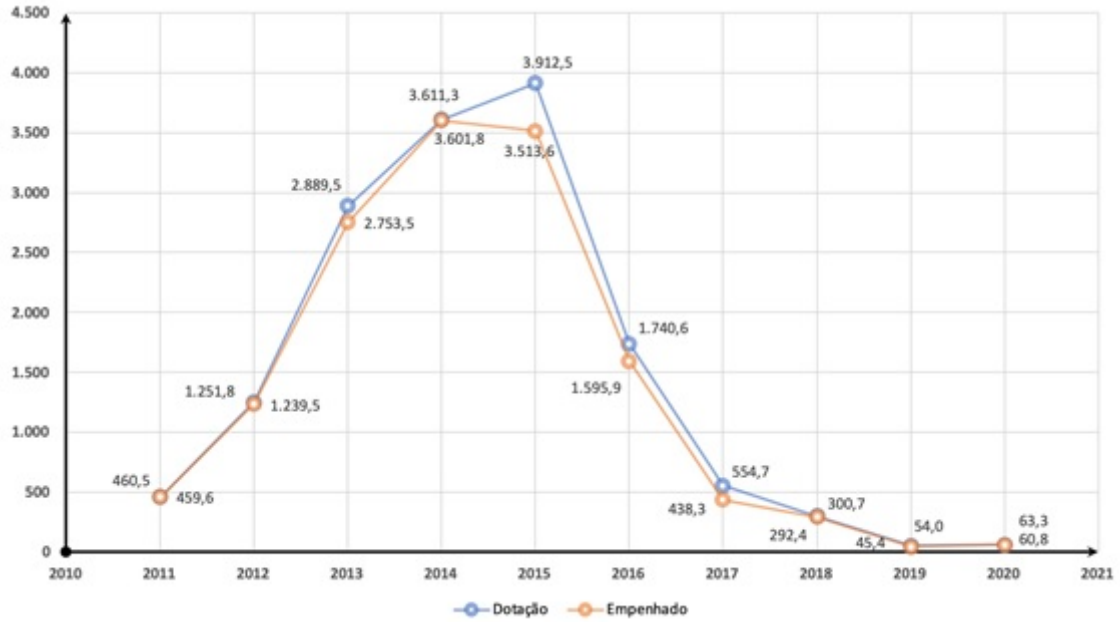
A partir de 2020, a ação orçamentária 20RW foi extinta, dando espaço à nova ação 21B4 – Fomento ao Desenvolvimento e Modernização dos Sistemas de Ensino de Educação Profissional e Tecnológica, por meio da qual as ações do Pronatec continuaram a ser financiadas. A diferença entre a 20RW e a 21B4 é que esta última permite o financiamento de despesas de custeio e de despesas de capital, enquanto a extinta 20RW permitia somente o financiamento de despesas de custeio, o que influenciou diretamente a implementação dos cursos do Pronatec, os quais não puderam contar com recursos de investimento ou de material permanente, por exemplo.

Sendo uma política pública nacional e utilizada como plataforma governamental, o Pronatec contou com dotações orçamentárias bilionárias. No Gráfico 1, pode-se verificar a evolução das dotações finais anuais destinadas ao programa, assim como os respectivos empenhos, os quais indicam a execução orçamentária anual efetiva.

De 2011 a 2020, foram destinados ao Pronatec R$ 14,8 bilhões de reais, dos quais R$ 14,0 bilhões foram empenhados, de acordo com os dados públicos disponíveis no Sistema Integrado de Planejamento e Orçamento – SIOP, o que indica cerca de 94,6% de execução orçamentária, ou seja, de todos os recursos destinados ao Pronatec, 94,6% foram efetivamente gastos pelo Ministério da Educação com o financiamento de cursos técnicos e de cursos de qualificação profissional. Pode-se considerar uma ótima execução orçamentária.

Para entender melhor a distribuição dos recursos do Pronatec, pode-se categorizar a execução orçamentária por rede de ensino, conforme ilustrado no Gráfico 2. Os dados de empenho orçamentário por ano e por tipo de rede ofertante podem ser consultados via extração de relatórios do Tesouro Gerencial – TG, que é o sistema de consulta à execução orçamentária, financeira e patrimonial do Governo Federal no âmbito do Orçamento Fiscal e da Seguridade Social. O TG é um sistema complementar ao Sistema Integrado de Administração Financeira do Governo Federal – SIAFI, por meio do qual é possível a emissão de relatórios gerenciais dos dados constantes no SIAFI. Como o TG não possui

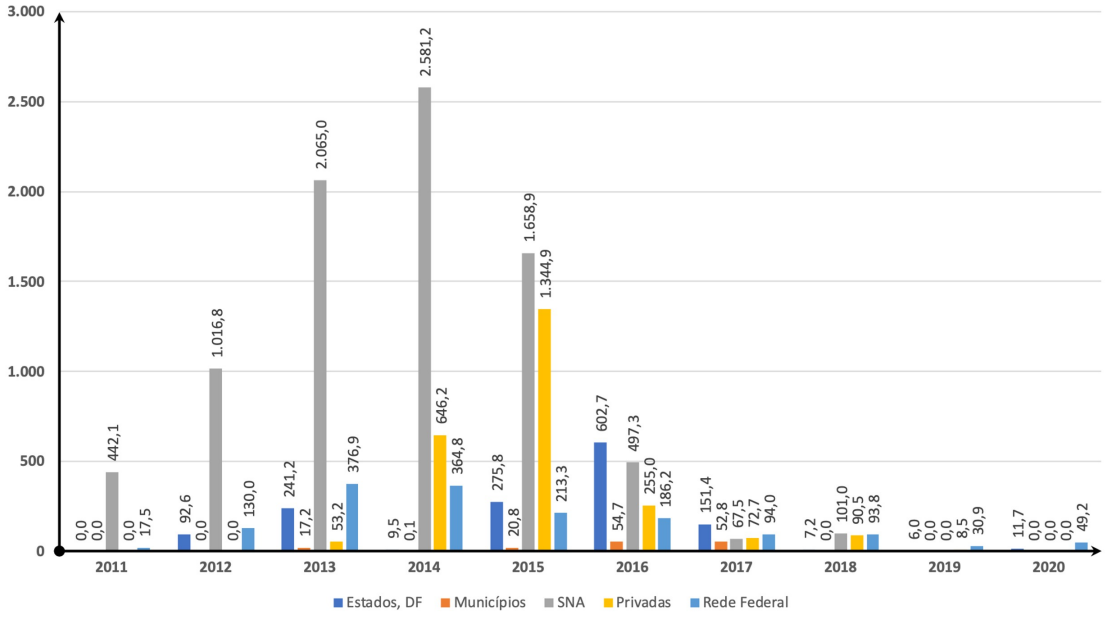
Gráfico 1: Evolução orçamentária do Pronatec (em milhões de R$).



Fonte: Elaborado pelo autor, com dados do Sistema Integrado de Planejamento e Orçamento – SIOP, acesso público.

acesso público, somente usuários devidamente cadastrados e autorizados podem acessar o sistema, os relatórios de execução de qualquer ação orçamentária podem ser solicitados aos órgãos e entidades executoras, com base na Lei nº 12.527, de 18 de novembro de 2011, conhecida como Lei de Acesso à Informação – LAI.

Gráfico 2: Evolução orçamentária do Pronatec por rede ofertante (em milhões de R$).



Fonte: Elaborado pelo autor, com dados do Tesouro Gerencial.

Conforme evidenciado no Gráfico 2, as instituições que compõem os SNA, Sistema Nacional de Aprendizagem Comercial – SENAC, Sistema Nacional de Aprendizagem Industrial – SENAI, Sistema Nacional de Aprendizagem Rural – SENAR e Sistema Nacional de Aprendizagem do Transporte – SENAT, receberam 60,2% de todos os recursos executados no âmbito do Pronatec, sendo seguidos pelas redes privadas (17,7%), pela RFEPCT (11,1%), pelas redes estaduais e distrital (10,0%) e pelas redes municipais de educação (1,0%). Porém, tal padrão de distribuição orçamentária pode ser observado de 2011 a 2015, sendo que em 2016 e 2017 as redes estaduais e distrital receberam a maior parte dos recursos, em 2018 os SNA, os Estados e a rede privada receberam praticamente o mesmo montante de recursos e, ao considerarmos somente os orçamentos de 2019 e de 2020, as instituições da RFEPCT receberam, juntas, mais de 75,3% de todos os recursos executados no Pronatec.

2.4 Pronatec na Rede Federal

A RFEPCT foi instituída pela Lei nº 11.892, de 29 de dezembro de 2008, e é atualmente composta por 38 Institutos Federais de Educação, Ciência e Tecnologia – IF, por 2 Centros Federais de Educação Tecnológica – CEFET, pela Universidade Tecnológica Federal do Paraná – UTFPR, pelo Colégio Pedro II e por 24 Escolas Técnicas vinculadas às Universidades Federais, assim, todas essas instituições podem atuar como parceiros ofertantes do Pronatec, ofertando cursos técnicos e de qualificação profissional à população financiados via Bolsa-Formação, mediando autorização e pactuação das vagas com a SETEC/MEC.

Em 2019, com a nova composição do Governo Federal, incluindo o corpo gestor da SETEC/MEC, o Pronatec continuou a ser utilizado como a principal política pública para o fomento de vagas em cursos de EPT à sociedade brasileira. Porém, dada a diminuição da dotação orçamentária disponível e o cenário de pandemia causada pela Covid-19 a partir do início de 2020, as instituições da RFEPCT passaram a assumir um papel central na oferta de cursos, dada a facilidade de interlocução e articulação com a SETEC/MEC.

No início de 2020, a SETEC/MEC havia planejado o fomento de novas vagas em cursos técnicos e em cursos de qualificação profissional, tanto na modalidade de educação presencial quanto na modalidade de EaD, a serem ofertadas por diversas redes de ensino. Porém, com a instauração do cenário pandêmico, o planejamento precisou ser ajustado, passando a priorizar a oferta de cursos de qualificação profissional, com média de 200 horas, na modalidade de EaD, com o objetivo de permitir que os cidadãos brasileiros fossem capacitados profissionalmente enquanto as medidas de distanciamento social estivessem em vigor, inclusive com a suspensão das atividades presenciais em diversas redes de ensino.

Atuando nessa diretriz, a SETEC/MEC lançou três chamadas às instituições da RFEPCT ao longo do primeiro semestre de 2020, abrindo a possibilidade para que tais instituições apresentassem propostas para a oferta de cursos de qualificação profissional na modalidade de EaD, a serem financiadas com recursos do Pronatec/Bolsa-Formação. Um dos critérios estabelecidos para as propostas foi o valor flexível, ou seja, cada instituição deveria apresentar a proposta de oferta de cursos considerando o valor efetivo da hora-aluno, desde que tal valor não ultrapassasse o valor máximo de referência, que é de R$ 4,50 (quatro reais e cinquenta centavos) para cursos EaD. Ao receber as propostas, a SETEC/MEC identificou que, das 30 instituições que tiveram as propostas aprovadas, somente 5 apresentaram valores para a hora-aluno menores do que o valor de referência, sendo que as 25 demais instituições utilizaram como base o valor de R$ 4,50 (quatro reais e cinquenta centavos) para a hora-aluno.

Assim, para as 167.100 vagas inicialmente aprovadas, foram previstos R$ 64,8 milhões. Para cada instituição ofertante, o valor aprovado é calculado pelo somatório da multiplicação do quantitativos de vagas em cada curso pela carga horária do respectivo curso, o que resulta no total de horas-aluno a serem ofertadas, o qual é multiplicado pelo valor de referência da hora-aluno. Por exemplo, se uma determinada instituição prevê a oferta de 1.000 vagas em curso de 200 horas, o objeto previsto é a oferta de 200.000 horas-aluno; se para a execução de cada hora-aluno é previsto o valor máximo de referência, que é de R$ 4,50 (quatro reais e cinquenta centavos) para cursos EaD, então, para essa oferta, é aprovado um valor de R$ 900 mil.

Como nem todas as vagas ofertadas são efetivamente ocupadas, a SETEC/MEC determinou que as descentralizações orçamentárias para as vagas aprovadas em 2020 ocorreriam em duas parcelas: a primeira, correspondente a 50% do montante total aprovado para viabilizar o início da oferta e liberado somente após a publicação do processo seletivo no sítio eletrônico da instituição; e a segunda, após a instituição registrar as matrículas no módulo "Pronatec" do Sistec, até o limite das matrículas efetivadas. Esse mecanismo de liberação dos recursos foi proposto pela SETEC/MEC com o objetivo de evitar sobras orçamentárias nas instituições, o que acarretaria na ineficiência do gasto público e na devolução de eventuais saldos não utilizados, que, dependendo da data de devolução, não poderiam ser aplicados em outras ações, dado o princípio da anualidade do orçamento público.

2.5 Problema Identificado

Foram três fases de pactuação no ano de 2020: a primeira, com 24 instituições ofertando 82.120 vagas cujos cursos estavam previstos para iniciar entre abril e maio de 2020; a segunda, com 18 instituições ofertando 60.220 vagas, com previsão de início dos

cursos em junho de 2020; e a terceira, com 11 instituições ofertando 24.760 vagas, com início dos cursos em agosto de 2020. Ressalta-se que uma instituição pode ter atuado como parceiro ofertante em mais de uma fase de pactuação.

Com a liberação inicial de 50% dos recursos totais previstos, que foram calculados de acordo com as propostas apresentadas pelos parceiros ofertantes e aprovadas pela SETEC/MEC, os cursos foram iniciados, sendo as respectivas matrículas devidamente cadastradas pelas instituições ofertantes no Sistec, permitindo o acompanhamento e o monitoramento dessa execução pela SETEC/MEC.

A partir do mês de agosto de 2020, as instituições de ensino ofertante começaram a solicitar o repasse da segunda, e última parcela, dos recursos previstos. Para tanto, a SETEC/MEC calculava o montante da segunda parcela de acordo com as matrículas válidas registradas no Sistec, assim, a segunda seria limitada pelas matrículas. Porém, foram identificadas duas situações: (i) algumas instituições solicitaram um montante menor de recursos do que poderiam receber pelas matrículas registradas no sistema; e (ii) algumas instituições efetuaram a devolução de recursos ao fim do exercício, pois houve sobra orçamentária.

Neste sentido, foi possível verificar que os valores propostos pelos parceiros ofertantes, para essas pactuações específicas, estavam superdimensionados, ou seja, acima do valor efetivo da hora-aluno praticada. Assim, um estudo mais aprofundado sobre essa temática se faz necessário.

3 Objetivos

3.1 Objetivo Geral

- Estimar o valor efetivo da hora-aluno para os cursos de qualificação profissional na modalidade de educação a distância ofertados pelas instituições da Rede Federal com os recursos do Pronatec.

3.2 Objetivos Específicos

- Analisar a diferença entre os valores propostos pelas instituições e os valores executados.
- Identificar os fatores que influenciam na diferença dos valores das horas-aluno efetivamente executados.
- Propor modelo para a estimação do valor efetivo da hora-aluno.

4 Metodologia

A pesquisa possui uma abordagem quantitativa, com a variável dependente o valor da hora-aluno, a qual será estudada sob a ótica de diversas categorias, como o tipo de instituição ofertante, a região brasileira do parceiro ofertante e o eixo tecnológico do curso, assim como outras categorias que sejam identificadas como de interesse ao longo da análise.

O método será indutivo, partindo de um estudo de caso, instituições da Rede Federal ofertando cursos EaD, podendo ser generalizado para as demais redes, dadas as limitações apresentadas pelas particularidades de cada uma. Portanto, o valor estimado será para a Rede Federal, sendo outros estudos necessários para testar o modelo proposto para as ofertas das demais redes.

A realidade estudada é multivariada, ou seja, uma determinada variável dependente de interesse é, ou pode ser causada por mais de um fator. Assim, embora a pesquisa proponha descrever a relação proposta entre alguma causa e algum efeito, deve-se ter em mente que um efeito pode ter várias causas. A noção de causalidade envolve, principalmente, relações determinísticas, aquelas onde, se uma causa ocorre, então o efeito ocorrerá com certeza.

4.1 Análise de Variância

A técnica de análise dos dados será a Análise de Variância - ANOVA. De acordo com Casella e Berger (2014), a ANOVA é uma das técnicas estatísticas mais amplamente utilizadas, cuja ideia principal é o particionamento da variabilidade. Por mais que o nome dessa técnica induza a pensar que seu principal objetivo é o estudo das variâncias, na verdade, o foco é analisar a variabilidade das médias.

Em sua forma mais simples, a ANOVA é um modelo de estimação das médias de diversas populações, frequentemente assumidas como normalmente distribuídas. No entanto, o foco central da ANOVA está no planejamento estatístico, ou seja, sobre como obter o máximo de informações sobre a maioria das populações com a menor quantidade de observações.

O modelo de ANOVA a ser adotado nas análises do presente estudo é dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \alpha_i\beta_j\gamma_k + \epsilon_{ijk} \tag{4.1.1}$$

onde:

i: representa a quantidade de populações/categorias, i, para a primeira variável categórica independente do modelo;

j: representa a quantidade de populações/categorias, j, para a segunda variável categórica independente do modelo;

k: representa a quantidade de populações/categorias, k, para a terceira variável categórica independente do modelo;

Y_{ijk}: valor que a variável de interesse (dependente) assume, em função dos valores das demais variáveis (independentes);

μ: representa a média de todos os valores de Y_{ijk}, independente da divisão das observações por população/categorias das variáveis independentes do modelo;

α_i: representa o efeito único devido às i categorias da primeira variável categórica do modelo;

β_j: representa o efeito único devido às j categorias da segunda variável categórica do modelo;

γ_k: representa o efeito único devido às k categorias da terceira variável categórica do modelo;

$\alpha_i\beta_j$: representa o efeito único devido às interações entre as categorias i da primeira variável categórica e as categorias j da segunda variável categórica do modelo;

$\alpha_i\gamma_k$: representa o efeito único devido às interações entre as categorias i da primeira variável categórica e as categorias k da terceira variável categórica do modelo;

$\beta_j\gamma_k$: representa o efeito único devido às interações entre as categorias j da segunda variável categórica e as categorias k da terceira variável categórica do modelo;

$\alpha_i\beta_j\gamma_k$: representa o efeito único devido às interações entre as categorias i da primeira variável categórica, as categorias j da segunda variável categórica e as categorias k da terceira variável categórica do modelo; e

ϵ_{ijk}: representa os erros aleatórios entre os valores observados e os valores previstos para a variável de interesse (dependente), Y_{ijk}.

4.1.1 Suposições do Modelo ANOVA

Dado o modelo apresentado na seção anterior, Casella e Berger (2014) definem que a suposição mínima necessária antes de ser realizada qualquer estimação é que $E[\epsilon_{ijk}] = 0$ e $Var[\epsilon_{ijk}] < \infty$, para todo i, j e k. Seguindo essas suposições básicas, pode-se estimar os parâmetros τ_i, β_j e γ_k; porém, para fazer qualquer estimação ou teste de intervalo de confiança, é necessário assumir algumas suposições distribucionais, quais sejam:

i $E[\epsilon_{ijk}] = 0$, para todo i, j, k;

ii $Var[\epsilon_{ijk}] = \sigma_i^2 < \infty$, para todo i, j, k;

iii $Cov\left(\epsilon_{i,j,k}, \epsilon_{i^{'},j^{'},k^{'}}\right)$, para todo i, j, k e $i^{'}, j^{'}, k^{'}$, a menos que $i = i^{'}$, $j = j^{'}$ e $k = k^{'}$;

iv Os ϵ_{ijk} são independentes e normalmente distribuídos (erros normais); e

v $\sigma_i^2 = \sigma^2$, para todo i, indicando igualdade de variância, ou homocedasticidade.

Sem a suposição iv, é possível fazer somente a estimação pontual e, possivelmente, procurar estimadores que minimizem a variância dentro de uma mesma categoria, mas não seria possível fazer a estimação intervalar ou teste de hipóteses. Caso os erros assumam uma distribuição que não seja a normal, mesmo possível, a definição de intervalos e a realização de testes podem ser difíceis de derivar, a não ser que haja tamanhos de amostras e de populações razoáveis, que não sejam muito assimétricos, tomando como base o Teorema do Limite Central - TLC.

A suposição de igualdade de variâncias também é bastante importante. É interessante observar que sua importância está associada à suposição de normalidade. Em geral, se houver suspeita de que os dados violam gravemente as suposições da ANOVA, a alternativa, em geral, é tentar transformar os dados não linearmente.

A transformação mais utilizada é a proposta por Box-Cox, da seguinte maneira:

$$Y_i(\lambda) = \begin{cases} ln(X_i), & \text{se } \lambda = 0 \\ \frac{X_i^{\lambda}-1}{\lambda}, & \text{se } \lambda \neq 0 \end{cases} \tag{4.1.2}$$

A transformação de Box-Cox consiste em encontrar um λ tal que os dados transformados $Y_1, Y_2, ..., Y_n$ se aproximem de uma distribuição normal.

4.1.2 Hipótese Clássica da ANOVA

O clássico teste da ANOVA possui como hipóteses:

$$H_0 : \ \mu_1 = \mu_2 = ... = \mu_k \tag{4.1.3}$$

$$H_1 : \ \mu_i \neq \mu_j, \ \text{para algum } i, j \tag{4.1.4}$$

É importante ressaltar que o simples fato de testar a igualdade das médias, muitas vezes, não contempla o objetivo do pesquisador, que pode estar interessado em descobrir,

por exemplo, qual é o melhor tratamento, qual fator apresenta maior eficácia ou produtividade. Assim, o interesse real da ANOVA não está no teste de hipótese, mas sim na estimação, pontual e intervalar, dos efeitos causados por cada tratamento ou categoria.

Combinações lineares, em particular os contrastes, representam um papel extremamente importante na análise de variância. Por meio do entendimento e da análise dos contrastes, pode-se fazer inferências significativas sobre os parâmetros do modelo. Para comparar os efeitos de dois tratamentos, 1 e 2, por exemplo, o contraste é dado por:

$$\left| \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| > t_{N-k,\alpha/2} \tag{4.1.5}$$

A partir dos resultados univariados sobre combinações lineares e a relação entre a hipótese nula na ANOVA, deriva-se um teste de hipótese para testar a hipótese nula da ANOVA, por meio da estatística F, calculada da seguinte forma:

$$F = \frac{\sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{\bar{Y}})/(k-1)}{S_p^2} > F_{k-1,N-k,\alpha}$$

Porém, uma outra maneira de calcular a estatística F e o p-valor das diferenças de médias entre grupos via ANOVA é a utilização da Tebala de Análise de Variância, a qual é dada por:

a: número de categorias do fator A;

b: número de categorias do fator B;

c: número de categorias do fator C;

p: número de variáveis independentes do modelo;

n: número total de observações;

$\bar{y}_{i..}$: média da i-ésima categoria do fator A;

$\bar{y}_{.j.}$: média da j-ésima categoria do fator B;

$\bar{y}_{..k}$: média da k-ésima categoria do fator C; e

$\bar{y}_{ijk}$: média global de todas as observações.

O p-valor relacionado à estatística F é dado por $1 - P(F \leq f)$. Para cada estatística F há um p-valor associado, o qual indicará a rejeição ou a não rejeição da hipótese nula, a qual indica igualdade nos efeitos/médias das diferentes categorias.

Tabela 1: Modelo de tabela de ANOVA.

Fonte de Variação	**Graus de Liberdade (GL)**	**Soma de Quadrados (SQ)**	**Quadrado Médio (QM)**	**Estatística F**
Fator A	$i-1$	$nbc\sum_i(\bar{y}_{i..}-\bar{y}_{...})^2$	$SQ(A)/GL(A)$	$QM(A)/QM(Res)$
Fator B	$j-1$	$nac\sum_j(\bar{y}_{.j.}-\bar{y}_{...})^2$	$SQ(B)/GL(B)$	$QM(B)/QM(Res)$
Fator C	$k-1$	$nab\sum_k(\bar{y}_{..k}-\bar{y}_{...})^2$	$SQ(C)/GL(C)$	$QM(C)/QM(Res)$
Interação AB	$(a-1)(b-1)$	$nc\sum_i\sum_j(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$	$SQ(AB)/GL(AB)$	$QM(AB)/QM(Res)$
Interação AC	$(a-1)(c-1)$	$nb\sum_i\sum_k(\bar{y}_{i.k}-\bar{y}_{i..}-\bar{y}_{..k}+\bar{y}_{...})^2$	$SQ(AC)/GL(AC)$	$QM(AC)/QM(Res)$
Interação BC	$(b-1)(c-1)$	$na\sum_j\sum_k(\bar{y}_{.jk}-\bar{y}_{.j.}-\bar{y}_{..k}+\bar{y}_{...})^2$	$SQ(BC)/GL(BC)$	$QM(BC)/QM(Res)$
Interação ABC	$(a-1)(b-1)(c-1)$	$n\sum_i\sum_j\sum_k(\bar{y}_{ijk}-\bar{y}_{ij.}-\bar{y}_{i.k}-\bar{y}_{.jk}+\bar{y}_{i..}+\bar{y}_{.j.}+\bar{y}_{..k}-\bar{y}_{...})^2$	$SQ(ABC)/GL(ABC)$	$QM(ABC)/QM(Res)$
Resíduos	$n-p-1$	$SQ(T)-\sum SQ(\text{demais termos})$	$SQ(Res)/GL(Res)$	
Total	$n-1$	$\sum_i\sum_j\sum_k(y_{ijk}^2-\bar{y}_{...})^2$		

Fonte: Elaborado pelo autor, com base em Casella e Berger (2014).

4.2 Base de Dados

Uma das regras definidas nos processos de pactuação de vagas com as instituições da Rede Federal em 2020, para cursos FIC EaD, foi a disponibilização, pelo parceiro ofertante, de vagas de contrapartida. Assim, para cada vaga financiada pelo MEC, a instituição parceira deveria ofertar, no mínimo, mais uma vaga nos mesmos cursos das vagas financiadas, como contrapartida institucional. Portanto, a oferta total é composta pelo somatório das vagas financiadas com as vagas de contrapartida. Porém, as vagas financiadas, e as respectivas matrículas, são registradas no módulo "Pronatec" do Sistec, enquanto as vagas de contrapartida são registradas no módulo "Ciclo de Matrículas" do mesmo sistema. Um fator importante a ser considerado para a análise, a consolidação e o saneamento dos dados, é que essas duas bases possuem formatos diferentes, ou seja, a maioria das variáveis são comuns, mas algumas variáveis, essenciais para a análise proposta, constam somente no módulo "Pronatec". Nesse sentido, foi realizada uma análise preliminar para a consolidação dessas duas bases de dados de vagas e matrículas.

Com o objetivo de relacionar essas duas bases de dados de vagas e matrículas, financiadas e de contrapartida, com os dados de execução orçamentária, considerando que o Sistec não possui as variáveis relacionadas à parte orçamentária e financeira das pactuações, foram inseridas, em cada uma das bases, três variáveis: número do TED; valor da hora-aluno proposto; e tipo de vagas (financiadas ou contrapartida). Assim,

será possível relacionar cada turma, e as respectivas vagas e matrículas, com os dados orçamentários, os quais constam em uma terceira base de dados, que é denominada como relatório de controle de execução orçamentária para cada TED que, assim como as duas primeiras bases de dados, foi consolidado e atualizado pela área técnica da SETEC/MEC.

Portanto, para a presente pesquisa, serão utilizadas três bases de dados: (i) relatório de pactuação das vagas financiadas, formatado pela área técnica da SETEC/MEC com base nas matrículas registradas no Sistec; (ii) relatório de pactuação das vagas de contrapartida, formatado pela área técnica da SETEC/MEC; e (iii) relatório de controle de execução orçamentária para cada TED, consolidado pela área técnica da SETEC/MEC. Nas duas primeiras bases, cada observação corresponde a uma turma, contendo todos os dados sobre a turma, como a identificação da instituição ofertante, a identificação do curso, a identificação das datas relacionadas à oferta e as quantidades de vagas ofertadas e de matrículas efetivadas.

As três bases de dados podem ser obtidas via solicitação a ser cadastrada na Plataforma Integrada de Ouvidoria e Acesso à Informação do Governo Federal, disponível em https://falabr.cgu.gov.br/, instrumento que visa a dar cumprimento à Lei nº 12.527, de 18 de novembro de 2011, conhecida como Lei de Acesso à Informação. O acesso aos dados foi possível pois não há variáveis identificadoras dos alunos, estando de acordo com a recente Lei nº 13.709, de 14 de agosto de 2018, denominada Lei Geral de Proteção de Dados – LGPD.

Nos Quadros 1, 2 e 3 são relacionadas as variáveis contidas em cada base de dados utilizada.

Conforme verificado nos Quadros 1 e 2, nas bases de dados referentes às vagas financiadas e às vagas de contrapartida, além das variáveis já disponibilizadas originalmente, foram inseridas as variáveis: "TED", indicando o número do Termo de Execução Descentralizada que possui aquela turma como objeto pactuado, constante na base de dados orçamentária; "valor hora-aluno proposto", indicando o valor da hora-aluno que a instituição parceira propôs na apresentação da proposta de oferta, também constante na base de dados orçamentária; e "tipo de vagas", indicando se as vagas são financiadas ou de contrapartida, variável criada para categorizar as turmas.

Após a inclusão dessas três variáveis, foi realizada uma análise de correspondência entre as variáveis das bases de dados 1 e 2, sendo que as variáveis que não possuíam correspondentes na outra base de dados foram identificadas como "NA" na coluna "Correspondente" dos Quadros 1 e 2. Ademais, algumas variáveis que, considerando a contextualização a priori do problema de pesquisa, não seriam necessárias para as análises pretendidas, foram marcadas como "Não" na coluna "Manter?" dos Quadros 1 e 2 e excluídas das bases de dados.

Quadro 1: Variáveis da base de dados 1, referente às vagas financiadas.

Item	Variável	Tipo	Base	Correspondente (2)	Manter
1	Sigla da instituição	Qualitativa nominal	1 - Financiadas	Instituição	Sim
2	Estado do local da oferta	Qualitativa nominal	1 - Financiadas	UF	Sim
3	Código IBGE do município	Qualitativa nominal	1 - Financiadas	Código IBGE do município	Sim
4	Município do local da oferta	Qualitativa nominal	1 - Financiadas	Município	Sim
5	Código da unidade de ensino	Qualitativa nominal	1 - Financiadas	Código da unidade de ensino	Não
6	Nome da unidade de ensino	Qualitativa nominal	1 - Financiadas	Nome da unidade de ensino	Sim
7	Tipo de curso	Qualitativa nominal	1 - Financiadas	Tipo de curso	Sim
8	Eixo tecnológico	Qualitativa nominal	1 - Financiadas	Eixo tecnológico	Sim
9	Código do curso	Qualitativa nominal	1 - Financiadas	NA	Não
10	Nome do curso	Qualitativa nominal	1 - Financiadas	Nome do curso	Sim
11	Período de pactuação	Qualitativa nominal	1 - Financiadas	NA	Sim (2)
12	Modalidade educacional	Qualitativa nominal	1 - Financiadas	Modalidade educacional	Sim
13	Carga horária	Quantitativa discreta	1 - Financiadas	Carga horária total	Sim
14	Modalidade de demanda	Qualitativa nominal	1 - Financiadas	NA	Não
15	Vagas	Quantitativa discreta	1 - Financiadas	NA	Sim (2)
16	Situação de turma	Qualitativa ordinal	1 - Financiadas	Situação do ciclo	Sim
17	Data de início	Qualitativa ordinal	1 - Financiadas	Data início do curso	Sim
18	Data prevista de término	Qualitativa ordinal	1 - Financiadas	Data fim previsto do curso	Sim
19	Data de publicação	Qualitativa ordinal	1 - Financiadas	NA	Não
20	Data fim da pré-matrícula	Qualitativa ordinal	1 - Financiadas	NA	Não
21	Matrículas	Quantitativa discreta	1 - Financiadas	Matrículas	Sim
22	TED	Qualitativa nominal	3 - Orçamento	TED	Sim
23	Valor hora-aluno proposto	Quantitativa contínua	3 - Orçamento	Valor hora-aluno proposto	Sim
24	Tipo de vagas	Qualitativa nominal	4 - Criada	Tipo de vagas	Sim

Fonte: Elaborado pelo autor.

Quadro 2: Variáveis da base de dados 2, referente às vagas de contrapartida.

Item	Variável	Tipo	Base	Correspondente (1)	Manter
1	Instituição	Qualitativa nominal	2 - Contrapartida	Sigla da instituição	Sim
2	UF	Qualitativa nominal	2 - Contrapartida	Estado do local da oferta	Sim
3	Código IBGE do município	Qualitativa nominal	2 - Contrapartida	Código IBGE do município	Sim
4	Município	Qualitativa nominal	2 - Contrapartida	Município do local da oferta	Sim
5	Nome da unidade de ensino	Qualitativa nominal	2 - Contrapartida	Nome da unidade de ensino	Sim
6	Código da unidade de ensino	Qualitativa nominal	2 - Contrapartida	Código da unidade de ensino	Não
7	Subtipo cursos	Qualitativa nominal	2 - Contrapartida	NA	Não
8	Eixo tecnológico	Qualitativa nominal	2 - Contrapartida	Eixo tecnológico	Sim
9	Nome do curso	Qualitativa nominal	2 - Contrapartida	Nome do curso	Sim
10	Código ciclo de matrícula	Qualitativa nominal	2 - Contrapartida	NA	Não
11	Ciclo de matrícula	Qualitativa nominal	2 - Contrapartida	NA	Não
12	Carga horária total	Quantitativa discreta	2 - Contrapartida	Carga horária	Sim
13	Data início do curso	Qualitativa ordinal	2 - Contrapartida	Data de início	Sim
14	Data fim previsto do curso	Qualitativa ordinal	2 - Contrapartida	Data prevista de término	Sim
15	Modalidade educacional	Qualitativa nominal	2 - Contrapartida	Modalidade educacional	Sim
16	Situação do ciclo	Qualitativa nominal	2 - Contrapartida	Situação de turma	Sim
17	Tipo de curso	Qualitativa nominal	2 - Contrapartida	Tipo de curso	Sim
18	Matrículas	Quantitativa discreta	2 - Contrapartida	Matrículas	Sim
19	TED	Qualitativa nominal	3 - Orçamento	TED	Sim
20	Valor hora-aluno proposto	Quantitativa contínua	3 - Orçamento	Valor hora-aluno proposto	Sim
21	Tipo de vagas	Qualitativa nominal	4 - Criada	Tipo de vagas	Sim

Fonte: Elaborado pelo autor.

Finalizada essa etapa inicial de análise das bases de dados, as bases 1 e 2 foram consolidadas, sendo a variável "Tipo de vagas" indicando a origem de cada turma, se "financiadas" constam na base de dados 1, se "contrapartida" constam na base de dados 2. No Quadro 4 são indicadas as variáveis que compõem a base de dados consolidada, a

Quadro 3: Variáveis da base de dados 3, referente ao controle orçamentário dos TED.

Item	Variável	Tipo	Base	Correspondente (1 e 2)	Manter
1	Instituição	Qualitativa nominal	3 - Orçamento	Sigla da instituição / Instituição	Sim
2	Número TED	Qualitativa nominal	3 - Orçamento	TED (incluída nas bases 1 e 2)	Sim
3	Ano	Qualitativa ordinal	3 - Orçamento	NA	Não
4	Processo	Qualitativa nominal	3 - Orçamento	NA	Não
5	SIAFI	Qualitativa nominal	3 - Orçamento	NA	Sim
6	Valor Aprovado (R$)	Quantitativa contínua	3 - Orçamento	NA	Sim
7	Valor Empenhado (R$)	Quantitativa contínua	3 - Orçamento	NA	Sim
8	Valor da Hora-Aluno (R$)	Quantitativa contínua	3 - Orçamento	NA	Sim

Fonte: Elaborado pelo autor.

ser utilizada para as análises.

Todas as etapas de consolidação dos dados foram realizadas via programação no software livre R-Studio, versão 1.1.456, sendo os respectivos códigos disponibilizados no Apêndice.

Quadro 4: Variáveis da base de dados consolidada.

Item	Variável	Descrição	Tipo	Base
1	x1_inst	Nome da instituição de ensino ofertante	Qualitativa nominal	Vagas (1, 2)
2	x2_uf	Unidade da federação da instituição ofertante.	Qualitativa nominal	Vagas (1, 2)
3	x3_cod_ibge_mun	Código do IBGE do município de oferta do curso	Qualitativa nominal	Vagas (1, 2)
4	x4_mun	Município de oferta do curso	Qualitativa nominal	Vagas (1, 2)
5	x5_nome_ue	Nome da unidade de ensino ofertante	Qualitativa nominal	Vagas (1, 2)
6	x6_tp_curso	Tipo de curso (FIC ou técnico)	Qualitativa nominal	Vagas (1, 2)
7	x7_eixo_tec	Eixo tecnológico do curso ofertado	Qualitativa nominal	Vagas (1, 2)
8	x8_nome_curso	Nome do curso	Qualitativa nominal	Vagas (1, 2)
9	x9_per_pact	Período de pactuação da oferta	Qualitativa nominal	Vagas (1)
10	x10_mod_of	Modalidade da oferta (presencial ou EaD)	Qualitativa nominal	Vagas (1, 2)
11	x11_ch	Carga horária do curso	Quantitativa discreta	Vagas (1, 2)
12	x12_vagas	Quantidade de vagas aprovadas no curso	Quantitativa discreta	Vagas (1)
13	x13_st_turma	Situação da turma	Qualitativa ordinal	Vagas (1, 2)
14	x14_dt_inicio	Data de início do curso	Qualitativa ordinal	Vagas (1, 2)
15	x15_dt_termino	Data prevista de término do curso	Qualitativa ordinal	Vagas (1, 2)
16	x16_mat	Quantidade de matrículas no curso	Quantitativa discreta	Vagas (1, 2)
17	x17_ted	Número do TED relacionado à turma	Qualitativa nominal	Orçamento (3)
18	x18_vl_ha_prop	Valor da hora-aluno proposto pela instituição	Quantitativa contínua	Orçamento (3)
19	x19_tp_vagas	Tipo de vagas (financiada ou contrapartida)	Qualitativa nominal	Criada (4)
20	x20_ha_prop	Quantidade de horas-aluno proposta	Quantitativa discreta	Criada (4)
21	x21_vl_prop	Valor total proposto por turma	Quantitativa contínua	Criada (4)
22	x22_ha_exec	Quantidade de horas-aluno executadas	Quantitativa discreta	Criada (4)
23	x23_siafi	Número SIAFI correspondente ao TED	Qualitativa nominal	Orçamento (3)
24	x24_vl_prop_ha_efet	Valor proposto efetivo da hora-aluno	Quantitativa contínua	Criada (4)
25	x25_vl_exec_ha_efet	Valor executado efetivo da hora-aluno	Quantitativa contínua	Criada (4)

Fonte: Elaborado pelo autor.

As variáveis "x9_per_pact", que indica o período de pactuação das vagas aprovadas, e "x12_vagas", que indica a quantidade de vagas aprovadas para cada curso, estão disponíveis somente na base de dados 1, das vagas financiadas, portanto, foram inseridas nas turmas da base de dados 2, das vagas de contrapartida, em analogia às vagas financiadas, dado que toda vaga de contrapartida está relacionada, necessariamente, a uma vaga financiada. Por isso, no Quadro 4 essas variáveis foram indicadas como "Vagas (1)" na

coluna "Base", ou seja, estão disponíveis somente na base de dados 1.

As variáveis "x17_ted", que indica o número do TED que financiou aquela turma, "x18_vl_ha_prop", que indica o valor da hora-aluno, em reais, proposto pela instituição de ensino na proposta de oferta, e "x23_siafi", que relaciona o código da transferência SIAFI do TED respectivo, constam na base de dados 3, referente ao controle da execução orçamentária dos TED, sendo todas marcadas como "Orçamento (3)" na coluna "Base" do Quadro 4. Tal controle é realizado pela área técnica da SETEC/MEC. Aqui é importante ressaltar que a variável "x17_ted" foi utilizada como vinculação entre as três bases de dados.

Já as variáveis "x19_tp_vagas", "x20_ha_prop", "x21_vl_prop", "x22_ha_exec", "x24_vl_prop_ha_efet" e "x25_vl_exec_ha_efet" foram criadas e incluídas na base de dados consolidada, conforme explicações a seguir.

- **x19_tp_vagas:** indica se as vagas daquela turma são "financiadas" ou "contrapartida", caso sejam do primeiro grupo, constam originalmente na base de dados 1, caso sejam do segundo grupo, constam originalmente na base de dados 2.

- **x20_ha_prop:** indica a quantidade de horas-aluno total por turma, foi calculada pela multiplicação entre a quantidade de vagas aprovadas (x12_vagas) e a carga horária do curso (x11_ch), assim, para cada turma, foi identificado o máximo de horas-aluno que podem ser executadas.

- **x21_vl_prop:** indica o valor total proposto para a execução de cada turma, que foi calculado pela multiplicação entre a quantidade total de horas-aluno por turma (x20_ha_prop) e o valor da hora-aluno proposto pela instituição (x18_vl_ha_prop).

- **x22_ha_exec:** indica a quantidade de horas-aluno executada em cada turma, foi obtida pela multiplicação entre a quantidade de matrículas da turma (x16_mat) e a carga horária do curso (x11_ch).

- **x24_vl_prop_ha_efet:** considerando que o valor da hora-aluno proposto pela instituição ofertante (x18_vl_ha_prop) corresponde somente às vagas financiadas, é necessário calcular o valor efetivo proposto para a hora-aluno para cada TED, que deverá considerar, além das vagas financiadas, as vagas de contrapartida; assim, foram calculados os somatórios das horas-alunos das vagas financiadas e das vagas de contrapartida para cada TED, esse total foi dividido pelo valor total aprovado (x21_vl_prop), resultando no valor efetivo proposto por cada instituição ofertante para o total de horas-aluno a serem executadas, tanto nas vagas financiadas quanto nas vagas de contrapartida.

- **x25_vl_exec_ha_efet:** para o cálculo dessa variável, primeiro foi calculado o total de horas-aluno executadas para cada TED, por meio da multiplicação entre as matrículas registradas (x16_mat) e as cargas horárias dos cursos (x11_ch); os resultados foram agrupados por TED (x17_ted), assim, considerando o valor empenhado para cada TED, constante na base de dados 3, de controle orçamentário, foi realizada a razão entre o valor empenhado e o total de horas-aluno executadas para cada TED, assim, chegou-se ao valor da hora-aluno efetivamente executado em cada TED.

Portanto, na base de dados consolidada, para cada turma, de cada curso ofertado, há a variável "x25_vl_exec_ha_efet", que indica o valor efetivo da hora-aluno executada pelo parceiro ofertante, valor que pode ser comparado com o valor proposto inicial (x18_vl_ha_prop), que considera somente as vagas financiadas, e o valor proposto efetivo (x24_vl_prop_ha_efet), que considera as vagas financiadas e as vagas de contrapartida.

Por fim, no que tange à dimensão da base de dados consolidada, esta conta com 2.613 observações, cada qual correspondendo a uma turma, e 25 variáveis, conforme especificado no Quadro 4. Essas 2.613 turmas totalizam 192.717 vagas aprovadas em 2020.

5 Análise e Resultados

Após a consolidação da base de dados, contendo 2.613 observações, cada uma representando uma turma, como unidade amostral, e 25 variáveis associadas, foram definidas as variáveis de interesse do estudo:

- **Instituição:** qualitativa nominal, representada por "x1_inst", indica a sigla da instituição da Rede Federal ofertante da turma; são 25 instituições de ensino que compõem a base de dados, sendo 21 Institutos Federais e 4 Universidades Federais.
- **UF:** qualitativa nominal, representada por "x2_UF", indica a unidade da federação onde a instituição de ensino ofertante da turma está localizada; as instituições estão localizadas em 20, das 27 unidades da federação; os Estados do Ceará, Espírito Santo, Mato Grosso, Paraná, Rio de Janeiro, São Paulo e Sergipe não ofertaram turmas do Pronatec no ano de 2020.
- **Eixo Tecnológico:** qualitativa nominal, representado por "x7_eixo_tec", indica o eixo tecnológico ao qual o curso pertence, com referência na 3ª edição do Catálogo Nacional de Cursos Técnicos - CNCT, o qual categoriza os cursos em 13 eixos tecnológicos; para os dados analisados, 11 eixos são representados, somente os eixos Militar e Segurança não tiveram cursos ofertados em 2020.

- **Nome do Curso:** qualitativa nominal, representado por "x28_nome_curso", indica o nome do curso ofertado, tendo como referência a 4ª edição do Guia Pronatec de Cursos FIC; nos dados analisados, há 176 cursos, dos 646 cursos existentes no Guia.
- **Valor Proposto Efetivo da Hora-Aula**: quantitativa contínua, dada em reais (R$) e indicada por "x24_vl_prop_ha_efet", representa o valor efetivo da hora-aula proposto pela instituição de insino na proposta de oferta, foi calculada, pois nas bases de dados originais, somente as vagas financiadas continham valor da hora-aluno, sendo necessário agrupar as vagas financiadas e as vagas de contrapartida ofertadas.
- **Valor Executado Efetivo da Hora-Aula:** quantitativa contínua, dada em reais (R$) e indicada por "x25_vl_exec_ha_efet", representa o valor executado efetivo recebido pela instituição ofertante para cada hora-aula de um determinado curso, é diferente do valor proposto pois, ao fim da oferta, as instituições informaram necessitar de um valo menor do que o proposto para custear todas as despesas do curso.

5.1 Análise Descritiva dos Dados

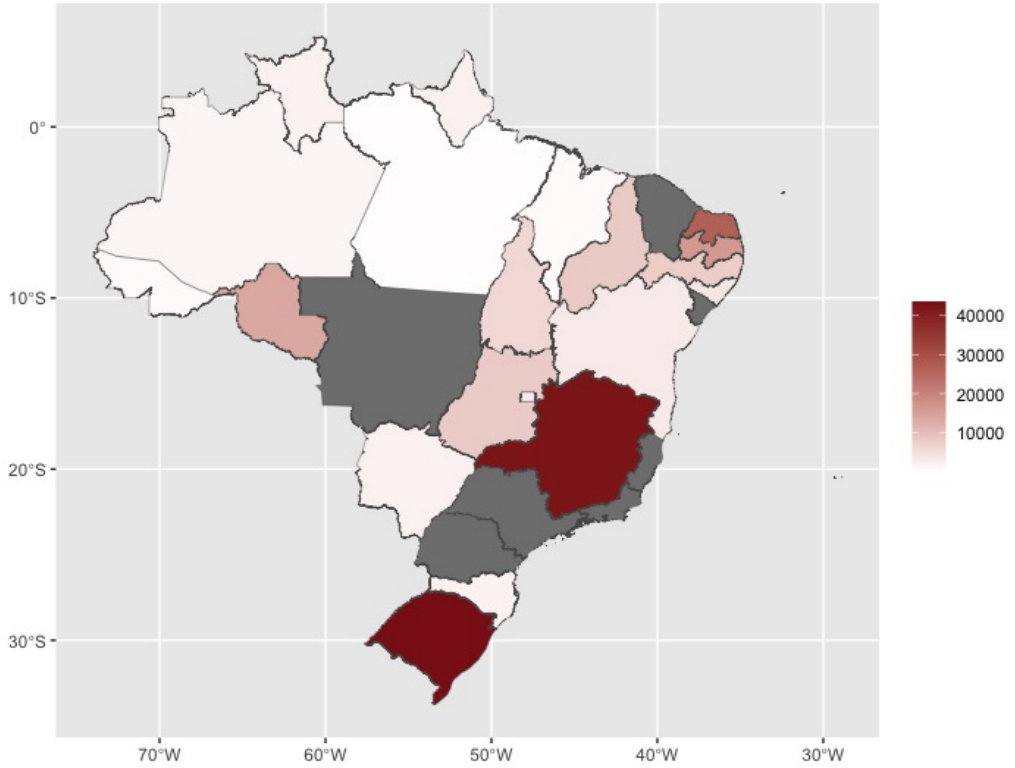
Ao iniciar a análise descritiva dos dados pelas duas variáveis quantitativas, "valor proposto efetivo da hora-aula"e "valor executado efetivo da hora-aula", verificou-se que, para o valor executado, havia 14 observações com o valor nulo, ou seja, turmas para as quais não houve pagamento, todas ofertadas pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina - IFSC, objeto de um mesmo TED, para o qual não houve repasse de recursos em 2020. Portanto, essas observações foram excluídas da base de dados, a qual passou a contar com 2.599 observações para as 6 variáveis elencadas como de interesse do estudo.

Para iniciar a análise dos dados, é importante entender a distribuição das vagas aprovadas pelo território brasileiro, considerando que cada uma das 27 unidades da federação possuem uma, ou mais de uma, instituição da Rede Federal apta a atuar como ofertante dos cursos do Pronatec.

As instituições de ensino sediadas nos Estados do Ceará, Sergipe, Mato Grosso, Espírito Santo, Rio de Janeiro, São Paulo e Paraná não apresentaram proposta para a oferta de vagas do Pronatec no ano de 2020, por isso são indicadas na cor cinza no gráfico acima. É importante ressaltar que há discricionaridade das instituições de ensino para apresentarem proposta de oferta de cursos. Assim, as instituições de ensino sediadas nas UF citadas anteriormente, e marcadas em cinza no gráfico 3, não apresentaram proposta de pactuação nem apresentaram justificativa para a não apresentação de proposta. Ademais, verifica-se que as cinco regiões brasileiras estão representadas na oferta de cursos, e

os Estados de Minas Gerais e Rio Grande do Sul apresentaram os maiores quantitativos de vagas aprovadas.

Gráfico 3: Total de vagas aprovadas por Unidade da Federação (UF), considerando as instituições da Rede Federal sediadas em cada Estado e no Distrito Federal.



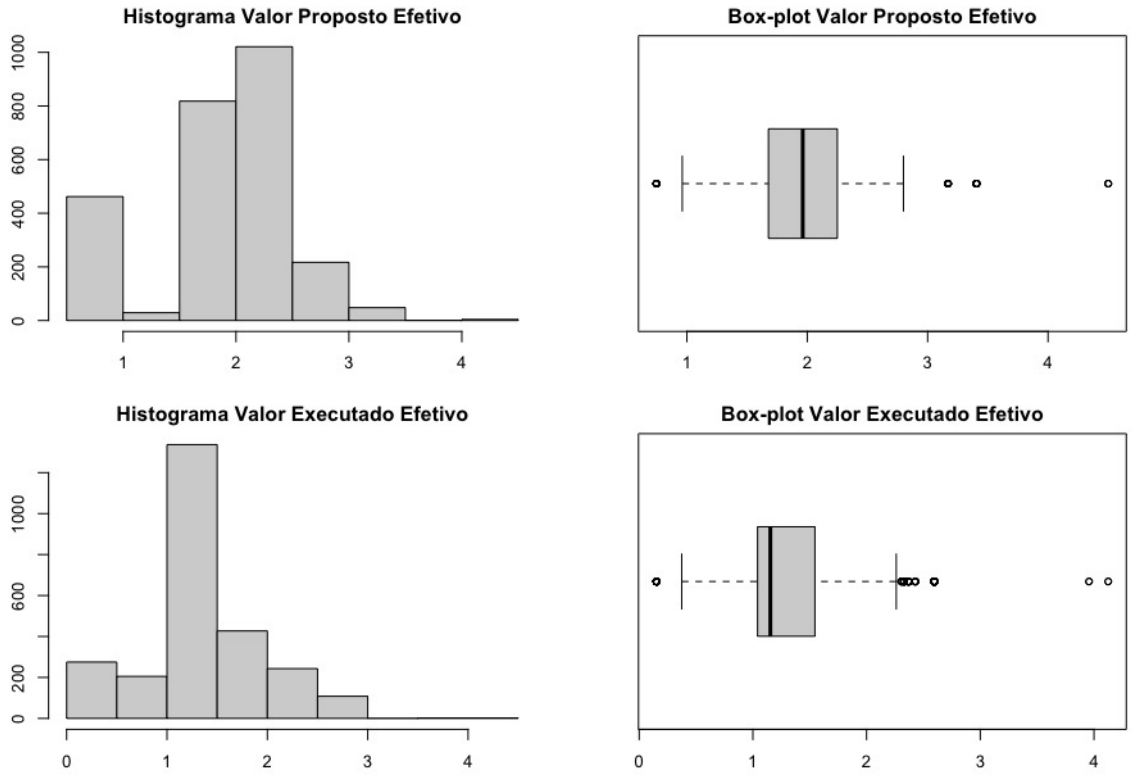
Fonte: Elaborado pelo autor.
Em cinza, os Estados cujas instituições da Rede Federal não tiveram vagas pactuadas em 2020.

Abaixo, são apresentados o histograma e o box-plot para os valores propostos e executados efetivos das horas-aluno dos cursos. É possível verificar que os dados não seguem uma distribuição normal.

5.1.1 Análise Descritiva por Eixo Tecnológico

Considerando que uma das hipóteses levantadas para o estudo indica que pode haver diferenças entre os valores das horas-alunos para cursos pertencentes a diferentes eixos tecnológicos, procedeu-se com a análise descritiva das variáveis quantitativas categorizadas por eixos tecnológicos.

Figura 1: Análise descritiva do valor proposto e do valor executado efetivo.



Fonte: Elaborado pelo autor.

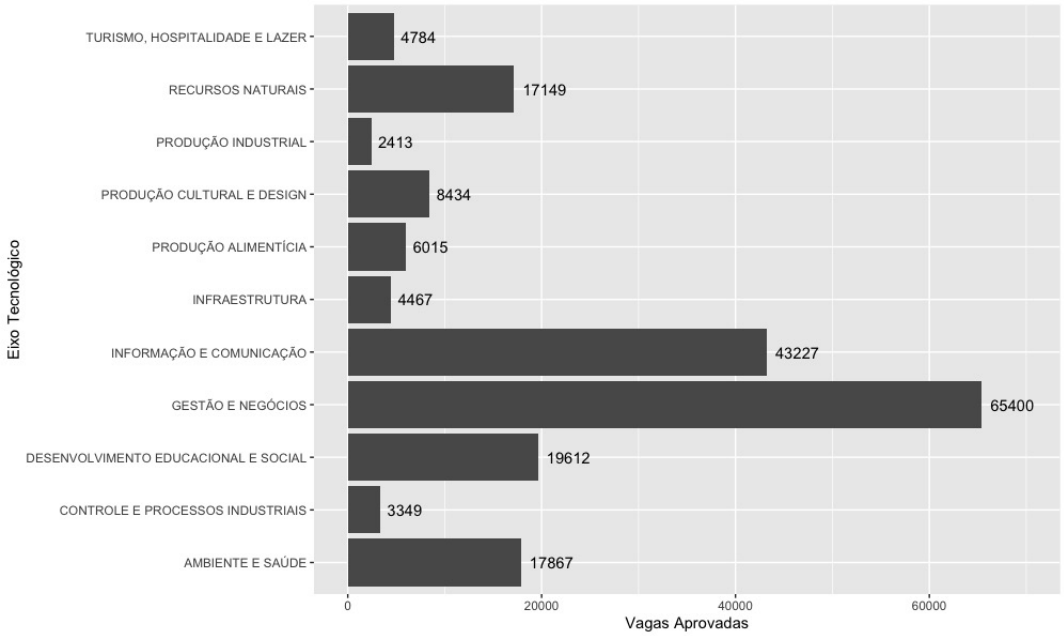
Tabela 2: Medidas descritivas do valor proposto efetivo da hora-aula (x_24).

	Eixo Tecnológico	n	med	var	dp	cv	min	max	q1	q3	li	ls
1	AS - Ambiente e Saúde	295	1.90	0.213	0.462	0.243	0.75	4.50	1.68	2.25	0.82	3.11
2	CPI - Controle e Processos Industriais	60	1.97	0.285	0.534	0.271	0.96	2.27	2.23	2.25	2.20	2.28
3	DES - Desenvolvimento Educacional e Social	231	1.99	0.218	0.467	0.235	0.75	3.17	1.75	2.25	1.00	3.00
4	GN - Gestão e Negócios	915	1.91	0.300	0.548	0.287	0.75	3.41	1.68	2.25	0.82	3.11
5	IC - Informação e Comunicação	495	1.86	0.367	0.606	0.326	0.96	4.50	1.53	2.25	0.44	3.33
6	INF - Infraestrutura	50	2.36	0.136	0.369	0.156	1.65	3.41	2.25	2.38	2.06	2.57
7	PA - Produção Alimentícia	77	1.96	0.341	0.584	0.298	0.75	2.62	1.61	2.37	0.48	3.51
8	PCD - Produção Cultural e Design	135	1.64	0.593	0.770	0.469	0.75	4.50	0.96	2.25	0.00	4.18
9	PI - Produção Industrial	41	1.60	0.441	0.664	0.415	0.96	2.62	0.96	2.25	0.00	4.18
10	RN - Recursos Naturais	206	2.05	0.276	0.525	0.256	0.96	3.17	1.61	2.28	0.62	3.27
11	THL - Turismo, Hospitalidade e Lazer	94	2.06	0.176	0.419	0.203	0.75	3.17	1.68	2.25	0.82	3.11
	Geral	**2.599**	**1.91**	**0.316**	**0.562**	**0.294**	**0.75**	**4.50**	**1.68**	**2.25**	**0.82**	**3.11**

Fonte: Elaborado pelo autor.

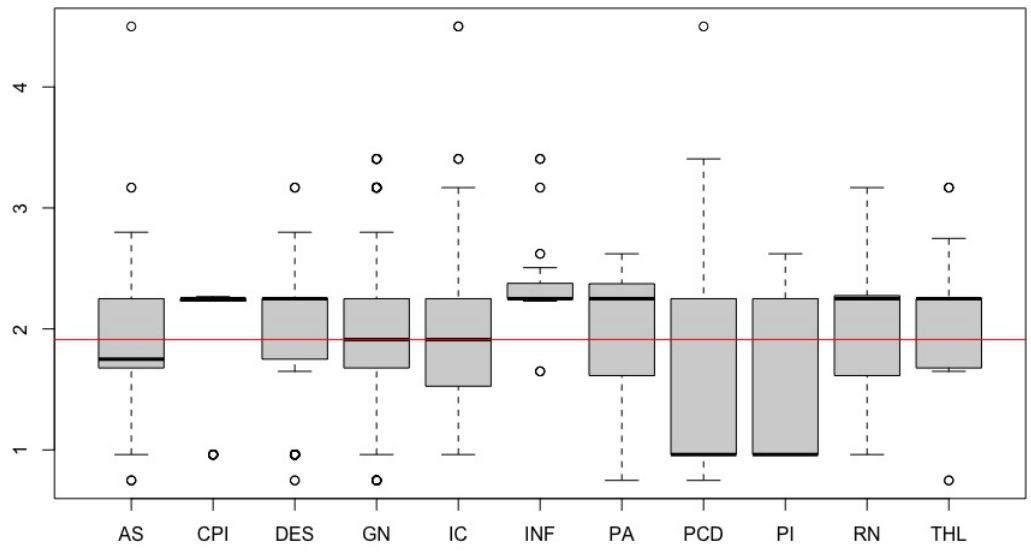
No gráfico 4, verifica-se que o eixo tecnológico Gestão e Negócios - GN apresentou o maior número de vagas aprovadas, seguido do eixo Informação e Comunicação - IC. No outro extremo, com os menores quantitativos de vagas aprovadas, estão os eixos de Produção Industrial - PI e de Controle e Processos Industriais - CPI.

Gráfico 4: Vagas aprovadas por eixo tecnológico.



Fonte: Elaborado pelo autor.

Gráfico 5: Box-plot do valor efetivo proposto da hora-aula por eixo tecnológico.



Fonte: Elaborado pelo autor.

O gráfico 5 evidencia que as médias dos valores propostos efetivos podem ser diferentes, de acordo com cada eixo tecnológico, principalmente quando as médias dos grupos são comparadas com a média global dos dados, indicada em vermelho. Porém, os

valores propostos pelas instituições de ensino não serão considerados para o modelo de estimação dos valores das horas-aluno, pois a definição desses valores propostos não segue uma metodologia única, sendo de competência de cada instituição.

Tabela 3: Medidas descritivas do valor executado efetivo da hora-aula (x_25).

	Eixo Tecnológico	n	med	var	dp	cv	min	max	q1	q3	li	ls
1	AS - Ambiente e Saúde	295	1.30	0.261	0.511	0.393	0.15	4.13	1.03	1.55	0.25	2.32
2	CPI - Controle e Processos Industriais	60	0.84	0.308	0.555	0.661	0.15	2.00	0.37	1.13	0.00	2.25
3	DES - Desenvolvimento Educacional e Social	231	1.27	0.280	0.529	0.416	0.15	2.43	1.04	1.56	0.26	2.35
4	GN - Gestão e Negócios	915	1.41	0.380	0.616	0.437	0.15	2.59	1.04	1.71	0.03	2.73
5	IC - Informação e Comunicação	495	1.28	0.272	0.522	0.408	0.15	3.96	1.04	1.55	0.28	2.30
6	INF - Infraestrutura	50	1.58	0.442	0.665	0.421	0.15	2.31	1.13	2.16	0.00	3.72
7	PA - Produção Alimentícia	77	1.21	0.363	0.602	0.498	0.15	2.37	0.95	1.56	0.03	2.48
8	PCD - Produção Cultural e Design	135	1.11	0.260	0.510	0.459	0.15	4.13	1.04	1.13	0.92	1.25
9	PI - Produção Industrial	41	0.90	0.229	0.478	0.531	0.15	2.20	0.46	1.04	0.00	1.91
10	RN - Recursos Naturais	206	1.22	0.399	0.632	0.518	0.15	2.37	0.69	1.71	0.00	3.26
11	THL - Turismo, Hospitalidade e Lazer	94	1.22	0.236	0.486	0.398	0.37	2.20	1.13	1.41	0.71	1.83
	Geral	**2.599**	**1.30**	**0.338**	**0.581**	**0.447**	**0.15**	**4.13**	**1.04**	**1.55**	**0.28**	**2.30**

Fonte: Elaborado pelo autor.

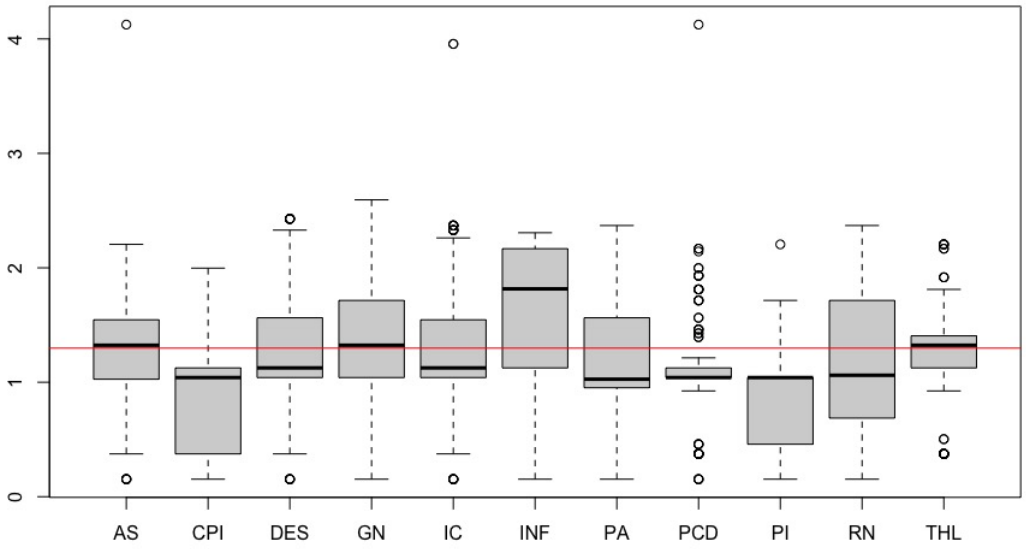
Assim como no gráfico 5, o gráfico 6 também indica que há evidências para supor que cada eixo tecnológico possui um valor médio executado por hora-aluno diferente dos demais, quando se analisa a variação das médias dos grupos (eixos) em relação à média global, marcada no gráfico pela linha vermelha. Alguns eixos tecnológicos poderão apresentar valores semelhantes a outros, porém, a hipótese adotada é de que o eixo tecnológico influencia no valor da hora-aluno executado, principalmente devido às diferentes necessidades de infraestrutura física e de pessoal para viabilizar a oferta dos cursos. Assim, além da evidência gráfica, a hipótese de diferença se baseia nas diferentes cargas horárias, infraestruturas e equipe destinadas aos cursos de cada eixo tecnológico.

5.1.2 Análise Descritiva por Região

Por mais que cada unidade da federação tenha uma ou mais instituições de ensino, e que a atuação dessas instituições ocorre, quase que integralmente, nos limites territoriais da unidade da federação, levanta-se a hipótese de que a região geográfica onde está sediada a instituição de ensino ofertante possa influenciar no valor da hora-aluno. Tal suposição se baseia no fato de que as regiões brasileiras possuem características econômicas, sociais e culturais diversas, o que pode afetar os custos da oferta.

De todas as vagas aprovadas para cursos FIC do Pronatec em 2020, ofertados pela Rede Federal, 34,2% foram para instituições de ensino sediadas na região nordeste, seguida da região sul, com 23,6%, da região sudeste, com 22,0%, da região norte, com 13,6%, e da região centro-oeste, com 6,6%. A região nordeste, mesmo com mais unidades da federação, apresentou um número maior de instituições da Rede Federal interessadas

Gráfico 6: Box-plot do valor efetivo executado da hora-aula por eixo tecnológico.



Fonte: Elaborado pelo autor.

em ofertar vagas em cursos FIC do Pronatec, enquanto na região sudeste, somente o Estado de Minas Gerais ofertou vagas, por meio de 4 (quatro) instituições de ensino.

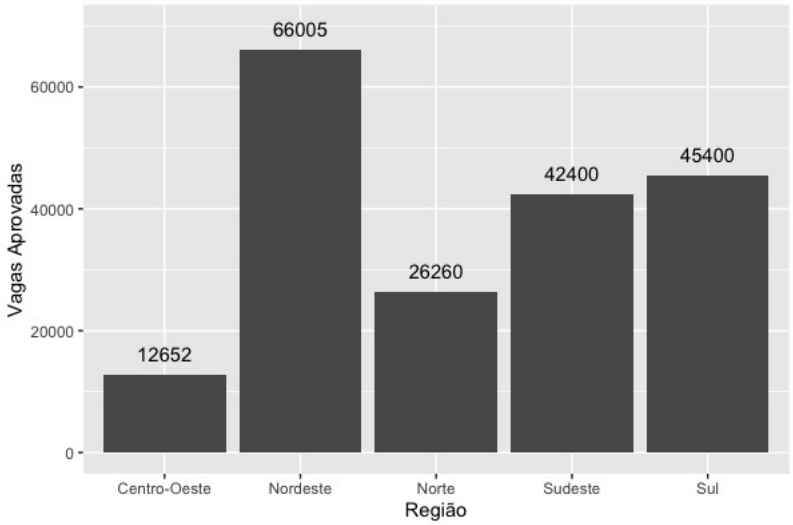
De acordo com o gráfico 8, assume-se a hipótese de que as regiões brasileiras apresentam valores diferentes para a hora-aluno executada, ao comparar as médias dos grupos com a média global, indicada em vermelho. No que tange as regiões centro-oeste e sul, a variabilidade da primeira é quase nula, mesmo ocorrendo oferta em três, das quatro unidades da federação, indicando uma uniformidade dos valores, enquanto que na segunda, região sul, a grande quantidade de vagas ofertadas pelo Rio Grande do Sul influenciou na média, pois Santa Catarina ofertou somente duas turmas, enquanto o Paraná não ofertou vagas na pactuação analisada.

Tabela 4: Medidas descritivas do valor executado efetivo da hora-aula (x_25) por região.

	Região	n	med	var	dp	cv	min	max	q1	q3	li	ls
1	Centro-Oeste	232	2.01	0.155	0.394	0.196	1.15	2.31	2.14	2.20	2.05	2.30
2	Nordeste	745	1.06	0.250	0.500	0.472	0.37	4.13	0.86	1.13	0.46	1.52
3	Norte	354	1.89	0.303	0.550	0.291	1.02	2.59	1.43	2.59	0.00	4.34
4	Sudeste	786	1.24	0.255	0.505	0.407	0.15	2.33	0.92	1.55	0.00	2.48
5	Sul	482	0.99	0.027	0.164	0.166	0.50	1.04	1.04	1.04	1.04	1.04
	Geral	**2.599**	**1.30**	**0.338**	**0.581**	**0.447**	**0.15**	**4.13**	**1.04**	**1.55**	**0.28**	**2.30**

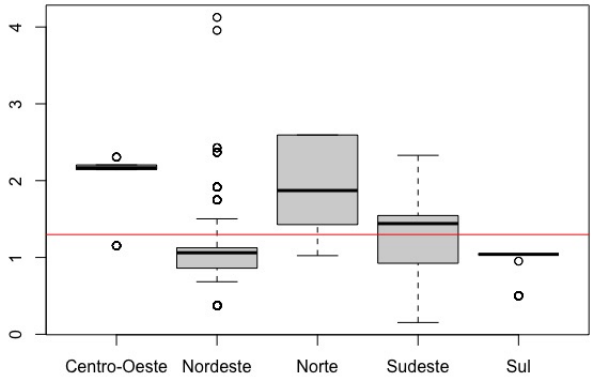
Fonte: Elaborado pelo autor.

Gráfico 7: Vagas aprovadas por região.



Fonte: Elaborado pelo autor.

Gráfico 8: Box-plot do valor efetivo executado da hora-aula por região.



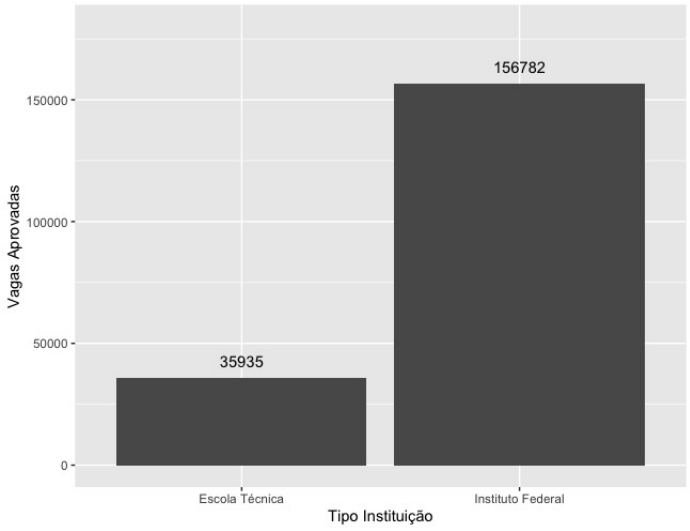
Fonte: Elaborado pelo autor.

5.1.3 Análise Descritiva por Tipo de Instituição

Considerando a diferença administrativa e organizacional entre os Institutos Federais - IF e as Escolas Técnicas Vinculadas às Universidades Federais - ETV, leva-se a pressupor que podem haver diferenças entre os custos desses tipos de instituições, o que pode refletir no valor da hora-aula. Portanto, serão analisadas essas eventuais diferenças.

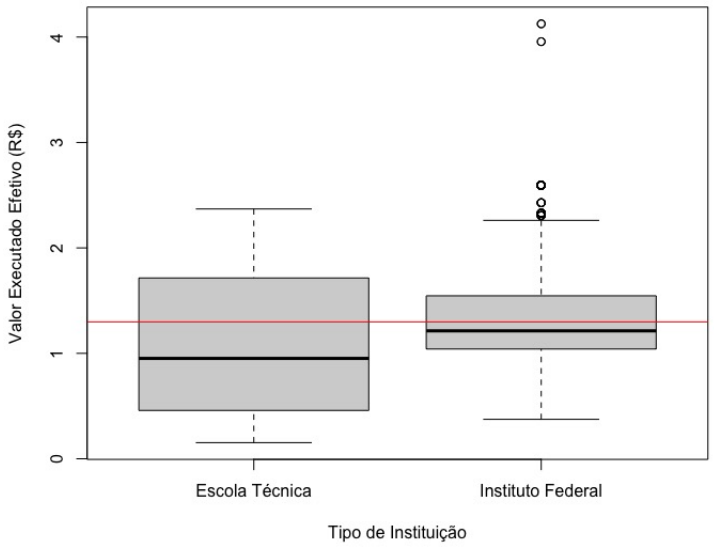
A terceira hipótese a ser considerada é a que indica diferenças entre os valores das horas-aluno executadas por tipo de instituição.

Gráfico 9: Vagas aprovadas por tipo de instituição.



Fonte: Elaborado pelo autor.

Gráfico 10: Box-plot do valor efetivo executado da hora-aula por tipo de instituição.



Fonte: Elaborado pelo autor.

Tabela 5: Medidas descritivas do valor executado efetivo da hora-aula (x_25) por tipo de instituição.

	Tipo de Instituição	**n**	**med**	**var**	**dp**	**cv**	**min**	**max**	**q1**	**q3**	**li**	**ls**
1	Escola Técnica	396	0.94	0.391	0.625	0.665	0.15	2.37	0.46	1.71	0.00	3.60
2	Instituto Federal	2.203	1.36	0.302	0.549	0.404	0.37	4.13	1.04	1.55	0.28	2.30
	Geral	**2.599**	**1.30**	**0.338**	**0.581**	**0.447**	**0.15**	**4.13**	**1.04**	**1.55**	**0.28**	**2.30**

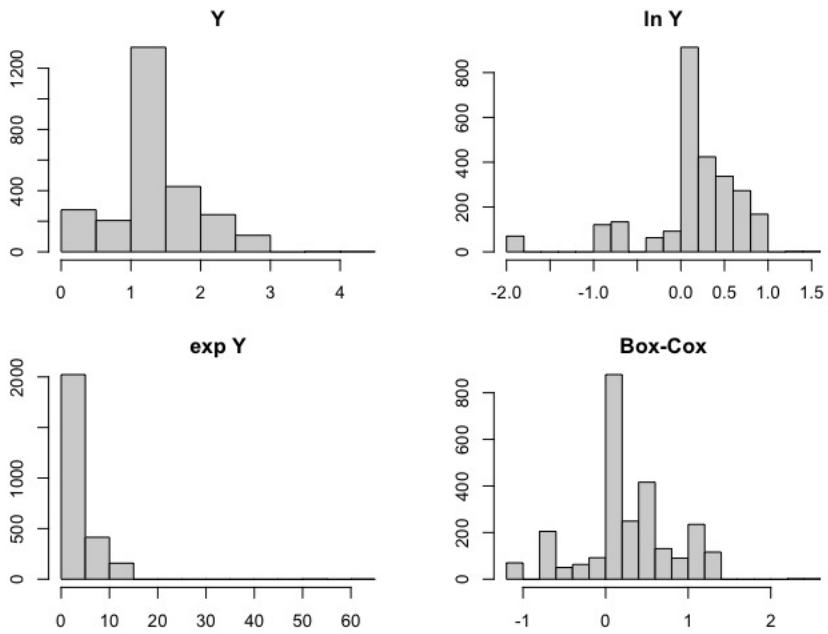
Fonte: Elaborado pelo autor.

5.1.4 Transformação da Variável Resposta

Conforme indicado na seção 4.1.1, a qual trata das suposições do modelo ANOVA, quando os dados da variável resposta (Y_{ijk}), quantitativa, que representa o valor executado efetivo da hora-aluno, dado um eixo tecnológico i, uma região geográfica j e um tipo de instituição k, indicam uma não normalidade, uma das alternativas é tentar transformar esses dados.

Para o caso desse estudo, especificamente, a Figura 1 indica, na parte inferior, que as observações dos valores executados efetivos (variável resposta) seguem uma distribuição com tendência assimétrica. Assim, foram realizadas três transformações, cujos histogramas, juntamente com o dos dados originais, constam na Figura 2.

Figura 2: Histogramas dos dados originais e das três tranformações realizadas.

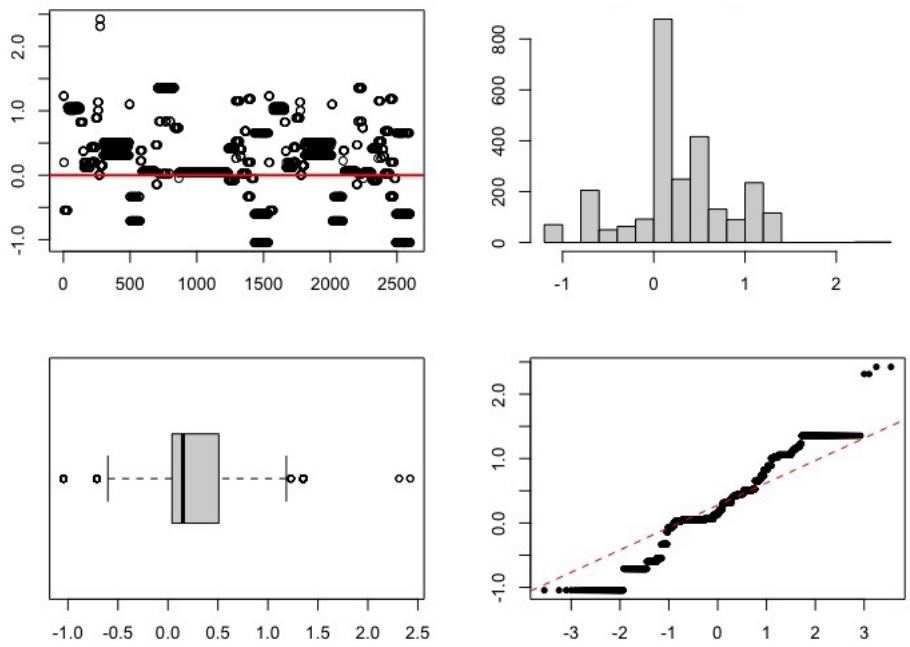


Fonte: Elaborado pelo autor.

No que tange à transformação de Box-Cox, dada pela equação 4.1.2, e a mais indicada para os casos de não normalidade dos dados, na Figura 3, é apresentada a análise gráfica dos dados transformados.

Tanto a análise gráfica quanto o teste de normalidade de Shapiro-Wilk indicam a rejeição da hipótese nula de normalidade dos dados. Porém, em casos de tamanhos de

Figura 3: Análise gráfica da normalidade da variável resposta após a transformação de Box-Cox.



Fonte: Elaborado pelo autor.

amostras e de populações razoáveis e não muito assimétricos, como é o caso dos dados analisados, a ANOVA pode ser utilizada, tendo como base o Teorema do Limite Central - TLC. Portanto, para a elaboração do modelo ANOVA para a estimação do valor executado da hora-aluno, serão utilizados os dados da variável resposta não transformados, ou seja, conforme foram observados.

5.1.5 Valores Discrepantes da Variável Resposta

Conforme análise descritiva das observações da variável resposta, valor executado efetivo da hora-aluno, verificou-se a presença de valores discrepantes, também denominados como *outliers*. Das 2599 observações, foram identificadas 70 observações abaixo do limite inferior e 162 observações acima do limite superior, sendo esses limites dados por uma vez e meia o desvio interquartílico em relação à mediana. É importante ressaltar que todas essas 232 observações discrepantes estão, também, representadas nas 2367 observações restantes, em termos de combinações entre eixos tecnológicos, regiões geográficas e tipos de instituição. Portanto, serão elaborados duas estimativas dos parâmetros do modelo ANOVA, a primeira considerando os dados discrepantes e a segunda desconsiderando essas observações, para que seja observado o modelo que será melhor ajustado.

5.2 Análise de Variância 1: com Dados Discrepantes

A Análise de Variância - ANOVA considerará a dependência do valor executado efetivo da hora-aluno em relação ao eixo tecnológico do curso, à região geográfica onde a instituição de ensino está localizada e o itpo de instituição de ensino. Portanto, o modelo é descrito como (descrito na seção 4.1):

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \alpha_i\beta_j\gamma_k + \epsilon_{ijk}$$

Para o caso analisado, μ, ou intercepto, corresponde à média global de todos os valores de Y_{ijk}, ou seja, a média geral do valor executado efetivo da hora-aluno. Os parâmetros α_i representam as categorias dos eixos tecnológicos dos cursos ($i = 1, 2, ..., 11$). Analogamente, os parâmetros β_j indicam as categorias das regiões geográficas ($j = 1, 2, ..., 5$). Já os parâmetros γ_k, consistem nas categorias do tipo de instituição ($k = 1, 2$).

Os parâmetros $\alpha_i\beta_j$, $\alpha_i\gamma_k$, $\beta_j\gamma_k$ e $\alpha_i\beta_j\gamma_k$ referem-se às interações entre os fatores AB, AC, BC e ABC, respectivamente, para cada valor de i, j, k. Por fim, o erro é representado pelo termo ϵ_{ijk}, consistindo na diferença entre o valor observado e o valor previsto.

Para realizar a estimativa pontual do valor da hora-aluno ($\hat{Y}_{ijk}$), ou seja, o valor esperado de Y_{ijk}, e considerando o pressuposto de normalidade dos erros/resíduos ($E[\epsilon_{ijk}] = 0$), utiliza-se:

$$E[Y_{ijk}] = \hat{Y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \alpha_i\beta_j\gamma_k$$

Assim, ao serem estimados os parâmetros do modelo, poderão ser previstos os valores esperados para a hora-aluno, dado o eixo tecnológico, a região geográfica e o tipo de instituição. Com a utilização do software livre $R - Studio$, foi realizada uma ANOVA com 80% das 2.599 observações, selecionadas aleatoriamente, como base de dados de treinamento, ou seja, para gerar as estimativas dos parâmetros do modelo. Os outros 20% das observações foram separadas em uma base de dados denominada de dados de teste, e serão utilizadas para testar as estimativas do modelo.

Abaixo, são apresentados os resultados da ANOVA, em forma de tabela.

Os resultados indicam a rejeição da hipótese nula de igualdade de médias/efeitos para os fatores A, B e C, assim como para suas respectivas interações, isto é, o eixo tecnológico, a região geográfica e o tipo de instituição onde o curso é ofertado são significativos para a estimativa do valor da hora-aluno para a oferta de um determinado curso

Tabela 6: ANOVA com dados discrepantes.

Fonte de Variação	**Graus Liberdade (GL)**	**Soma de Quadrados (SQ)**	**Quadrado Médio (QM)**	**Estatística F**	**p-valor**
Fator A (Eixo Tecnológico)	10	31.7	3.17	19.552	$< 2e^{-16}$ ***
Fator B (Região)	4	238.1	59.53	367.100	$< 2e^{-16}$ ***
Fator C (Tipo Instituição)	1	18.6	18.58	114.565	$< 2e^{-16}$ ***
Interação AB	36	44.0	1.22	7.538	$< 2e^{-16}$ ***
Interação AC	9	9.8	1.08	6.681	$< 1.79e^{-9}$ ***
Interação BC	1	24.8	24.82	153.074	$< 2e^{-16}$ ***
Interação ABC	7	4.2	0.60	3.673	0.000599 ***
Resíduos	2010	326	0.16	–	–
Total	2078	697.2	–	–	–

Fonte: Elaborado pelo autor.

de formação inicial e continuada - FIC, na modalidade de Educação a Distância - EaD, pela Rede Federal de Educação Profissional, Científica e Tecnológica - RFEPCT.

O coeficiente de determinação (R^2) estima a proporção da variabilidade da variável dependente (valor da hora-aluno) que é explicada pelo conjunto dos três fatores do modelo (eixo tecnológico, região geográfica e tipo de instituição), sendo calculado por:

$$R^2 = 1 - \frac{SQRes}{SQT} \tag{5.2.1}$$

Considerando os resultados apresentados na tabela ANOVA, o R^2 do modelo é de 0.5324, indicando que o valor da hora-aluno é explicado 53% pelos três fatores propostos, e os outros 47% são explicados por outros fatores, os quais constam como residuais. Por mais que, a princípio, pareça ser pouca a explicação, é necessário considerar que modelos com variáveis dependentes qualitativas/categóricas apresentam valores menores para R^2 do que modelos com variáveis independentes quantitativas. Conclui-se, portanto, que o ajuste é suficiente para o problema estudado.

5.2.1 Análise dos Resíduos ANOVA 1

Dado que todos os fatores foram considerados significativos para o modelo, faz-se necessário realizar a análise dos resíduos do modelo, para verificar o cumprimento dos pressupostos/suposições da ANOVA, relacionados no item 4.1.1.

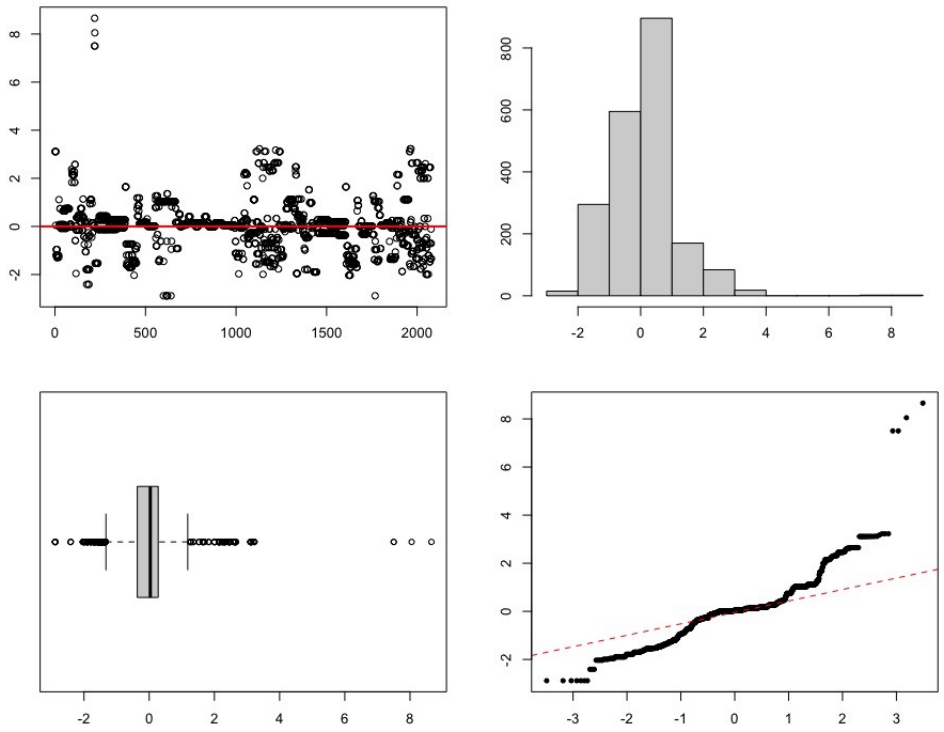
Para a análise dos resíduos, foram utilizados os resíduos studentizados, os quais incorparam as variâncias individuais dos resíduos no escalonamento, sendo representados por:

$$t_i = \frac{r_i}{\sqrt{QMRes(1 - h_{ii})}} \tag{5.2.2}$$

onde i representa cada observação da base de dados e $QMRes$ o quadrado médio dos resíduos, da tabela ANOVA, o qual indica a estimativa de σ_2 gerada pelo modelo ajustado.

Abaixo, são apresentados quatro gráficos. O primeiro é um gráfico de dispersão dos resíduos studentizados, por meio do qual é possível identificar uma tendência dos resíduos estarem distribuídos em relação à media zero, marcada pela linha vermelha. A indicação de que os resíduos possuem média zero também pode ser verificada no segundo gráfico, um histograma, o qual demonstra uma concentração dos resíduos em torno do valor nulo. Essa mesma tendência pode ser visualizada no terceiro gráfico, no boxplot, cuja média está marcada no valor zero.

Figura 4: Análise gráfica dos resíduos studentizados do modelo 1, com dados discrepantes(dados treinamento).



Fonte: Elaborado pelo autor.

A independência dos resíduos pode ser verificada no gráfico de dispersão, onde não é possível verificar nenhum padrão de comportamento dos resíduos, indicando um

comportamento aleatório.

Em relação ao pressuposto da variância constante, o gráfico de dispersão também indica uma variação distribuída uniformemente em torno da média, com exceção de somente três resíduos.

Por fim, ao tratar da normalidade dos resíduos, há indícios de que a distribuição dos resíduos se aproxime a uma distribuição com média zero e variância constante, o que pode ser observado pelo histograma, pelo boxplot e pelo gráfico dos quantis teóricos normais. Sobre esse último, as extremidades não convergem para a normalidade, conforme evidenciado no boxplot, porém, a parte central evidencia um ajuste adequado à normalidade.

Pela análise gráfica dos resíduos, há evidências para concluir pelo atendimentos dos pressupostos de normalidade, independência e variância constante dos resíduos.

Outra maneira de verificar o atendimento dos pressupostos do modelo é a utilização de testes estatísticos. Porém, quando o número de observações é elevado, como é o caso em estudo, modelo estimado com 2079 observações, os testes indicam a rejeição das hipóteses nulas de normalidade e de igualdade de variâncias, pois qualquer divergência ou variação possui um grande peso no ajuste. Neste sentido, para o presente estudo, os resíduos serão analisados apenas graficamente.

5.3 Análise de Variância 2: sem Dados Discrepantes

Seguindo a mesma metodologia utilizada na ANOVA 1, com os dados discrepantes, as 2367 observações da base de dados sem os dados discrepantes da variável resposta foram divididas em base de teste, com 80% dos dados, e em base de treino, com 20% dos dados. A ANOVA 2, com 80% da base de dados sem os valores discrepantes de Y_{ijk}, apresentou os seguintes resultados:

5.3.1 Análise dos Resíduos ANOVA 2

A Figura 4 corresponde à análise gráfica dos resíduos studentizados do modelo 2, gerado com 80% dos dados da base sem os valores discrepantes de Y_{ijk}. Ao compararmos com a Figura 3, que contém a análise gráfica dos resíduos studentizados do modelo 1, verifica-se uma maior homogeneidade na dispersão dos resíduos do modelo 2, além desse último modelo apresentar uma distribuição mais evidenciada para a normal, verificada no histograma, em comparação com a distribuição assimétrica do modelo 1. Outro fator de análise refere-se aos pontos discrepantes dos resíduos. Enquanto no modelo 1 o boxplot apresentava peso maior para a cauda direita, os resíduos do modelo 2 distribuem-se mais

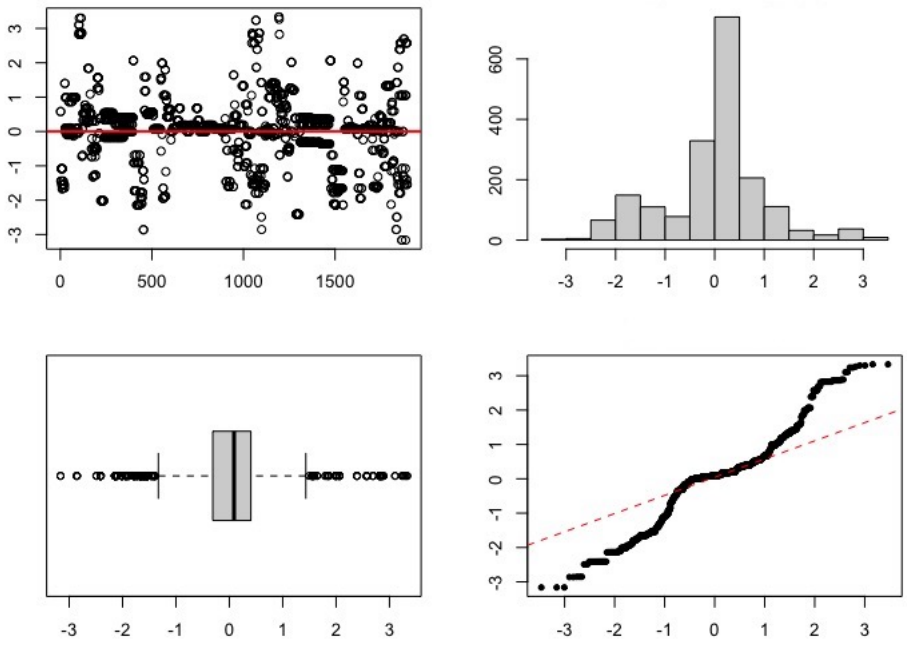
Tabela 7: ANOVA sem dados discrepantes.

Fonte de Variação	**Graus Liberdade (GL)**	**Soma de Quadrados (SQ)**	**Quadrado Médio (QM)**	**Estatística F**	**p-valor**
Fator A (Eixo Tecnológico)	10	8.7	0.87	8.934	$1.75e^{-14}$ ***
Fator B (Região)	4	186.0	46.51	477.793	$< 2e^{-16}$ ***
Fator C (Tipo Instituição)	1	2.3	2.28	23.474	$1.37e^{-6}$ ***
Interação AB	37	19.2	0.52	5.336	$< 2e^{-16}$ ***
Interação AC	9	5.3	0.59	6.042	$2.16e^{-8}$ ***
Interação BC	1	7.7	7.71	79.181	$< 2e^{-16}$ ***
Interação ABC	7	2.6	0.37	3.826	0.000389 ***
Resíduos	1824	177.5	0.10	–	–
Total	1893	409.3	–	–	–

Fonte: Elaborado pelo autor.

igualmente em torno da média inclusive os valores discrepantes. Por fim, a convergência dos resíduos com os quantis da normal é melhor ajustada no modelo 2, que possui caudas mais pesadas, mas igualmente distribuídas para ambos os lados da distribuição.

Figura 5: Análise gráfica dos resíduos studentizados do modelo 2, sem dados discrepantes (dados treinamento).



Fonte: Elaborado pelo autor.

Portanto, de acordo com a análise gráfica dos resíduos studentizados dos modelos,

conclui-se que o modelo 2, gerado com 80% das observações, amostradas aleatoriamente, da base de dados sem os valores discrepantes da variável resposta, está melhor ajustado, sendo, portanto, o modelo escolhido para a estimação do valor executado efetivo da hora-aluno dos cursos FIC EaD ofertados pelas instituições da Rede Federal, no âmbito do Pronatec.

5.4 Estimação dos Coeficientes do Modelo 2

No Apêndice 1, são apresentados todos os coeficientes/parâmetros estimados para o modelo 2, os quais serão utilizados para a previsão/estimação do valor da hora-aluno. São apresentadas a estimação pontual e a estimação intervalar, considerando um intevalo de confiança de 95% e o erro-padrão do modelo, de 0.3168.

Como exemplo, para um curso do eixo tecnológico Informação e Comunicação ($i = 5$), ofertado por uma Escola Técnica vinculada à Universidade Federal ($k = 2$) que está sediada na região Nordeste ($j = 1$), o modelo de previsão do valor da hora-aluno é dado por:

$$E[Y_{ijk}] = \hat{Y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k + \alpha_i\beta_j\gamma_k \tag{5.4.1}$$

$$= \mu + \alpha_5 + \beta_1 + \gamma_2 + \alpha_5\beta_1 + \alpha_5\gamma_2 + \beta_1\gamma_2 + \alpha_5\beta_1\gamma_2 \tag{5.4.2}$$

$$= 1.6470 + 0.0918 - 0.3068 + 0.0 - 0.4014 + 0.0 + 0.0 + 0.0 \tag{5.4.3}$$

$$= 1.0306 \tag{5.4.4}$$

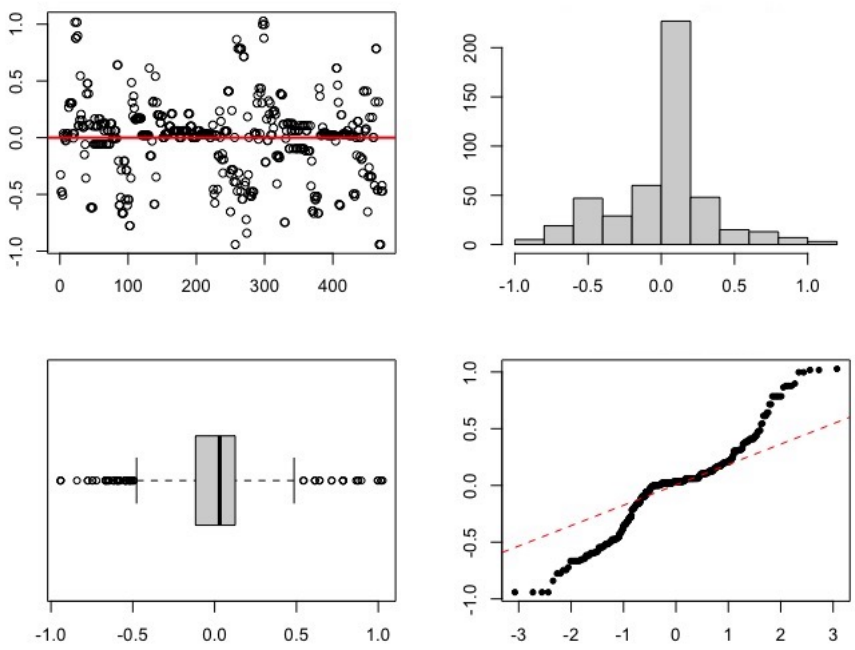
Para essa combinação de categorias dos fatores, verifica-se que todos os parâmetros que possuem γ_2 são nulos, ou seja, o fato de a instituição de ensino ofertante ser Escola Técnica não influencia no valor da hora-aluno. Assim, para as categorias indicadas, o valor estimado da hora-aluno é de R$ 1,03. Após a definição das categorias dos três fatores, os coeficientes podem ser obtidos via consulta à tabela do apêndice 1 e, substituindo-os no modelo, o valor da hora-aluno pode ser estimado.

5.5 Aplicação do Modelo 2 nos Dados de Teste

Com o objetivo de testar o modelo estimado com as 1.894 observações (80%) selecionadas aleatoriamente, foram reservadas, como base de dados de teste, as 473 observações restantes que não foram selecionadas para a estimação do modelo. Ao aplicar os coeficientes/parâmetros estimados nessas 473 observações, foram analisados, graficamente, o comportamento dos erros/resíduos, que consistem na diferença entre o valor da

hora-aluno observado e o estimado. Na Figura 6, são apresentados os resultados.

Figura 6: Análise gráfica dos erros/resíduos dos dados de teste.



Fonte: Elaborado pelo autor.

Assim como na análise gráfica dos erros/resíduos do modelo, é possível verificar a evidência de normalidade dos resíduos dos dados de teste, com média zero e variância constante em torno da média. Tais resultados confirmam a adequação do modelo.

5.6 Valores das Horas-Aluno

Os valores estimados pelo modelo podem não ser muito práticos para a adoção em larga escala, no âmbito de políticas públicas federais de financiamento de cursos de Educação Profissional e Tecnológica, por isso, no apêndice 3, é apresentada uma relação com todas as combinações possíveis dos três fatores que compõem a estimativa do valor da hora-aluno, contendo os respectivos valores estimados, estimativas pontuais e intervalares, para uma consulta mais rápida e prática.

Na mesma tabela, são indicados os valores sugeridos para serem aplicados na prática, os quais variam a cada R$ 0,50. Assim, no exemplo do item 5.4, para o valor da estimação pontual de R$ 1,03, verifica-se o limite máximo do intervalo de confiança de 95%, que é de R$ 1,65 e atribui-se o valor múltiplo de R$ 0,50 mais próximo desse limite

superior. Assim, nesse exemplo, é sugerida a adoção do valor de R$ 1,50 como referência. Todos os valores sugeridos no apêndice 3 consideram o múltiplo de R$ 0,50 mais próximo do limite superior da estimação intervalar, com o objetivo de maximizar a padronizar o valor da hora-aluno a ser repassado às instituições ofertantes.

Por fim, é importante esclarecer que nem todas as combinações possíveis entre os três fatores estavam disponíveis na base de dados analisada. Por isso, na tabela do apêndice 3, é indicado se a combinação estava ou não disponível na base de dados. A disponibilização dessa informação é relevante para a tomada de decisão com base nos valores estimados e sugeridos pelo presente estudo, pois, caso a combinação de interesse não tenha sido contemplada nos dados utilizados para a estimação dos parâmetros/coeficientes do modelo, pode ser que o valor estimado não esteja muito alinhado à realidade.

6 Conclusão

A oferta de cursos de Educação Profissional e Tecnológica - EPT pelas instituições de ensino que compõem a Rede Federal de Educação Profissional, Científica e Tecnológica - RFEPCT é um dos principais instrumentos de difusão da EPT no Brasil. O Programa Nacional de Acesso ao Ensino Técnico e Emprego - Pronatec, instituído em 2011, tornou-se uma importante fonte de financiamento da oferta desses cursos. Porém, ao longo dos anos, o Pronatec precisou se adaptar às mudanças das demandas sociais e econômicas, principalmente durante o ano de 2020, marcado pela pandemia causada pela Covid-19.

O presente estudo analisou os dados referentes aos cursos ofertados durante o ano de 2020 pelas instituições de ensino da RFEPCT, que foram ministrados na modalidade de Educação a Distância - EaD e consistiram em cursos de curta duração, denominados de Formação Inicial e Continuada - FIC ou de Qualificação Profissional.

A principal motivação dessa análise foi o fato de que, ao final da oferta dos cursos, as instituições de ensino ofertantes informaram uma necessidade menor de recursos para custear as despesas da oferta do que os valores aprovados/pactuados, sendo que esses útimos foram apresentados pelas próprias instituições de ensino. Portanto, há evidências de que o valor previsto para a hora-aluno, atualmente com parâmetro máximo de R$ 4,50 (quatro reais e cinquenta centavos), seja efetivamente menor do que esse máximo estipulado em normativos federais.

As análises descritivas dos dados mostraram que os eixos tecnológicos, as regiões geográficas onde as instituições de ensino estão sediadas e o tipo de instituição, se Escola Técnica vinculada às Universidades Federais - ET ou Institutos Federais de Educação Profissional, Científica e Tecnológica - IF, constituem fatores categóricos que podem influenciar no valor da hora-aluno executado.

A técnica estatística mais adequada para o delineamento do problema identificado é a Análise de Variância - ANOVA, cujo objetivo principal é identificar se os efeitos (médias) das categorias, ou fatores, são iguais ou se divergem entre si. Como resultado, verificou-se que os três fatores, eixo tecnológico, região geográfica e tipo de instituição, apresentam efeitos diferentes entre pelo menos duas categorias, de todas as possíveis, assim como a interação entre as diversas categorias desses fatores.

Dos resultados obtidos, é importante destacar que, dos onze eixos tecnológicos analisados, somente o eixo de Ambiente e Saúde possui influência nula sobre o valor da hora-aluno. Em relação à região geográfica, a região Centro-Oeste também não apresenta efeito significativo na estimação do valor da hora-aluno. Por fim, no que tange ao tipo de instituição, a oferta do curso por um IF apresenta um efeito positivo no valor da hora-aluno, enquanto a oferta por um ET não apresenta efeito significativo.

Como resultado da pesquisa, foi disponibilizada uma relação de todos os valores estimados e sugeridos para as horas-alunos, considerando todas as combinações possíveis entre os onze eixos tecnológicos, as cinco regiões geográficas e os dois tipos de instituições de ensino. Assim, os tomadores de decisão do Pronatec podem ter acesso ao valor sugerido da hora-aluno para uma determinado curso, inserido em um certo eixo tecnológico, ofertado por uma ET ou um IF, sediado em dada região do país.

É importante ressaltar que a pesquisa apresentou algumas limitações. A primeira é que se restringe às instituições da RFEPCT, não abrangendo a oferta de cursos por outros tipos de instituições pública, como as redes estaduais, distrial e municipais de educação, nem as instituições privadas. A segunda limitação corresponde aos eixos tecnológicos, que são treze, porém, nos dados analisados, não houve oferta para dois desses eixos, os quais não foram contemplados nos resultados. A terceira limitação consiste na não representatividade, nos dados, de todas as combinações possíveis entre os três fatores de interesse, assim, na tabela que contém os valores estimados e sugeridos, é indicada se aquela combinação estava ou não contemplada nos dados analisados. Por fim, a quarta limitação refere-se ao próprio ano de 2020, que apresentou características peculiares devido ao cenário de pandemia pela Covid-19, o que pode ter afetado o padrão de execução dos cursos e, consequentemente, os valores efetivos executados das horas-aluno.

O modelo proposto poderá ser testado para outras variáveis independentes de interesse da política pública, contemplando, em estudos futuros, ofertas de cursos presenciais, de cursos técnicos e ministrados para outras redes públicas e privada de ensino. Ressalta-se que a conclusão de que os valores das horas-alunos dos cursos de EPT diferem devido a fatores específicos amplia a discussão sobre o financiamento do Pronatec, levando a uma reflexão sobre a eficiência e a efetividade da aplicação dos recursos públicos.

7 Referências

BRASIL. Lei nº 12.513, de 26 de outubro de 2011.
Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12513.htm.
Acessado em: 11 de março de 2021.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011.
Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.
Acessado em: 13 de março de 2021.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018.
Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.
Acessado em: 14 de março de 2021.

BRASIL. Ministério da Educação. Portaria nº 817, de 13 de agosto de 2015.
Disponível em: https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/20462540/do1-2017-02-13-portaria-no-817-de-13-de-agosto-de-2015–20462295.
Acessado em: 4 de março de 2021.

KELLSTEDT, P. & WHITTEN, G. D. Fundamentos da Pesquisa em Ciência Política. 2ª edição. São Paulo: Bluncher, 2015.

BUSSAB, W. de O. & MORETTIN, P. A. Estatística Básica. 8ª edição. São Paulo: Saraiva, 2013.

CASELLA, G. & BERGER, R. L. Inferência Estatística. 2ª edição. São Paulo: Cengage Learning, 2014.

SANTOS, G. S. & MARCHESAN, M. T. N. Educação Profissional e Tecnológica no Brasil e seus Docentes: Trajetos e Desafios. Revista de Letras, Artes e Comunicação, v. 11, n. 1, p. 357-374. Blumenau, 2017.

8 Apêndices

8.1 Apêndice 1: Estimativas Coeficientes do Modelo

Tabela 7: Estimativas pontuais dos coeficientes do modelo.

Descrição	Coeficiente	Estimativa	LI (95%)	LS (95%)
Intercepto	μ	1,6470317440	1,0261382400	2,2679252480
Eixo – Ambiente e Saúde	α_1	0,0000000000	-0,6208935040	0,6208935040
Eixo – Controle e Processos Industriais	α_2	0,0137048730	-0,6071886310	0,6345983770
Eixo – Desenvolvimento Educacional e Social	α_3	0,1150356060	-0,5058578980	0,7359291100
Eixo – Gestão e Negócios	α_4	-0,2383842890	-0,8592777930	0,3825092150
Eixo – Informação e Comunicação	α_5	0,0918538470	-0,5290396570	0,7127473510
Eixo – Infraestrutura	α_6	0,7860469910	0,1651534870	1,4069404950
Eixo – Produção Alimentícia	α_7	-0,1100154730	-0,7309089770	0,5108780310
Eixo – Produção Cultural e Design	α_8	0,6001652900	-0,0207282140	1,2210587940
Eixo – Produção Industrial	α_9	-0,2329341820	-0,8538276860	0,3879593220
Eixo – Recursos Naturais	α_{10}	-0,0991592530	-0,7200527570	0,5217342510
Eixo – Turismo, Hospitalidade e Lazer	α_{11}	0,9310493300	0,3101558260	1,5519428340
Região – Nordeste	β_1	-0,3067949300	-0,9276884340	0,3140985740
Região – Norte	β_2	-0,6497415280	-1,2706350320	-0,0288480240
Região – Sudeste	β_3	-0,7958436890	-1,4167371930	-0,1749501850
Região – Sul	β_4	-1,1622097910	-1,7831032950	-0,5413162870
Região – Centro-Oeste	β_5	0,0000000000	-0,6208935040	0,6208935040
Instituição – Instituto Federal	γ_1	0,5263463930	-0,0945471110	1,1472398970
Instituição – Escola Técnica	γ_2	0,0000000000	-0,6208935040	0,6208935040
Nordeste – Ambiente e Saúde	$\alpha_1\beta_1$	0,0000000000	-0,6208935040	0,6208935040
Nordeste – Controle e Processos Industriais	$\alpha_2\beta_1$	-0,1820011710	-0,8028946750	0,4388923330
Nordeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_1$	-0,5033951060	-1,1242886100	0,1174983980
Nordeste – Gestão e Negócios	$\alpha_4\beta_1$	-0,1263747810	-0,7472682850	0,4945187230
Nordeste – Informação e Comunicação	$\alpha_5\beta_1$	-0,4014251660	-1,0223186700	0,2194683380
Nordeste – Infraestrutura	$\alpha_6\beta_1$	0,1949479680	-0,4259455360	0,8158414720
Nordeste – Produção Alimentícia	$\alpha_7\beta_1$	-0,2809200090	-0,9018135130	0,3399734950
Nordeste – Produção Cultural e Design	$\alpha_8\beta_1$	-0,9885247900	-1,6094182940	-0,3676312860
Nordeste – Produção Industrial	$\alpha_9\beta_1$	0,1288517870	-0,4920417170	0,7497452910
Nordeste – Recursos Naturais	$\alpha_{10}\beta_1$	-0,2725387700	-0,8934322740	0,3483547340
Nordeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_1$	-1,3194088300	-1,9403023340	-0,6985153260
Norte – Ambiente e Saúde	$\alpha_1\beta_2$	0,0000000000	-0,6208935040	0,6208935040
Norte – Controle e Processos Industriais	$\alpha_2\beta_2$	0,4599450370	-0,1609484670	1,0808385410
Norte – Desenvolvimento Educacional e Social	$\alpha_3\beta_2$	0,1461967710	-0,4746967330	0,7670902750
Norte – Gestão e Negócios	$\alpha_4\beta_2$	0,3568162860	-0,2640772180	0,9777097900
Norte – Informação e Comunicação	$\alpha_5\beta_2$	0,5002681420	-0,1206253620	1,1211616460
Norte – Infraestrutura	$\alpha_6\beta_2$	-0,1192354770	-0,7401289810	0,5016580270
Norte – Produção Alimentícia	$\alpha_7\beta_2$	-0,1075674400	-0,7284609440	0,5133260640
Norte – Produção Cultural e Design	$\alpha_8\beta_2$	0,1095691440	-0,5113243600	0,7304626480
Norte – Produção Industrial	$\alpha_9\beta_2$	0,0000000000	-0,6208935040	0,6208935040
Norte – Recursos Naturais	$\alpha_{10}\beta_2$	-0,1502630780	-0,7711565820	0,4706304260
Norte – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_2$	0,0258597190	-0,5950337850	0,6467532230

Fonte: Elaborado pelo autor.

Continuação da Tabela 7.

Descrição	Coeficiente	Estimativa	LI (95%)	LS (95%)
Sudeste – Ambiente e Saúde	$\alpha_1\beta_3$	0,0000000000	-0,6208935040	0,6208935040
Sudeste – Controle e Processos Industriais	$\alpha_2\beta_3$	0,4312720690	-0,1896214350	1,0521655730
Sudeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_3$	0,0307971730	-0,5900963310	0,6516906770
Sudeste – Gestão e Negócios	$\alpha_4\beta_3$	0,3169096320	-0,3039838720	0,9378031360
Sudeste – Informação e Comunicação	$\alpha_5\beta_3$	0,4578235530	-0,1630699510	1,0787170570
Sudeste – Infraestrutura	$\alpha_6\beta_3$	0,0777317810	-0,5431617230	0,6986252850
Sudeste – Produção Alimentícia	$\alpha_7\beta_3$	0,1742634800	-0,4466300240	0,7951569840
Sudeste – Produção Cultural e Design	$\alpha_8\beta_3$	-0,2748460130	-0,8957395170	0,3460474910
Sudeste – Produção Industrial	$\alpha_9\beta_3$	0,1544088390	-0,4664846650	0,7753023430
Sudeste – Recursos Naturais	$\alpha_{10}\beta_3$	0,1253343670	-0,4955591370	0,7462278710
Sudeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_3$	-0,0672705590	-0,6881640630	0,5536229450
Sul – Ambiente e Saúde	$\alpha_1\beta_4$	0,0000000000	-0,6208935040	0,6208935040
Sul – Controle e Processos Industriais	$\alpha_2\beta_4$	0,0000000000	-0,6208935040	0,6208935040
Sul – Desenvolvimento Educacional e Social	$\alpha_3\beta_4$	-0,0795335890	-0,7004270930	0,5413599150
Sul – Gestão e Negócios	$\alpha_4\beta_4$	0,2471585990	-0,3737349050	0,8680521030
Sul – Informação e Comunicação	$\alpha_5\beta_4$	0,4022072510	-0,2186862530	1,0231007550
Sul – Infraestrutura	$\alpha_6\beta_4$	0,0000000000	-0,6208935040	0,6208935040
Sul – Produção Alimentícia	$\alpha_7\beta_4$	-0,5197913870	-1,1406848910	0,1011021170
Sul – Produção Cultural e Design	$\alpha_8\beta_4$	0,0484837280	-0,5724097760	0,6693772320
Sul – Produção Industrial	$\alpha_9\beta_4$	-0,0016458620	-0,6225393660	0,6192476420
Sul – Recursos Naturais	$\alpha_{10}\beta_4$	-0,1763435670	-0,7972370710	0,4445499370
Sul – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_4$	-0,5244875530	-1,1453810570	0,0964059510
Centro Oeste – Ambiente e Saúde	$\alpha_1\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Controle e Processos Industriais	$\alpha_2\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Gestão e Negócios	$\alpha_4\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Informação e Comunicação	$\alpha_5\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Infraestrutura	$\alpha_6\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Produção Alimentícia	$\alpha_7\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Produção Cultural e Design	$\alpha_8\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Produção Industrial	$\alpha_9\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Recursos Naturais	$\alpha_{10}\beta_5$	0,0000000000	-0,6208935040	0,6208935040
Centro Oeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_5$	0,0000000000	-0,6208935040	0,6208935040
IF – Ambiente e Saúde	$\alpha_1\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Controle e Processos Industriais	$\alpha_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Desenvolvimento Educacional e Social	$\alpha_3\gamma_1$	-0,1034313320	-0,7243248360	0,5174621720
IF – Gestão e Negócios	$\alpha_4\gamma_1$	-0,0359460360	-0,6568395400	0,5849474680
IF – Informação e Comunicação	$\alpha_5\gamma_1$	-0,4916211870	-1,1125146910	0,1292723170
IF – Infraestrutura	$\alpha_6\gamma_1$	-0,8004349880	-1,4213284920	-0,1795414840
IF – Produção Alimentícia	$\alpha_7\gamma_1$	0,1216955910	-0,4991979130	0,7425890950
IF – Produção Cultural e Design	$\alpha_8\gamma_1$	-0,6187601200	-1,2396536240	0,0021333840
IF – Produção Industrial	$\alpha_9\gamma_1$	0,2644689420	-0,3564245620	0,8853624460
IF – Recursos Naturais	$\alpha_{10}\gamma_1$	0,0984685770	-0,5224249270	0,7193620810
IF – Turismo, Hospitalidade e Lazer	$\alpha_{11}\gamma_1$	-0,9146730460	-1,5355665500	-0,2937795420

Fonte: Elaborado pelo autor.

Continuação da Tabela 7.

Descrição	Coeficiente	Estimativa	LI (95%)	LS (95%)
ET – Ambiente e Saúde	$\alpha_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Controle e Processos Industriais	$\alpha_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Desenvolvimento Educacional e Social	$\alpha_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Gestão e Negócios	$\alpha_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Informação e Comunicação	$\alpha_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Infraestrutura	$\alpha_6\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Produção Alimentícia	$\alpha_7\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Produção Cultural e Design	$\alpha_8\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Produção Industrial	$\alpha_9\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Recursos Naturais	$\alpha_{10}\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Turismo, Hospitalidade e Lazer	$\alpha_{11}\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste	$\beta_1\gamma_1$	-0,9015720860	-1,5224655900	-0,2806785820
IF – Norte	$\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste	$\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul	$\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste	$\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste	$\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte	$\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste	$\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul	$\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste	$\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste – Ambiente e Saúde	$\alpha_1\beta_1\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste – Controle e Processos Industriais	$\alpha_2\beta_1\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_1\gamma_1$	0,5514015480	-0,0694919560	1,1722950520
IF – Nordeste – Gestão e Negócios	$\alpha_4\beta_1\gamma_1$	0,4747274170	-0,1461660870	1,0956209210
IF – Nordeste – Informação e Comunicação	$\alpha_5\beta_1\gamma_1$	0,7234481760	0,1025546720	1,3443416800
IF – Nordeste – Infraestrutura	$\alpha_6\beta_1\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste – Produção Alimentícia	$\alpha_7\beta_1\gamma_1$	0,4884579600	-0,1324355440	1,1093514640
IF – Nordeste – Produção Cultural e Design	$\alpha_8\beta_1\gamma_1$	0,6211242960	0,0002307920	1,2420178000
IF – Nordeste – Produção Industrial	$\alpha_9\beta_1\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Nordeste – Recursos Naturais	$\alpha_{10}\beta_1\gamma_1$	0,2128000150	-0,4080934890	0,8336935190
IF – Nordeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_1\gamma_1$	1,2562037570	0,6353102530	1,8770972610
IF – Norte – Ambiente e Saúde	$\alpha_1\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Controle e Processos Industriais	$\alpha_2\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Desenvolvimento Educacional e Social	$\alpha_3\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Gestão e Negócios	$\alpha_4\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Informação e Comunicação	$\alpha_5\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Infraestrutura	$\alpha_6\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Produção Alimentícia	$\alpha_7\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Produção Cultural e Design	$\alpha_8\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Produção Industrial	$\alpha_9\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Recursos Naturais	$\alpha_{10}\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Norte – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_2\gamma_1$	0,0000000000	-0,6208935040	0,6208935040

Fonte: Elaborado pelo autor.

Continuação da Tabela 7.

Descrição	Coeficiente	Estimativa	LI (95%)	LS (95%)
IF – Sudeste – Ambiente e Saúde	$\alpha_1\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Controle e Processos Industriais	$\alpha_2\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Gestão e Negócios	$\alpha_4\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Informação e Comunicação	$\alpha_5\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste - Infraestrutura	$\alpha_6\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Produção Alimentícia	$\alpha_7\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Produção Cultural e Design	$\alpha_8\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Produção Industrial	$\alpha_9\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Recursos Naturais	$\alpha_{10}\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sudeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_3\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Ambiente e Saúde	$\alpha_1\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Controle e Processos Industriais	$\alpha_2\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Desenvolvimento Educacional e Social	$\alpha_3\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Gestão e Negócios	$\alpha_4\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Informação e Comunicação	$\alpha_5\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Infraestrutura	$\alpha_6\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Produção Alimentícia	$\alpha_7\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Produção Cultural e Design	$\alpha_8\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Produção Industrial	$\alpha_9\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Recursos Naturais	$\alpha_{10}\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Sul – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_4\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Ambiente e Saúde	$\alpha_1\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Controle e Processos Industriais	$\alpha_2\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Gestão e Negócios	$\alpha_4\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Informação e Comunicação	$\alpha_5\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Infraestrutura	$\alpha_6\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Produção Alimentícia	$\alpha_7\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Produção Cultural e Design	$\alpha_8\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Produção Industrial	$\alpha_9\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Recursos Naturais	$\alpha_{10}\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
IF – Centro Oeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_5\gamma_1$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Ambiente e Saúde	$\alpha_1\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Controle e Processos Industriais	$\alpha_2\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Gestão e Negócios	$\alpha_4\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Informação e Comunicação	$\alpha_5\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Infraestrutura	$\alpha_6\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Produção Alimentícia	$\alpha_7\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Produção Cultural e Design	$\alpha_8\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Produção Industrial	$\alpha_9\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Recursos Naturais	$\alpha_{10}\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Nordeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_1\gamma_2$	0,0000000000	-0,6208935040	0,6208935040

Fonte: Elaborado pelo autor.

Continuação da Tabela 7.

Descrição	Coeficiente	Estimativa	LI (95%)	LS (95%)
ET – Norte – Ambiente e Saúde	$\alpha_1\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Controle e Processos Industriais	$\alpha_2\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Desenvolvimento Educacional e Social	$\alpha_3\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Gestão e Negócios	$\alpha_4\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Informação e Comunicação	$\alpha_5\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Infraestrutura	$\alpha_6\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Produção Alimentícia	$\alpha_7\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Produção Cultural e Design	$\alpha_8\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Produção Industrial	$\alpha_9\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Recursos Naturais	$\alpha_{10}\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Norte – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_2\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Ambiente e Saúde	$\alpha_1\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Controle e Processos Industriais	$\alpha_2\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Gestão e Negócios	$\alpha_4\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Informação e Comunicação	$\alpha_5\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste - Infraestrutura	$\alpha_6\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Produção Alimentícia	$\alpha_7\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Produção Cultural e Design	$\alpha_8\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Produção Industrial	$\alpha_9\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Recursos Naturais	$\alpha_{10}\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sudeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_3\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Ambiente e Saúde	$\alpha_1\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Controle e Processos Industriais	$\alpha_2\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Desenvolvimento Educacional e Social	$\alpha_3\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Gestão e Negócios	$\alpha_4\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Informação e Comunicação	$\alpha_5\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Infraestrutura	$\alpha_6\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Produção Alimentícia	$\alpha_7\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Produção Cultural e Design	$\alpha_8\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Produção Industrial	$\alpha_9\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Recursos Naturais	$\alpha_{10}\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Sul – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_4\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Ambiente e Saúde	$\alpha_1\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Controle e Processos Industriais	$\alpha_2\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Desenvolvimento Educacional e Social	$\alpha_3\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Gestão e Negócios	$\alpha_4\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Informação e Comunicação	$\alpha_5\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Infraestrutura	$\alpha_6\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Produção Alimentícia	$\alpha_7\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Produção Cultural e Design	$\alpha_8\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Produção Industrial	$\alpha_9\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Recursos Naturais	$\alpha_{10}\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040
ET – Centro Oeste – Turismo, Hospitalidade e Lazer	$\alpha_{11}\beta_5\gamma_2$	0,0000000000	-0,6208935040	0,6208935040

Fonte: Elaborado pelo autor.

8.2 Apêndice 2: Valores Estimados e Sugeridos para Hora-Aluno

Tabela 8: Valor da hora-aluno estimado e sugerido para Escolas Técnicas.

Instituição	Região	Eixo	Valor Estimado Modelo (R$)	Valor Mínimo Modelo (95%)	Valor Máximo Modelo (95%)	Valor Sugerido (R$)	Oferta na base de dados?
Escola Técnica	Sul	Ambiente e Saúde	0,48	0,00	1,10	1,00	Não
Escola Técnica	Sul	Controle e Processos Industriais	0,50	0,00	1,12	1,00	Não
Escola Técnica	Sul	Desenvolvimento Educacional e Social	0,52	0,00	1,14	1,00	Não
Escola Técnica	Sul	Gestão e Negócios	0,49	0,00	1,11	1,00	Não
Escola Técnica	Sul	Informação e Comunicação	0,98	0,36	1,60	1,50	Não
Escola Técnica	Sul	Infraestrutura	1,27	0,65	1,89	1,50	Não
Escola Técnica	Sul	Produção Alimentícia	0,00	0,00	0,62	0,50	Não
Escola Técnica	Sul	Produção Cultural e Design	1,13	0,51	1,75	1,50	Não
Escola Técnica	Sul	Produção Industrial	0,25	0,00	0,87	0,50	Não
Escola Técnica	Sul	Recursos Naturais	0,21	0,00	0,83	0,50	Não
Escola Técnica	Sul	Turismo, Hospitalidade e Lazer	0,89	0,27	1,51	1,50	Não
Escola Técnica	Sudeste	Ambiente e Saúde	0,85	0,23	1,47	1,00	Sim
Escola Técnica	Sudeste	Controle e Processos Industriais	1,30	0,68	1,92	1,50	Sim
Escola Técnica	Sudeste	Desenvolvimento Educacional e Social	1,00	0,38	1,62	1,50	Sim
Escola Técnica	Sudeste	Gestão e Negócios	0,93	0,31	1,55	1,50	Sim
Escola Técnica	Sudeste	Informação e Comunicação	1,40	0,78	2,02	2,00	Sim
Escola Técnica	Sudeste	Infraestrutura	1,71	1,09	2,33	2,00	Sim
Escola Técnica	Sudeste	Produção Alimentícia	0,92	0,30	1,54	1,50	Sim
Escola Técnica	Sudeste	Produção Cultural e Design	1,18	0,56	1,80	1,50	Sim
Escola Técnica	Sudeste	Produção Industrial	0,77	0,15	1,39	1,00	Sim
Escola Técnica	Sudeste	Recursos Naturais	0,88	0,26	1,50	1,50	Sim
Escola Técnica	Sudeste	Turismo, Hospitalidade e Lazer	1,71	1,09	2,33	2,00	Sim
Escola Técnica	Centro Oeste	Ambiente e Saúde	1,65	1,03	2,27	2,00	Não
Escola Técnica	Centro Oeste	Controle e Processos Industriais	1,66	1,04	2,28	2,00	Não
Escola Técnica	Centro Oeste	Desenvolvimento Educacional e Social	1,76	1,14	2,38	2,00	Não
Escola Técnica	Centro Oeste	Gestão e Negócios	1,41	0,79	2,03	2,00	Não
Escola Técnica	Centro Oeste	Informação e Comunicação	1,74	1,12	2,36	2,00	Não
Escola Técnica	Centro Oeste	Infraestrutura	2,43	1,81	3,05	3,00	Não
Escola Técnica	Centro Oeste	Produção Alimentícia	1,54	0,92	2,16	2,00	Não
Escola Técnica	Centro Oeste	Produção Cultural e Design	2,25	1,63	2,87	2,50	Não
Escola Técnica	Centro Oeste	Produção Industrial	1,41	0,79	2,03	2,00	Não
Escola Técnica	Centro Oeste	Recursos Naturais	1,55	0,93	2,17	2,00	Não
Escola Técnica	Centro Oeste	Turismo, Hospitalidade e Lazer	2,58	1,96	3,20	3,00	Não
Escola Técnica	Nordeste	Ambiente e Saúde	1,34	0,72	1,96	1,50	Sim
Escola Técnica	Nordeste	Controle e Processos Industriais	1,17	0,55	1,79	1,50	Não
Escola Técnica	Nordeste	Desenvolvimento Educacional e Social	0,95	0,33	1,57	1,50	Sim
Escola Técnica	Nordeste	Gestão e Negócios	0,97	0,35	1,59	1,50	Sim
Escola Técnica	Nordeste	Informação e Comunicação	1,03	0,41	1,65	1,50	Sim
Escola Técnica	Nordeste	Infraestrutura	2,32	1,70	2,94	2,50	Não
Escola Técnica	Nordeste	Produção Alimentícia	0,95	0,33	1,57	1,50	Sim
Escola Técnica	Nordeste	Produção Cultural e Design	0,95	0,33	1,57	1,50	Sim
Escola Técnica	Nordeste	Produção Industrial	1,24	0,62	1,86	1,50	Não
Escola Técnica	Nordeste	Recursos Naturais	0,97	0,35	1,59	1,50	Sim
Escola Técnica	Nordeste	Turismo, Hospitalidade e Lazer	0,95	0,33	1,57	1,50	Sim
Escola Técnica	Norte	Ambiente e Saúde	1,00	0,38	1,62	1,50	Não
Escola Técnica	Norte	Controle e Processos Industriais	1,47	0,85	2,09	2,00	Não
Escola Técnica	Norte	Desenvolvimento Educacional e Social	1,26	0,64	1,88	1,50	Não
Escola Técnica	Norte	Gestão e Negócios	1,12	0,50	1,74	1,50	Não
Escola Técnica	Norte	Informação e Comunicação	1,59	0,97	2,21	2,00	Não
Escola Técnica	Norte	Infraestrutura	1,66	1,04	2,28	2,00	Não
Escola Técnica	Norte	Produção Alimentícia	0,78	0,16	1,40	1,00	Não
Escola Técnica	Norte	Produção Cultural e Design	1,71	1,09	2,33	2,00	Não
Escola Técnica	Norte	Produção Industrial	0,76	0,14	1,38	1,00	Não
Escola Técnica	Norte	Recursos Naturais	0,75	0,13	1,37	1,00	Não
Escola Técnica	Norte	Turismo, Hospitalidade e Lazer	1,95	1,33	2,57	2,50	Não

Fonte: Elaborado pelo autor.

Tabela 9: Valor da hora-aluno estimado e sugerido para Institutos Federais.

Instituição	Região	Eixo	Valor Estimado Modelo (R$)	Valor Mínimo Modelo (95%)	Valor Máximo Modelo (95%)	Valor Sugerido (R$)	Oferta na base de dados?
Instituto Federal	Sul	Ambiente e Saúde	1,01	0,39	1,63	1,50	Sim
Instituto Federal	Sul	Controle e Processos Industriais	1,02	0,40	1,64	1,50	Sim
Instituto Federal	Sul	Desenvolvimento Educacional e Social	0,94	0,32	1,56	1,50	Sim
Instituto Federal	Sul	Gestão e Negócios	0,98	0,36	1,60	1,50	Sim
Instituto Federal	Sul	Informação e Comunicação	1,01	0,39	1,63	1,50	Sim
Instituto Federal	Sul	Infraestrutura	1,00	0,38	1,62	1,50	Não
Instituto Federal	Sul	Produção Alimentícia	0,50	0,00	1,12	1,00	Sim
Instituto Federal	Sul	Produção Cultural e Design	1,04	0,42	1,66	1,50	Sim
Instituto Federal	Sul	Produção Industrial	1,04	0,42	1,66	1,50	Sim
Instituto Federal	Sul	Recursos Naturais	0,83	0,21	1,45	1,00	Sim
Instituto Federal	Sul	Turismo, Hospitalidade e Lazer	0,50	0,00	1,12	1,00	Sim
Instituto Federal	Sudeste	Ambiente e Saúde	1,38	0,76	2,00	2,00	Sim
Instituto Federal	Sudeste	Controle e Processos Industriais	1,82	1,20	2,44	2,00	Não
Instituto Federal	Sudeste	Desenvolvimento Educacional e Social	1,42	0,80	2,04	2,00	Sim
Instituto Federal	Sudeste	Gestão e Negócios	1,42	0,80	2,04	2,00	Sim
Instituto Federal	Sudeste	Informação e Comunicação	1,44	0,82	2,06	2,00	Sim
Instituto Federal	Sudeste	Infraestrutura	1,44	0,82	2,06	2,00	Sim
Instituto Federal	Sudeste	Produção Alimentícia	1,56	0,94	2,18	2,00	Sim
Instituto Federal	Sudeste	Produção Cultural e Design	1,08	0,46	1,70	1,50	Sim
Instituto Federal	Sudeste	Produção Industrial	1,56	0,94	2,18	2,00	Sim
Instituto Federal	Sudeste	Recursos Naturais	1,50	0,88	2,12	2,00	Sim
Instituto Federal	Sudeste	Turismo, Hospitalidade e Lazer	1,33	0,71	1,95	1,50	Sim
Instituto Federal	Centro Oeste	Ambiente e Saúde	2,17	1,55	2,79	2,50	Sim
Instituto Federal	Centro Oeste	Controle e Processos Industriais	2,19	1,57	2,81	2,50	Não
Instituto Federal	Centro Oeste	Desenvolvimento Educacional e Social	2,18	1,56	2,80	2,50	Sim
Instituto Federal	Centro Oeste	Gestão e Negócios	1,90	1,28	2,52	2,50	Sim
Instituto Federal	Centro Oeste	Informação e Comunicação	1,77	1,15	2,39	2,00	Sim
Instituto Federal	Centro Oeste	Infraestrutura	2,16	1,54	2,78	2,50	Sim
Instituto Federal	Centro Oeste	Produção Alimentícia	2,18	1,56	2,80	2,50	Sim
Instituto Federal	Centro Oeste	Produção Cultural e Design	2,15	1,53	2,77	2,50	Sim
Instituto Federal	Centro Oeste	Produção Industrial	2,20	1,58	2,82	2,50	Sim
Instituto Federal	Centro Oeste	Recursos Naturais	2,17	1,55	2,79	2,50	Sim
Instituto Federal	Centro Oeste	Turismo, Hospitalidade e Lazer	2,19	1,57	2,81	2,50	Sim
Instituto Federal	Nordeste	Ambiente e Saúde	0,97	0,35	1,59	1,50	Sim
Instituto Federal	Nordeste	Controle e Processos Industriais	0,80	0,18	1,42	1,00	Sim
Instituto Federal	Nordeste	Desenvolvimento Educacional e Social	1,02	0,40	1,64	1,50	Sim
Instituto Federal	Nordeste	Gestão e Negócios	1,04	0,42	1,66	1,50	Sim
Instituto Federal	Nordeste	Informação e Comunicação	0,89	0,27	1,51	1,50	Sim
Instituto Federal	Nordeste	Infraestrutura	1,15	0,53	1,77	1,50	Sim
Instituto Federal	Nordeste	Produção Alimentícia	1,18	0,56	1,80	1,50	Sim
Instituto Federal	Nordeste	Produção Cultural e Design	0,58	0,00	1,20	1,00	Sim
Instituto Federal	Nordeste	Produção Industrial	1,12	0,50	1,74	1,50	Sim
Instituto Federal	Nordeste	Recursos Naturais	0,90	0,28	1,52	1,50	Sim
Instituto Federal	Nordeste	Turismo, Hospitalidade e Lazer	0,92	0,30	1,54	1,50	Sim
Instituto Federal	Norte	Ambiente e Saúde	1,52	0,90	2,14	2,00	Sim
Instituto Federal	Norte	Controle e Processos Industriais	2,00	1,38	2,62	2,50	Sim
Instituto Federal	Norte	Desenvolvimento Educacional e Social	1,68	1,06	2,30	2,00	Sim
Instituto Federal	Norte	Gestão e Negócios	1,61	0,99	2,23	2,00	Sim
Instituto Federal	Norte	Informação e Comunicação	1,62	1,00	2,24	2,00	Sim
Instituto Federal	Norte	Infraestrutura	1,39	0,77	2,01	2,00	Sim
Instituto Federal	Norte	Produção Alimentícia	1,43	0,81	2,05	2,00	Sim
Instituto Federal	Norte	Produção Cultural e Design	1,61	0,99	2,23	2,00	Sim
Instituto Federal	Norte	Produção Industrial	1,56	0,94	2,18	2,00	Não
Instituto Federal	Norte	Recursos Naturais	1,37	0,75	1,99	1,50	Sim
Instituto Federal	Norte	Turismo, Hospitalidade e Lazer	1,57	0,95	2,19	2,00	Sim

Fonte: Elaborado pelo autor.

8.3 Apêndice 3: Código do R-Studio

```
### TCC 2 - ESTATISTICA - 2021/1 ###

install.packages("readxl")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("nortest")
install.packages("geobr")
install.packages("forecast")

library(readxl)
library(dplyr)
library(ggplot2)
library(nortest)
library(geobr)

### (1) Importando Base de Dados Turmas - Dados Consolidados ###

dados <- read_excel("../0-TCC-1-DADOS-CONSOLIDADOS.xlsx",sheet=4,col_names=TRUE)

head(dados) # com nomes das colunas
dim(dados) # 2.613 observacoes / 19 variaveis
class(dados) # tibble / data frame
colnames(dados) # nomes das variaveis

# dados x1: Sigla da instituicao (chr)
# dados x2: Estado do local da oferta (chr)
# dados x3: Codigo IBGE do municipio (num)
# dados x4: Municipio do local da oferta (chr)
# dados x5: Nome da unidade de ensino (chr)
# dados x6: Tipo de curso (chr)
# dados x7: Eixo tecnologico (chr)
# dados x8: Nome do curso (chr)
# dados x9: Periodo de pactuacao (chr)
# dados x10: Modalidade educacional (chr)
# dados x11: Carga horaria (num)
# dados x12: Vagas (num)
# dados x13: Situacao de turma (chr)
```

```
# dados x14: Data de inicio (posix)
# dados x15: Data prevista de termino (posix)
# dados x16: Matriculas (num)
# dados x17: TED - incluida (num)
# dados x18: Valor hora-aluno proposto - incluida (num)
# dados x19: Tipo de vagas - incluida (chr)

### (2) Renomeando as Variaveis ###

colnames(dados)[1] <- 'x1_inst'
colnames(dados)[2] <- 'x2_uf'
colnames(dados)[3] <- 'x3_cod_ibge_mun'
colnames(dados)[4] <- 'x4_mun'
colnames(dados)[5] <- 'x5_nome_ue'
colnames(dados)[6] <- 'x6_tp_curso'
colnames(dados)[7] <- 'x7_eixo_tec'
colnames(dados)[8] <- 'x8_nome_curso'
colnames(dados)[9] <- 'x9_per_pact'
colnames(dados)[10] <- 'x10_mod_of'
colnames(dados)[11] <- 'x11_ch'
colnames(dados)[12] <- 'x12_vagas'
colnames(dados)[13] <- 'x13_st_turma'
colnames(dados)[14] <- 'x14_dt_inicio'
colnames(dados)[15] <- 'x15_dt_termino'
colnames(dados)[16] <- 'x16_mat'
colnames(dados)[17] <- 'x17_ted'
colnames(dados)[18] <- 'x18_vl_ha_prop'
colnames(dados)[19] <- 'x19_tp_vagas'

head(dados)
names(dados)

### (3) Analisando os Valores das Variaveis e Ajustando a Base de Dados ###

# x1_inst: sigla da instituicao
class(dados$x1_inst) # character - ok
table(dados$x1_inst) # ok
sum(table(dados$x1_inst)) # 2.613 - ok
```

```
# x2_uf: estado do local da oferta
class(dados$x2_uf) # character - ok
table(dados$x2_uf) # ok
sum(table(dados$x2_uf)) # 2.613 - ok

# x3_cod_ibge_mun: codigo IBGE do municipio
class(dados$x3_cod_ibge_mun) # numeric - alterar para character
dados$x3_cod_ibge_mun <- as.character(dados$x3_cod_ibge_mun)
class(dados$x3_cod_ibge_mun) # character - ok
table(dados$x3_cod_ibge_mun) # ok
sum(table(dados$x3_cod_ibge_mun)) # 2.613 - ok

# x4_mun: municipio do local da oferta
class(dados$x4_mun) # character - ok
table(dados$x4_mun) # ok
sum(table(dados$x4_mun)) # 2.613 - ok

# x5_nome_ue: nome da unidade de ensino
class(dados$x5_nome_ue) # character - ok
table(dados$x5_nome_ue) # ok
sum(table(dados$x5_nome_ue)) # 2.613 - ok

# x6_tp_curso: tipo de curso
class(dados$x6_tp_curso) # character - ok
table(dados$x6_tp_curso) # ok - somente FIC
sum(table(dados$x6_tp_curso)) # 2.613 - ok

# x7_eixo_tec: eixo tecnologico
class(dados$x7_eixo_tec) # character - ok
table(dados$x7_eixo_tec) # ok - 11 eixos
sum(table(dados$x7_eixo_tec)) # 2.613 - ok

# x8_nome_curso: nome do curso
class(dados$x8_nome_curso) # character - ok
table(dados$x8_nome_curso) # ok - nomes dos cursos padronizados
sum(table(dados$x8_nome_curso)) # 2.613 - ok

# x9_per_pact: periodo de pactuacao
class(dados$x9_per_pact) # character - ok
```

```
table(dados$x9_per_pact) # ok
sum(table(dados$x9_per_pact)) # 2.613 - ok

# x10_mod_of: modalidade educacional
class(dados$x10_mod_of) # character - ok
table(dados$x10_mod_of) # ok - somente "a distancia"
sum(table(dados$x10_mod_of)) # 2.613 - ok

# x11_ch: carga horaria
class(dados$x11_ch) # numeric - ok
table(dados$x11_ch) # ok
sum(table(dados$x11_ch)) # 2.613 - ok

# x12_vagas: vagas aprovadas
class(dados$x12_vagas) # numeric - ok
table(dados$x12_vagas) # ok
sum(table(dados$x12_vagas)) # 2.613 - ok

# x13_st_turma: situacao de turma
class(dados$x13_st_turma) # character - ok
table(dados$x13_st_turma) # padronizar
dados$x13_st_turma[dados$x13_st_turma %in% "ATIVO"] <- "INICIADA"
dados$x13_st_turma[dados$x13_st_turma %in% "CONCLUÍDO"] <- "CONCLUÍDA"
dados$x13_st_turma[dados$x13_st_turma %in% "PUBLICADA"] <- "INICIADA"
table(dados$x13_st_turma) # ok
sum(table(dados$x13_st_turma)) # 2.613 - ok

# x14_dt_inicio: data de inicio
class(dados$x14_dt_inicio) # posixct - formato de data - ok
table(dados$x14_dt_inicio) # ok
sum(table(dados$x14_dt_inicio)) # 2.376 - ha 237 valores em branco (sem data)

# x15_dt_termino: data prevista de termino
class(dados$x15_dt_termino) # posixct - formato de data - ok
table(dados$x15_dt_termino) # ok
sum(table(dados$x15_dt_termino)) # 2.376 - ha 237 valores em branco (sem data)

# x16_mat: matrículas
class(dados$x16_mat) # numeric - ok
```

```
table(dados$x16_mat) # ok
sum(table(dados$x16_mat)) # 2.613 - ok

# x17_ted: numero do termo de execucaco descentralizada
class(dados$x17_ted) # numeric - alterar para character
dados$x17_ted <- as.character(dados$x17_ted)
class(dados$x17_ted) # character - ok
table(dados$x17_ted) # ok
sum(table(dados$x17_ted)) # 2.613 - ok

# x18_vl_ha_prop: valor da hora-aula proposto
class(dados$x18_vl_ha_prop) # numeric - ok
table(dados$x18_vl_ha_prop) # ok
sum(table(dados$x18_vl_ha_prop)) # 2.613 - ok

# x19_tp_vagas: tipo de vagas
class(dados$x19_tp_vagas) # character - ok
table(dados$x19_tp_vagas) # ok
sum(table(dados$x19_tp_vagas)) # 2.613 - ok

### (4) Incluindo Variaveis na Base de Dados ###

colnames(dados)

# x20_ha_prop: quantidade de horas-aluno propostas por turma
for(i in 1:nrow(dados)){
  dados$x20_ha_prop[i] <- dados$x11_ch[i]*dados$x12_vagas[i]
}
colnames(dados) # variavel x20_ha_prop incluida
class(dados$x20_ha_prop) # numeric - ok
summary(dados$x20_ha_prop) # ok

# x21_vl_total_prop: valor total proposto por turma
for(i in 1:nrow(dados)){
  dados$x21_vl_prop[i] <- dados$x20_ha_prop[i]*dados$x18_vl_ha_prop[i]
}
colnames(dados) # variavel x21_vl_prop
class(dados$x21_vl_prop) # numeric - ok
summary(dados$x21_vl_prop) # ok
```

```
# x22_ha_exec: quantidade de horas-aluno executadas por turma
for(i in 1:nrow(dados)){
  dados$x22_ha_exec[i] <- dados$x11_ch[i]*dados$x16_mat[i]
}
colnames(dados) # variavel x22_ha_exec
class(dados$x22_ha_exec) # numeric - ok
summary(dados$x22_ha_exec) # ok

### (5) Criando a Base de Dados dos TEDs ###

dados$x17_ted <- as.factor(dados$x17_ted)

teds_1 <- dados %>% select(x17_ted,x20_ha_prop,x22_ha_exec) %>%
  group_by(x17_ted) %>%
  summarise(ha_prop_total=sum(x20_ha_prop),ha_exec_total=sum(x22_ha_exec))
teds_2 <- dados %>% select(x17_ted,x21_vl_prop) %>%
  group_by(x17_ted) %>%
  summarise(vl_prop_total=sum(x21_vl_prop))
teds_3 <- left_join(teds_1,teds_2,by="x17_ted")

dados_teds <- teds_3 %>% mutate(vl_prop_ha_efet=vl_prop_total/ha_prop_total)

class(dados_teds)
summary(dados_teds)

class(dados_teds$x17_ted) # factor - ok
class(dados_teds$ha_prop_total) # numeric - ok
class(dados_teds$ha_exec_total) # numeric - ok
class(dados_teds$vl_prop_total) # numeric - ok
class(dados_teds$vl_prop_ha_efet) # numeric - ok

# Importando a Base de Dados Orcamentarios
dados_teds_orc <- read_excel(".../0-TCC-1-DADOS-CONSOLIDADOS.xlsx",sheet=5,
col_names=TRUE)
head(dados_teds_orc)

class(dados_teds_orc$`Valor Aprovado (R$)`) # numeric - ok
class(dados_teds_orc$`Valor Empenhado (R$)`) # numeric - ok
```

```
class(dados_teds_orc$‘Número TED‘) # numeric - alterar para factor
dados_teds_orc$‘Número TED‘ <- as.factor(dados_teds_orc$‘Número TED‘)
class(dados_teds_orc$‘Número TED‘) # factor - ok

# Juntando as Bases de Dados "dados_teds" e "dados_teds_orc"

teds_orc_1 <- dados_teds_orc %>%
  select(‘Número TED‘,SIAFI,‘Valor Aprovado (R$)‘,‘Valor Empenhado (R$)‘)
colnames(teds_orc_1)[1] <- "x17_ted"

dados_teds_1 <- left_join(dados_teds,teds_orc_1,by="x17_ted")
colnames(dados_teds_1)[6] <- "siafi"
colnames(dados_teds_1)[7] <- "vl_aprov"
colnames(dados_teds_1)[8] <- "vl_empen"
colnames(dados_teds_1)

# Conferindo se o valor aprovado calculado (vl_prop_total) e o valor aprovado
indicado (vl_aprov) sao iguais

dados_teds_1 <- dados_teds_1 %>% mutate(delta_vl=vl_prop_total-vl_aprov)
dados_teds_1 %>% filter(delta_vl!=0) # teds 9488, 9644 e 9662
# ted 9488: delta_vl = R$ 32.448,00 (alteracao nos cursos)
# ted 9644: delta_vl = R$ 1.110.600,00 (aditivo no fim do ano)
# ted 9661: delta_vl = - R$ 6.750,00 (diferencao nao identificada)
# Serao considerados os valores aprovados calculados (vl_prop_total)

# Calculando o valor da hora-aluno executado

dados_teds_1 <- dados_teds_1 %>% mutate(vl_exec_ha_efet=vl_empen/ha_exec_total)
colnames(dados_teds_1)

### (6) Inserindo as Variaveis Calculadas TEDs na Base de Dados das Turmas ###

# ha_prop_total: nao inserir, poir e por TED, e nao por turma
# ha_exec_total: nao inserir, poir e por TED, e nao por turma
# vl_prop_total: nao inserir, pois e por TED, e nao por turma
# vl_prop_ha_efet: inserir ***
# siafi: inserir ***
# vl_aprov: nao inserir
```

```
# vl_empen: nao inserir, pois e por TED, e nao por turma
# delta_vl: nao inserir, somente para conferencia
# vl_exec_ha_efet: inserir ***

dados_teds_2 <- dados_teds_1 %>%
select(x17_ted,siafi,vl_prop_ha_efet,vl_exec_ha_efet)

dados <- left_join(dados,dados_teds_2,by="x17_ted")
colnames(dados)

colnames(dados)[23] <- "x23_siafi"
colnames(dados)[24] <- "x24_vl_prop_ha_efet"
colnames(dados)[25] <- "x25_vl_exec_ha_efet"

colnames(dados)
class(dados$x23_siafi) # character - alterar para factor
dados$x23_siafi <- as.factor(dados$x23_siafi)
class(dados$x23_siafi) # factor - ok
class(dados$x24_vl_prop_ha_efet) # numeric - ok
class(dados$x25_vl_exec_ha_efet) # numeric - ok

dados$regiao <- NA
for(i in 1:nrow(dados)){
  if(dados$x2_uf[i] %in% c("AC","AM","AP","PA","RO","RR","TO")){
    dados$regiao[i] <- "Norte"
  }
  if(dados$x2_uf[i] %in% c("AL","BA","MA","PB","PE","PI","RN")){
    dados$regiao[i] <- "Nordeste"
  }
  if(dados$x2_uf[i] %in% c("DF","GO","MS")){
    dados$regiao[i] <- "Centro-Oeste"
  }
  if(dados$x2_uf[i] %in% c("MG")){
    dados$regiao[i] <- "Sudeste"
  }
  if(dados$x2_uf[i] %in% c("SC","RS")){
    dados$regiao[i] <- "Sul"
  }
}
```

```
table(dados$regiao)
class(dados$regiao)
dados$regiao <- as.factor(dados$regiao)

table(dados$x1_inst)
dados$tipo_inst <- NA
for(i in 1:nrow(dados)){
  if(dados$x1_inst[i] %in% c("UFPB","UFRN","UFRPE","UFV")){
    dados$tipo_inst[i] <- "Escola Técnica"
  } else {
    dados$tipo_inst[i] <- "Instituto Federal"
  }
}
table(dados$tipo_inst)
class(dados$tipo_inst)
dados$tipo_inst <- as.factor(dados$tipo_inst)

### (7) Analise Exploratoria dos Dados  ###

# Variaveis de interesse

dados2 <- dados %>% select(x1_inst,x2_uf,x7_eixo_tec,x8_nome_curso,
                           x24_vl_prop_ha_efet,x25_vl_exec_ha_efet,
                           regiao,tipo_inst)

table(dados2$x1_inst)
table(dados2$x2_uf)
table(dados2$x7_eixo_tec)
table(dados2$x8_nome_curso)
table(dados2$regiao)
table(dados2$tipo_inst)
summary(dados2$x24_vl_prop_ha_efet) # nao possui valores estranhos
summary(dados2$x25_vl_exec_ha_efet) # valor minimo zero
dados2[dados2$x25_vl_exec_ha_efet==0,] # 14 observacoes do IFSC - retirar
dados2[dados2$x1_inst=="IFSC",]
dados2 <- dados2[dados2$x25_vl_exec_ha_efet!=0,] # retirar observacoes x_25=0

# Mapa com vagas aprovadas por UF
```

```
list_geobr()

estados <- read_state(code_state="all",year=2019)
names(estados)

dados$x2_uf <- as.factor(dados$x2_uf)
mapa1_dados <- dados %>% select(x2_uf,x12_vagas) %>% group_by(x2_uf) %>%
summarise(vagas_uf=sum(x12_vagas))
estados_1 <- left_join(estados,mapa1_dados,by=c("abbrev_state"="x2_uf"))
names(estados_1)

estados_1 %>%
  ggplot() +
  geom_sf(aes(fill=vagas_uf),size=.15)+
  scale_fill_gradient2(mid="white",high="firebrick4")+
  labs(fill="")

# Valor efetivo proposto (x24)

x24 <- dados2$x24_vl_prop_ha_efet
hist(x24)
boxplot(x24)
qqnorm(x24,main="",xlab="Quantis teóricos N(0,1)",pch=20,
       ylab="Valor efetivo proposto h/a (R$)")
qqline(x24,lty=2,col="red")

ks.test(x24,"pnorm") # Kolmogorov-Smirnov
lillie.test(x24) # Lilliefors
cvm.test(x24) # Cramer-von Mises
shapiro.test(x24) # Shapiro-Wilk
sf.test(x24) # Shapiro-Francia
ad.test(x24) # Anderson-Darling

# Valor efetivo executado (x25)

x25 <- dados2$x25_vl_exec_ha_efet
hist(x25)
boxplot(x25)
```

```
qqnorm(x25,main="",xlab="Quantis teóricos N(0,1)",pch=20,
       ylab="Valor efetivo executado h/a (R$)")
qqline(x25,lty=2,col="red")

ks.test(x25,"pnorm") # Kolmogorov-Smirnov
lillie.test(x25) # Lilliefors
cvm.test(x25) # Cramer-von Mises
shapiro.test(x25) # Shapiro-Wilk
sf.test(x25) # Shapiro-Francia
ad.test(x25) # Anderson-Darling

# Transformacao de Box-Cox

library(forecast)

lambda <- BoxCox.lambda(x25,
                        method=c("loglik"),
                        lower=-5,
                        upper=5)
lambda

x25_transformada <- BoxCox(x25,lambda)
x25_transformada

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
plot(x25_transformada)
abline(h=0,col="red",lwd=2)
hist(x25_transformada)
boxplot(x25_transformada,horizontal=TRUE)
qqnorm(x25_transformada,main="",xlab="Quantis teóricos N(0,1)",pch=20,
       ylab="Valor efetivo executado h/a (R$)")
qqline(x25_transformada,lty=2,col="red")

dados2$x25_transf <- NA
dados2$x25_transf <- as.numeric(x25_transformada)

shapiro.test(dados2$x25_transf) # mesmo p-valor: não normal
bartlett.test(dados2$x25_transf~dados2$x7_eixo_tec)
bartlett.test(dados2$x25_transf~dados2$regiao)
```

```
bartlett.test(dados2$x25_transf~dados2$tipo_inst)

hist(dados2$x25_vl_exec_ha_efet,main="Y")
hist(log(dados2$x25_vl_exec_ha_efet),main="ln Y")
hist(exp(dados2$x25_vl_exec_ha_efet),main="exp Y")
hist(dados2$x25_transf,main="Box-Cox")

### Descritivas por eixo tecnologico (x7)

eixos_nomes <- c("AMBIENTE E SAÚDE","CONTROLE E PROCESSOS INDUSTRIAIS",
"DESENVOLVIMENTO EDUCACIONAL E SOCIAL","GESTÃO E NEGÓCIOS",
"INFORMAÇÃO E COMUNICAÇÃO","INFRAESTRUTURA","PRODUÇÃO ALIMENTÍCIA",
"PRODUÇÃO CULTURAL E DESIGN","PRODUÇÃO INDUSTRIAL",
"RECURSOS NATURAIS","TURISMO, HOSPITALIDADE E LAZER")
eixos_nomes_abrev <- c("AS","CPI","DES","GN","IC","INF","PA","PCD","PI","RN","THL")

vagas_eixo <- dados %>% select(x7_eixo_tec,x12_vagas,x16_mat) %>%
  group_by(x7_eixo_tec) %>% summarise(total_ap=sum(x12_vagas),
                                      total_mat=sum(x16_mat))
vagas_eixo[,1] <- eixos_nomes_abrev

ggplot(vagas_eixo,aes(x=x7_eixo_tec,y=total_ap))+
  geom_col(aes(y=total_ap))+
  geom_text(aes(label=round(total_ap,0)),hjust=0.5,vjust=-1)+
  labs(x="Eixo Tecnológico",y="Vagas Aprovadas")+
  ylim(c(0,70000))

desc_x24 <- dados2 %>% select(x7_eixo_tec,x24_vl_prop_ha_efet) %>%
  group_by(x7_eixo_tec) %>%
  summarise(n=sum(x24_vl_prop_ha_efet)/mean(x24_vl_prop_ha_efet),
                                      med=mean(x24_vl_prop_ha_efet),
                                      var=var(x24_vl_prop_ha_efet),
                                      min=min(x24_vl_prop_ha_efet),
                                      max=max(x24_vl_prop_ha_efet),
                                      q1=quantile(x24_vl_prop_ha_efet)[2],
                                      q3=quantile(x24_vl_prop_ha_efet)[4],
                                      li=q1-1.5*(q3-q1),
                                      ls=q3+1.5*(q3-q1))
```

```
sum(desc_x24$n)
mean(dados2$x24_vl_prop_ha_efet)
var(dados2$x24_vl_prop_ha_efet)
min(dados2$x24_vl_prop_ha_efet)
max(dados2$x24_vl_prop_ha_efet)
quantile(dados2$x24_vl_prop_ha_efet)
quantile(dados2$x24_vl_prop_ha_efet)[2]-
1.5*(quantile(dados2$x24_vl_prop_ha_efet)[4]-
quantile(dados2$x24_vl_prop_ha_efet)[2])
quantile(dados2$x24_vl_prop_ha_efet)[4]+
1.5*(quantile(dados2$x24_vl_prop_ha_efet)[4]-
quantile(dados2$x24_vl_prop_ha_efet)[2])

dados2$x7_eixo_tec <- as.factor(dados2$x7_eixo_tec)

desc_x25 <- dados2 %>% select(x7_eixo_tec,x25_vl_exec_ha_efet) %>%
  group_by(x7_eixo_tec) %>%
  summarise(n=sum(x25_vl_exec_ha_efet)/mean(x25_vl_exec_ha_efet),
                                      med=mean(x25_vl_exec_ha_efet),
                                      var=var(x25_vl_exec_ha_efet),
                                      min=min(x25_vl_exec_ha_efet),
                                      max=max(x25_vl_exec_ha_efet),
                                      q1=quantile(x25_vl_exec_ha_efet)[2],
                                      q3=quantile(x25_vl_exec_ha_efet)[4],
                                      li=q1-1.5*(q3-q1),
                                      ls=q3+1.5*(q3-q1))

sum(desc_x25$n)
mean(dados2$x25_vl_exec_ha_efet)
var(dados2$x25_vl_exec_ha_efet)
min(dados2$x25_vl_exec_ha_efet)
max(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)[2]-
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]
-quantile(dados2$x25_vl_exec_ha_efet)[2])
quantile(dados2$x25_vl_exec_ha_efet)[4]+
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]
-quantile(dados2$x25_vl_exec_ha_efet)[2])
```

```
aggregate(dados2$x25_vl_exec_ha_efet~dados2$x7_eixo_tec,dados2,var)
bartlett.test(dados2$x25_vl_exec_ha_efet~dados2$x7_eixo_tec,dados2)

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
hist(dados2$x24_vl_prop_ha_efet,main="Histograma Valor Proposto Efetivo")
boxplot(dados2$x24_vl_prop_ha_efet,
main="Box-plot Valor Proposto Efetivo",horizontal=TRUE)
hist(dados2$x25_vl_exec_ha_efet,main="Histograma Valor Executado Efetivo")
boxplot(dados2$x25_vl_exec_ha_efet,
main="Box-plot Valor Executado Efetivo",horizontal=TRUE)

par(mfrow=c(1,1))
boxplot(dados2$x24_vl_prop_ha_efet ~ dados2$x7_eixo_tec,
        names=c(eixos_nomes_abrev),horizontal=FALSE,outline=TRUE,
        xlab="Eixo Tecnológico",ylab="Valor Proposto Efetivo (R$)")
abline(h=mean(dados2$x24_vl_prop_ha_efet),col="red",lwd=1)

boxplot(dados2$x25_vl_exec_ha_efet ~ dados2$x7_eixo_tec,
        names=c(eixos_nomes_abrev),horizontal=FALSE,outline=TRUE,
        xlab="Eixo Tecnológico",ylab="Valor Executado Efetivo (R$)")
abline(h=mean(dados2$x25_vl_exec_ha_efet),col="red",lwd=1)

### Descritivas por regiao

vagas_regiao <- dados %>% select(regiao,x12_vagas,x16_mat) %>%
  group_by(regiao) %>% summarise(total_ap=sum(x12_vagas),
                                 total_mat=sum(x16_mat))

vagas_regiao$perc_ap <- NA
vagas_regiao$perc_ap <- (vagas_regiao$total_ap/sum(vagas_regiao$total_ap))*100

ggplot(vagas_regiao,aes(x=regiao,y=total_ap))+
  geom_col(aes(y=total_ap))+
  geom_text(aes(label=round(total_ap,0)),hjust=0.5,vjust=-1)+
  labs(x="Região",y="Vagas Aprovadas")+
  ylim(c(0,70000))

desc_regiao <- dados2 %>% select(regiao,x25_vl_exec_ha_efet) %>%
```

```
  group_by(regiao) %>%
  summarise(n=sum(x25_vl_exec_ha_efet)/mean(x25_vl_exec_ha_efet),
                                        med=mean(x25_vl_exec_ha_efet),
                                        var=var(x25_vl_exec_ha_efet),
                                        min=min(x25_vl_exec_ha_efet),
                                        max=max(x25_vl_exec_ha_efet),
                                        q1=quantile(x25_vl_exec_ha_efet)[2],
                                        q3=quantile(x25_vl_exec_ha_efet)[4],
                                        li=q1-1.5*(q3-q1),
                                        ls=q3+1.5*(q3-q1))

sum(desc_regiao$n)
mean(dados2$x25_vl_exec_ha_efet)
var(dados2$x25_vl_exec_ha_efet)
min(dados2$x25_vl_exec_ha_efet)
max(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)[2]-
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]-
quantile(dados2$x25_vl_exec_ha_efet)[2])
quantile(dados2$x25_vl_exec_ha_efet)[4]+
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]-
quantile(dados2$x25_vl_exec_ha_efet)[2])

table(dados2$x1_inst)

boxplot(dados2$x25_vl_exec_ha_efet ~ dados2$regiao,
        horizontal=FALSE,outline=TRUE,
        xlab="Região",ylab="Valor Executado Efetivo (R$)")
abline(h=mean(dados2$x25_vl_exec_ha_efet),col="red",lwd=1)

### Descritivas por tipo de instituicao

table(dados$tipo_inst)

vagas_inst <- dados %>% select(tipo_inst,x12_vagas,x16_mat) %>%
  group_by(tipo_inst) %>% summarise(total_ap=sum(x12_vagas),
                                    total_mat=sum(x16_mat))
```

```
ggplot(vagas_inst,aes(x=tipo_inst,y=total_ap))+
  geom_col(aes(y=total_ap))+
  geom_text(aes(label=round(total_ap,0)),hjust=0.5,vjust=-1)+
  labs(x="Tipo Instituição",y="Vagas Aprovadas")+
  ylim(c(0,180000))

desc_inst <- dados2 %>% select(tipo_inst,x25_vl_exec_ha_efet) %>%
  group_by(tipo_inst) %>%
  summarise(n=sum(x25_vl_exec_ha_efet)/mean(x25_vl_exec_ha_efet),
                                  med=mean(x25_vl_exec_ha_efet),
                                  var=var(x25_vl_exec_ha_efet),
                                  min=min(x25_vl_exec_ha_efet),
                                  max=max(x25_vl_exec_ha_efet),
                                  q1=quantile(x25_vl_exec_ha_efet)[2],
                                  q3=quantile(x25_vl_exec_ha_efet)[4],
                                  li=q1-1.5*(q3-q1),
                                  ls=q3+1.5*(q3-q1))

sum(desc_inst$n)
mean(dados2$x25_vl_exec_ha_efet)
var(dados2$x25_vl_exec_ha_efet)
min(dados2$x25_vl_exec_ha_efet)
max(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)
quantile(dados2$x25_vl_exec_ha_efet)[2]-
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]-
quantile(dados2$x25_vl_exec_ha_efet)[2])
quantile(dados2$x25_vl_exec_ha_efet)[4]+
1.5*(quantile(dados2$x25_vl_exec_ha_efet)[4]-
quantile(dados2$x25_vl_exec_ha_efet)[2])

table(dados2$x1_inst)

boxplot(dados2$x25_vl_exec_ha_efet ~ dados2$tipo_inst,
        horizontal=FALSE,outline=TRUE,
        xlab="Tipo de Instituição",ylab="Valor Executado Efetivo (R$)")
abline(h=mean(dados2$x25_vl_exec_ha_efet),col="red",lwd=1)

# Valores discrepantes do valor executado efetivo da hora-aluno (x25)
```

```
par(mfrow=c(1,1))
boxplot(x25,horizontal=TRUE)

li_x25 <- as.numeric(quantile(x25)[2]-1.5*(quantile(x25)[4]-quantile(x25)[2]))
ls_x25 <- as.numeric(quantile(x25)[4]+1.5*(quantile(x25)[4]-quantile(x25)[2]))

dados_inf <- dados2 %>% select(x1_inst,regiao,x7_eixo_tec,x25_vl_exec_ha_efet)
%>% filter(x25_vl_exec_ha_efet<li_x25) # 70 observacoes abaixo limite inferior
table(dados_inf$x7_eixo_tec) # 10 eixos / THL não consta
table(dados_inf$regiao) # somente sudeste
table(dados_inf$x1_inst) # somente UFV

dados_sup <- dados2 %>% select(x1_inst,regiao,x7_eixo_tec,x25_vl_exec_ha_efet)
%>% filter(x25_vl_exec_ha_efet>ls_x25) # 162 observacoes abaixo limite inferior
table(dados_sup$x7_eixo_tec) # 8 eixos / CPI, PI e THL não constam
table(dados_sup$regiao) # sul não consta
table(dados_sup$x1_inst) # IF Baiano / IFB / IFMA / IFRO / IFTM / UFRN

x25_sem_inf <- dados2[which(dados2$x25_vl_exec_ha_efet>=li_x25),]
dados3 <- x25_sem_inf[which(x25_sem_inf$x25_vl_exec_ha_efet<=ls_x25),]
table(dados3$x1_inst)

summary(x25_sem_inf$x25_vl_exec_ha_efet)
summary(dados3$x25_vl_exec_ha_efet)

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
plot(dados3$x25_vl_exec_ha_efet)
abline(h=mean(dados3$x25_vl_exec_ha_efet),col="red",lwd=2)
hist(dados3$x25_vl_exec_ha_efet)
boxplot(dados3$x25_vl_exec_ha_efet,horizontal=TRUE)
qqnorm(dados3$x25_vl_exec_ha_efet,main="",xlab="Quantis teóricos N(0,1)",pch=20,
       ylab="Valor efetivo executado h/a (R$)")
qqline(dados3$x25_vl_exec_ha_efet,lty=2,col="red")

# Dividindo a base de dados 1: treinamento (80%) e teste (20%)

nrow(dados2) # 2.599 observações
n1 <- round(0.8*nrow(dados2),0) # 2.079
```

```
n2 <- round(nrow(dados2)-0.8*nrow(dados2),0) # 520

amostra_n1 <- sort(sample(nrow(dados2),n1))

dados_treino <- dados2[amostra_n1,]
dados_teste <- dados2[-amostra_n1,]

# Analise de variancia 1: eixo tecnologico, regiao e tipo de instituicao

anova_eixo_reg_inst <- aov(dados_treino$x25_vl_exec_ha_efet ~
                           dados_treino$x7_eixo_tec +
                           dados_treino$regiao + dados_treino$tipo_inst +
                           dados_treino$x7_eixo_tec * dados_treino$regiao +
                           dados_treino$x7_eixo_tec * dados_treino$tipo_inst +
                           dados_treino$regiao * dados_treino$tipo_inst +
                           dados_treino$x7_eixo_tec *
                           dados_treino$regiao * dados_treino$tipo_inst)
summary(anova_eixo_reg_inst)

res_2 <- anova_eixo_reg_inst$residuals
plot(res_2)
abline(h=0,col="red",lwd=2)
hist(res_2)
boxplot(res_2)
qqnorm(res_2,pch=20)
qqline(res_2,lty=2,col="red")

res_2_n <- rstandard(anova_eixo_reg_inst)
plot(res_2_n)
abline(h=0,col="red",lwd=2)
hist(res_2_n)
boxplot(res_2_n)
qqnorm(res_2_n,pch=20)
qqline(res_2_n,lty=2,col="red")

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
res_2_s <- rstudent(anova_eixo_reg_inst)
plot(res_2_s)
abline(h=0,col="red",lwd=2)
```

```
hist(res_2_s)
boxplot(res_2_s,horizontal=TRUE)
qqnorm(res_2_s,pch=20)
qqline(res_2_s,lty=2,col="red")

# Dividindo a base de dados 2: treinamento (80%) e teste (20%)

nrow(dados3) # 2.367 observações
n21 <- round(0.8*nrow(dados3),0) # 1.894
n22 <- round(nrow(dados3)-0.8*nrow(dados3),0) # 473

amostra_n21 <- sort(sample(nrow(dados3),n21))

dados_treino_2 <- dados3[amostra_n21,]
dados_teste_2 <- dados3[-amostra_n21,]

# Analise de variancia 1: eixo tecnologico, regiao e tipo de instituicao

modelo2 <- aov(dados_treino_2$x25_vl_exec_ha_efet ~ dados_treino_2$x7_eixo_tec +
               dados_treino_2$regiao + dados_treino_2$tipo_inst +
               dados_treino_2$x7_eixo_tec * dados_treino_2$regiao +
               dados_treino_2$x7_eixo_tec * dados_treino_2$tipo_inst +
               dados_treino_2$regiao * dados_treino_2$tipo_inst +
               dados_treino_2$x7_eixo_tec * dados_treino_2$regiao *
               dados_treino_2$tipo_inst)
summary(modelo2)

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
res_22_s <- rstudent(modelo2)
plot(res_22_s)
abline(h=0,col="red",lwd=2)
hist(res_22_s)
boxplot(res_22_s,horizontal=TRUE)
qqnorm(res_22_s,pch=20)
qqline(res_22_s,lty=2,col="red")

## Estimacao dos Coeficientes do Modelo 2

modelo2$coefficients # coeficientes da ANOVA
```

```
coef2 <- modelo2$coefficients

colnames(dados_teste_2)
dim(dados_teste_2)

dados_treino %>% select(x25_vl_exec_ha_efet,x7_eixo_tec,regiao,tipo_inst) %>%
  filter(tipo_inst=="Instituto Federal",regiao=="Norte") %>%
  group_by(x7_eixo_tec) %>% summarise(media=mean(x25_vl_exec_ha_efet))

table(dados_treino$tipo_inst)
table(dados_treino$regiao)

# Diferenca de medias

summary(modelo2)
tukey <- TukeyHSD(modelo2)
tukey$`dados_treino$x7_eixo_tec`

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
plot(TukeyHSD(modelo2))

# Base de dados de teste

names(coef2)
dados_teste_2
table(dados_teste_2$tipo_inst)

dados_teste_2$y_prev <- NA

for(i in 1:nrow(dados_teste_2)){
  if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
  dados_teste_2$regiao[i]=="Sudeste" &
     dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
     dados_teste_2$y_prev[i] <- 0.85}
  if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
  dados_teste_2$regiao[i]=="Sudeste" &
     dados_teste_2$x7_eixo_tec[i]=="CONTROLE E PROCESSOS INDUSTRIAIS"){
     dados_teste_2$y_prev[i] <- 1.30}
  if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
```

```
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 1.00}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 0.93}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
   dados_teste_2$y_prev[i] <- 1.40}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFRAESTRUTURA"){
   dados_teste_2$y_prev[i] <- 1.71}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
   dados_teste$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 0.92}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
   dados_teste_2$y_prev[i] <- 1.18}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO INDUSTRIAL"){
   dados_teste_2$y_prev[i] <- 0.77}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 0.88}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 1.71}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 1.34}
```

```
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 0.97}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
   dados_teste_2$y_prev[i] <- 1.03}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 0.95}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 0.97}
if(dados_teste_2$tipo_inst[i]=="Escola Técnica" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 0.95}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 1.01}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="CONTROLE E PROCESSOS INDUSTRIAIS"){
   dados_teste_2$y_prev[i] <- 1.02}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 0.94}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 0.98}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
```

```
   dados_teste_2$y_prev[i] <- 1.01}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 0.50}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
   dados_teste_2$y_prev[i] <- 1.04}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO INDUSTRIAL"){
   dados_teste_2$y_prev[i] <- 1.04}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 0.83}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
   dados_teste_2$regiao[i]=="Sul" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 0.50}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 1.38}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 1.42}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 1.42}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
   dados_teste_2$y_prev[i] <- 1.44}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
```

```
   dados_teste_2$x7_eixo_tec[i]=="INFRAESTRUTURA"){
   dados_teste_2$y_prev[i] <- 1.44}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 1.56}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
   dados_teste_2$y_prev[i] <- 1.08}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO INDUSTRIAL"){
   dados_teste_2$y_prev[i] <- 1.56}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 1.50}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Sudeste" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 1.33}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 2.17}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 2.18}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 1.90}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
   dados_teste_2$y_prev[i] <- 1.77}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
```

```
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFRAESTRUTURA"){
   dados_teste_2$y_prev[i] <- 2.16}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 2.18}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
   dados_teste_2$y_prev[i] <- 2.15}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 2.17}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Centro-Oeste" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 2.19}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 0.97}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="CONTROLE E PROCESSOS INDUSTRIAIS"){
   dados_teste_2$y_prev[i] <- 0.80}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 1.02}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 1.04}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
   dados_teste_2$y_prev[i] <- 0.89}
```

```
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="INFRAESTRUTURA"){
   dados_teste_2$y_prev[i] <- 1.15}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
   dados_teste_2$y_prev[i] <- 1.18}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
   dados_teste_2$y_prev[i] <- 0.58}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
   dados_teste_2$y_prev[i] <- 0.90}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Nordeste" &
   dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
   dados_teste_2$y_prev[i] <- 0.92}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Norte" &
   dados_teste_2$x7_eixo_tec[i]=="AMBIENTE E SAÚDE"){
   dados_teste_2$y_prev[i] <- 1.52}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Norte" &
   dados_teste_2$x7_eixo_tec[i]=="CONTROLE E PROCESSOS INDUSTRIAIS"){
   dados_teste_2$y_prev[i] <- 2.00}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Norte" &
   dados_teste_2$x7_eixo_tec[i]=="DESENVOLVIMENTO EDUCACIONAL E SOCIAL"){
   dados_teste_2$y_prev[i] <- 1.68}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Norte" &
   dados_teste_2$x7_eixo_tec[i]=="GESTÃO E NEGÓCIOS"){
   dados_teste_2$y_prev[i] <- 1.61}
if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
dados_teste_2$regiao[i]=="Norte" &
   dados_teste_2$x7_eixo_tec[i]=="INFORMAÇÃO E COMUNICAÇÃO"){
```

```
     dados_teste_2$y_prev[i] <- 1.62}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="INFRAESTRUTURA"){
     dados_teste_2$y_prev[i] <- 1.39}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO ALIMENTÍCIA"){
     dados_teste_2$y_prev[i] <- 1.43}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
     dados_teste_2$y_prev[i] <- 1.61}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="PRODUÇÃO CULTURAL E DESIGN"){
     dados_teste_2$y_prev[i] <- 1.61}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="RECURSOS NATURAIS"){
     dados_teste_2$y_prev[i] <- 1.37}
  if(dados_teste_2$tipo_inst[i]=="Instituto Federal" &
  dados_teste_2$regiao[i]=="Norte" &
     dados_teste_2$x7_eixo_tec[i]=="TURISMO, HOSPITALIDADE E LAZER"){
     dados_teste_2$y_prev[i] <- 1.57}
}

summary(dados_teste_2$y_prev)
table(dados_teste_2$y_prev)
dados_teste_2 %>% filter(is.na(dados_teste_2$y_prev))

dados_teste_2$erro <- dados_teste_2$x25_vl_exec_ha_efet-dados_teste_2$y_prev
summary(dados_teste_2$erro)

par(mfrow=c(2,2),mar=c(3,3,2,2),oma=c(3,3,2,2))
plot(dados_teste_2$erro)
abline(h=0,col="red",lwd=2)
hist(dados_teste_2$erro)
boxplot(dados_teste_2$erro,horizontal=TRUE)
```

```
qqnorm(dados_teste_2$erro,pch=20)
qqline(dados_teste_2$erro,lty=2,col="red")
```