



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CURSO DE GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO
PROBABILIDADE E ESTATÍSTICA (INE5405)

Leonardo de Sousa Marques
Rafael Veronezi Ribeiro
Ruan Alboni Ferreira
Thayse Estevo Teixeira

**Análise Estatística de Dados do Exame Nacional de Desempenho dos Estudantes
(ENADE)**

Florianópolis
2025

1 Introdução

O Exame Nacional de Desempenho dos Estudantes (ENADE) foi instituído em 2004 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) com o objetivo de avaliar o desempenho dos concluintes de cursos de graduação no Brasil. A cada ciclo trienal, determinados cursos em todo o país realizam a prova, que avalia o conhecimento dos estudantes sobre os conteúdos programáticos definidos nas diretrizes curriculares de cada curso (SEAVI, 2020).

O ciclo trienal do ENADE é a forma de organização do exame em períodos de três anos, de modo que diferentes cursos de graduação sejam avaliados em anos distintos, garantindo a cobertura de todas as áreas do conhecimento de forma rotativa. Cada ano do ciclo avalia grupos específicos de cursos, conforme suas áreas de atuação, conforme descrito na Tabela 1.

Tabela 1: Distribuição dos cursos avaliados por ano do ciclo trienal do ENADE

Ano	Cursos Avaliados
Ano I	Cursos de bacharelado nas áreas de Ciências Agrárias, Ciências da Saúde e áreas afins. Cursos de bacharelado nas áreas de Engenharias e Arquitetura e Urbanismo. Cursos Superiores de Tecnologia nas áreas de Ambiente e Saúde, Produção Alimentícia, Recursos Naturais, Militar e Segurança.
Ano II	Cursos de bacharelado nas áreas de Ciências Biológicas, Ciências Exatas e da Terra, Linguística, Letras e Artes e áreas afins. Cursos de licenciatura nas áreas de Ciências da Saúde, Ciências Humanas, Ciências Biológicas, Ciências Exatas e da Terra, Linguística, Letras e Artes. Cursos de bacharelado nas áreas de Ciências Humanas e Ciências da Saúde, com cursos avaliados no âmbito das licenciaturas. Cursos Superiores de Tecnologia nas áreas de Controle e Processos Industriais, Informação e Comunicação, Infraestrutura e Produção Industrial.
Ano III	Cursos de bacharelado nas áreas de Ciências Sociais Aplicadas e áreas afins. Cursos de bacharelado nas áreas de Ciências Humanas e áreas afins que não tenham cursos avaliados no âmbito das licenciaturas. Cursos Superiores de Tecnologia nas áreas de Gestão e Negócios, Apoio Escolar, Hospitalidade e Lazer, Produção Cultural e Design.

Fonte: Adaptado de (INEP, 2025)

O formato da avaliação pode variar de ano para ano, mas, considerando os dados mais recentes disponíveis que serão analisados, relativos à prova de 2023, conforme o edital, a avaliação foi composta por 40 questões: 10 de Formação Geral (FG) e 30 de Conhecimento Específico (CE) (INEP, 2023). Além disso, os estudantes também respondem ao Questionário do Estudante, que coleta informações qualitativas sobre a organização didático-pedagógica, infraestrutura, instalações físicas da instituição e oportunidades de ampliação da formação acadêmica e profissional (SEAVI, 2020).

A fim de avaliar o desempenho dos estudantes e a qualidade das instituições de ensino, o ENADE utiliza as provas e os questionários para calcular índices quantitativos e qualitativos, tais como o Conceito Enade (CE), o Conceito Preliminar de Curso (CPC), o Indicador de Diferença entre os Desempenhos Observados e Esperado (IDD) e o Índice Geral de Cursos (IGC), que serão detalhados na Seção de Materiais e Métodos, juntamente com a Categoria Administrativa e a Modalidade de Ensino dos cursos avaliados.

1.1 Objetivos

O presente relatório tem como objetivo apresentar uma análise estatística dos dados referentes aos índices de qualidade coletados pelo INEP no período de 2021 a 2023, permitindo observar características essenciais e atualizadas das instituições de ensino públicas e privadas, que se utilizam da modalidade presencial ou a distância, do estado de Santa Catarina. Para isso, serão calculadas estatísticas de tendência central e de dispersão, além de análises gráficas, buscando evidenciar padrões e possibilitar conclusões relevantes a respeito das variáveis analisadas.

1.2 Objetivos Específicos

1. Definir e classificar de forma clara e objetiva as variáveis de interesse e as motivações para sua escolha;
2. Apresentar e interpretar gráficos e tabelas que tratam da frequência das variáveis de interesse;
3. Dissertar sobre medidas de resumo aplicadas às variáveis de interesse;
4. Expor a relação entre as variáveis de interesse, duas a duas.

2 Materiais e Métodos

Nesta seção, vamos discutir as variáveis quantitativas e qualitativas selecionadas para análise de dados. Além disso, definimos a base de dados que será utilizada para a análise exploratória como os cursos participantes do ENADE no ciclo trienal de 2021 à 2023 pertencentes ao território de Santa Catarina, de forma com que alcancemos uma base de dados mais compacta cujas conclusões geradas são mais precisas dentro da sua área de alcance, Santa Catarina.

2.1 Variáveis Quantitativas

Conceito Enade Contínuo (CE) - Variável Contínua: o Conceito Enade (CE) foi uma variável escolhida por avaliar o desempenho dos cursos de uma instituição de ensino. Sua métrica é baseada nos resultados obtidos por seus estudantes no Enade em relação ao desempenho geral da respectiva área acadêmica em todo o território nacional. Para fins de cálculo, um curso tem que ter pelo menos 2 participantes com resultados válidos no exame para gerar o Conceito Enade, um valor entre 0 e 5 (INEP, 2024b).

Indicador de Diferença entre os Desempenhos Observados e Esperado Contínuo (IDD) - Variável Contínua: já o IDD foi escolhido por avaliar o quanto um curso de graduação impactou no desenvolvimento de seus estudantes. Ele leva em conta o desempenho dos estudantes no Enade, sua nota de ingresso no ensino superior, a qualidade do curso, entre outras medidas de desempenho, para então comparar o resultado aos dos demais cursos do país. Tudo isso é então compactado em uma variável entre 0 e 5.

Conceito Preliminar de Curso Contínuo (CPC) - Variável Contínua: por outro lado, o CPC foi uma escolha por avaliar a qualidade dos cursos de graduação. As métricas levadas em conta para o seu cálculo são o desempenho dos estudantes no Enade; a contribuição do curso ao desenvolvimento do aluno (IDD); a titulação e regime de trabalho do corpo docente (Censo da Educação

Superior); e a opinião dos estudantes sobre a didática, infraestrutura e oportunidades de evolução acadêmica e profissional que o curso proporciona (Questionário do Estudante) (INEP, 2024a). Essa medida é relativa ao resultado médio da área de avaliação em todo o país, gerando uma faixa contínua definida de 1 a 5.

Índice Geral de Cursos Contínuo (IGC) - Variável Contínua: o IGC é uma média ponderada das notas contínuas dos Conceitos Preliminares de Curso (CPC) de cada curso de graduação e dos Conceitos Capes dos programas de pós-graduação das Instituições de Educação Superior (IES), pelo número de matrículas nos referidos cursos. As médias são então reduzidas a uma faixa contínua de 1 a 5. Sua escolha como variável vem do seu enfoque na qualidade da instituição de ensino perante as demais ao redor do país.

2.2 Variáveis Qualitativas

Categoria Administrativa - Variável Nominal: aborda o órgão responsável por gerir a instituição de ensino de determinado curso, sendo as 7 categorias possíveis: Pública Municipal, Pública Estadual, Pública Federal, Comunitária/Confessional, Privada sem Fins Lucrativos, Privada com Fins Lucrativos e Especial. Sua escolha vem do interesse de comparar a presença e o desempenho das instituições públicas e privadas no exame.

Modalidade de Ensino - Variável Nominal : representa como as aulas do curso são administradas aos estudantes, tendo 2 modalidades possíveis: Educação Presencial e Educação a Distância. Por ser um tema recente, essa variável foi incluída para analisar o impacto dessas modalidades no desempenho e qualidade das instituições.

2.3 Bases de Dados

A fim de analisar um ciclo trienal completo do ENADE, utilizaremos dados de 2021, 2022 e 2023. Para isso, fizemos o download, no website do ENADE¹, dos arquivos XLSX referentes ao CPC, IDD e IGC. Em seguida, os arquivos foram formatados para conversão em CSV, padronizando os nomes das colunas: letras minúsculas, sem acentuação e separadas por underline. O *script* Python utilizado nesse processo está referenciado no Anexo 7.1.

Com as bases de dados devidamente formatadas, o primeiro passo foi uni-las em um único conjunto, contemplando os dados dos três anos. Em seguida, ao analisar a frequência das variáveis qualitativas – categoria administrativa (Figura 1a) e modalidade de ensino (Figura 1b) –, constatou-se que o dataset totaliza 19.703 linhas, relativas à cada curso avaliado, com instituições de ensino aparecendo mais de uma vez. Para essa análise e geração dos gráficos, utilizamos o *script* R detalhado no Anexo 7.2. Diante do grande volume de dados, optou-se por filtrar apenas as informações referentes ao estado de Santa Catarina, de modo a viabilizar uma análise mais detalhada.

¹<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>

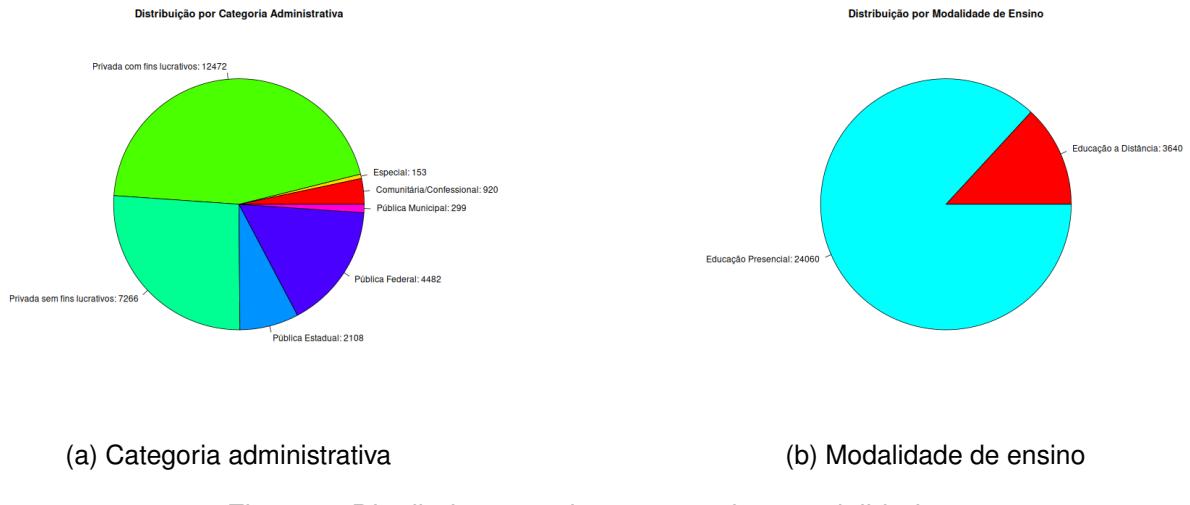


Figura 1: Distribuição geral por categoria e modalidade

3 Resultados e Discussões

Ao longo desta seção serão apresentadas e discutidas as tabelas de frequência, os gráficos, os cálculos estatísticos e a relação das variáveis quantitativas e qualitativas descritas na seção de Materiais e Métodos.

3.1 Tabelas de Frequência e Gráficos

3.1.1 Variáveis Qualitativas

Para a geração das tabelas de frequência das variáveis quantitativas, basta mapear o número de vezes que cada valor aparece no dataset. Nesse sentido, a Tabela 2 e a Tabela 3 apresentam, respectivamente, os dados referentes à modalidade de ensino e à categoria administrativa.

Além disso, destaca-se que a tabela de frequência da modalidade de ensino foi construída a partir da base acumulada do CPC — considerando cursos avaliados nos anos de 2021, 2022 e 2023. Já a tabela de frequência da categoria administrativa foi elaborada a partir da base do IGC, em que cada instituição aparece apenas uma vez — naturalmente, ao juntar as bases de dados, uma mesma instituição poderia aparecer uma, duas ou até três vezes.

Portanto, a partir das tabelas, conclui-se que, em Santa Catarina, há 1.349 cursos de ensino superior avaliados, distribuídos em 116 instituições de ensino.

Quando observamos mais diretamente o que a frequência dessas variáveis nos mostra, podemos observar que por volta de 20% dos cursos avaliados são da modalidade Educação a Distância (EaD), o que nos fornece algumas conclusões. A primeira sendo que essa tecnologia foi realmente adotada para cursos de ensino superior em Santa Catarina, ganhando espaço em relação à educação presencial. Já para a segunda conclusão obtida, temos o fato de que o ensino EaD não chegou a substituir o presencial, estando longe disso, dada a predominância do segundo com 80% dos cursos de Santa Catarina.

Distribuição por Modalidade de Ensino

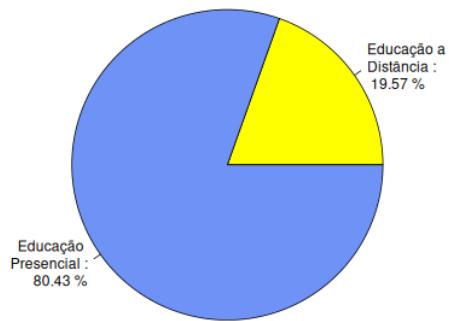


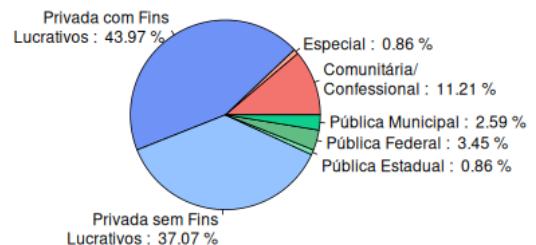
Tabela 2: Frequência por Modalidade de Ensino por Curso (SC)

Modalidade de Ensino	Frequência
Educação a Distância	264
Educação Presencial	1085
Total	1349

Tabela 3: Frequência por Categoria Administrativa por Curso (SC)

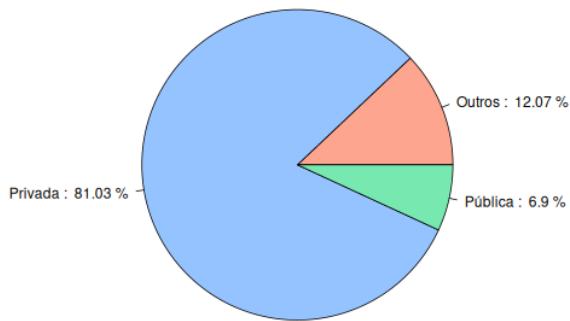
Categoria Administrativa	Frequência
Comunitária/Confessional	13
Especial	1
Privada com fins lucrativos	51
Privada sem fins lucrativos	43
Pública Estadual	1
Pública Federal	4
Pública Municipal	3
Total	116

Distribuição por Categoria Administrativa



Tendo em vista a Categoria Administrativa, é possível perceber uma grande predominância de instituições privadas, enquanto públicas, comunitárias e categorias especiais representam menos de 20%. O gráfico abaixo representa essa relação de forma mais clara. Por nosso foco ser na comparação entre instituições públicas e privadas, o gráfico dá destaque a essas categorias, deixando as demais compactadas em Outros. Com esses dados é possível perceber que a presença de instituições privadas em Santa Catarina supera fortemente a de públicas, representando mais de 80% enquanto a segunda constitui menos de 7%. O restante dos gráficos de distribuição para as variáveis qualitativas podem ser observados no Anexo 6.

Distribuição por Categoria Administrativa - Agrupada



3.1.2 Variáveis Quantitativas

Para a criação das tabelas de frequência das variáveis quantitativas, utilizou-se o método de Sturges para se determinar o número de classes (k) em que os intervalos seriam distribuídos. Para isso, utilizamos a função Sturges da classe nclass da linguagem R, que faz uso da fórmula $k = 1 + 1.322\log_{10}N$, em que N é o número total de elementos. Com isso, fez-se uso do método cut do R para separar os dados em k intervalos. Os detalhes do algoritmo podem ser observados no Anexo 7.3.

Nas Tabelas 4, 5, 6 e 7, respectivamente, temos os valores de frequência divididos por intervalo para o Conceito Enade, IDD, CPC e IGC. Observa-se que o limite inferior do primeiro intervalo pode aparecer como um valor negativo, embora o valor mínimo real seja 0. Isso ocorre porque o algoritmo de definição das classes calcula os intervalos de forma automática a partir da amplitude total dividida pelo número de classes, o que pode gerar uma extração ligeiramente abaixo do valor mínimo observado. Esse ajuste não compromete a análise, pois todos os dados efetivos permanecem dentro da faixa correta.

Em vista dos organizados, podemos tirar algumas conclusões. O Conceito Enade, como mostrado abaixo, tem uma distribuição bastante regular, com bastantes notas concentradas nos valores médios do centro (entre 2 e 3), gradativamente diminuindo para os extremos, mas há algumas divergências. Uma delas é a concentração levemente maior de notas menores que o centro (2,5), enquanto a outra é o salto observado no 0, que acumula uma quantia considerável de conceitos. Isso tudo indica que o Conceito Enade está inclinado para valores menores que o valor central.

Quanto ao IDD, ele também possui uma distribuição relativamente regular em relação ao valor médio (2,5), mas diferente do Conceito Enade, sua inclinação é para os valores acima do centro, com valores próximos do 5 tendo uma leve tendência de crescimento, indicando uma discrepância mais significativa entre o desempenho obtido e o esperado dos cursos avaliados.

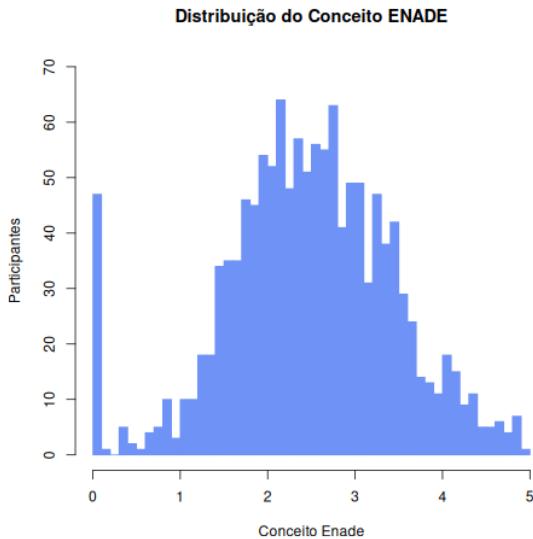
O CPC foi mais um parâmetro que apresentou um comportamento regular em relação ao valor médio (3), sem possuir discrepâncias em nenhum dos extremos, indicando que os cursos de Santa Catarina apresentaram um desempenho próximo à média nacional.

Por fim, o IGC apresenta uma grande concentração de valores centrais, com as extremidades tendo pouquíssimos valores, indicando que as instituições de ensino de Santa Catarina são avalia-

das como tendo uma qualidade média em relação ao restante do país. O restante dos gráficos de distribuição para as variáveis podem ser observados no Anexo 6.

Tabela 4: Frequência por Conceito Enade Contínuo por Curso (SC)

Limites	Frequência
[−0.00497, 0.414)	54
[0.414, 0.829)	15
[0.829, 1.24)	35
[1.24, 1.66)	119
[1.66, 2.07)	198
[2.07, 2.49)	225
[2.49, 2.9)	224
[2.9, 3.31)	184
[3.31, 3.73)	132
[3.73, 4.14)	56
[4.14, 4.56)	35
[4.56, 4.98]	21



Distribuição do IDD

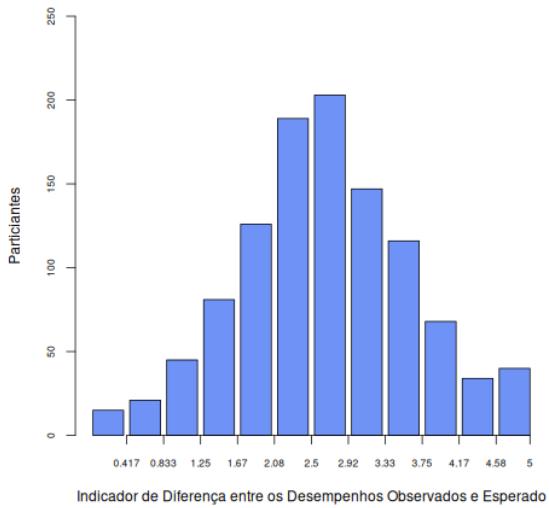


Tabela 5: Frequência por IDD Contínuo (SC)

Limites	Frequência
[−0.005, 0.417)	15
[0.417, 0.833)	21
[0.833, 1.25)	45
[1.25, 1.67)	81
[1.67, 2.08)	126
[2.08, 2.5)	189
[2.5, 2.92)	203
[2.92, 3.33)	147
[3.33, 3.75)	116
[3.75, 4.17)	68
[4.17, 4.58)	34
[4.58, 5]	40

Tabela 6: Frequência por Conceito Preliminar do Curso Contínuo (SC)

Limites	Frequência
[0.973, 1.28)	5
[1.28, 1.58)	16
[1.58, 1.88)	45
[1.88, 2.19)	104
[2.19, 2.49)	166
[2.49, 2.79)	233
[2.79, 3.09)	229
[3.09, 3.39)	169
[3.39, 3.7)	148
[3.7, 4)	78
[4, 4.3)	42
[4.3, 4.61]	16

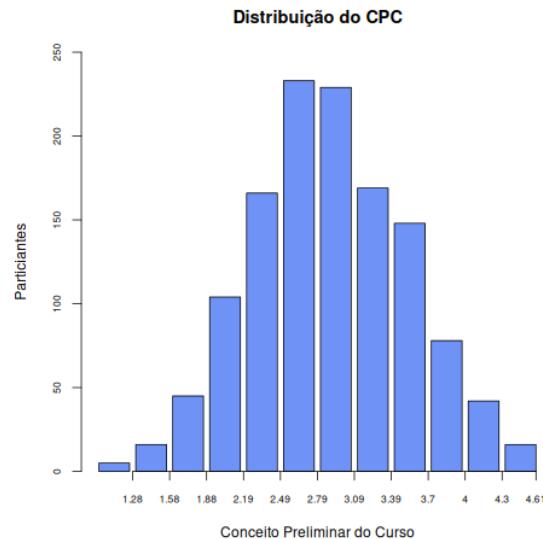
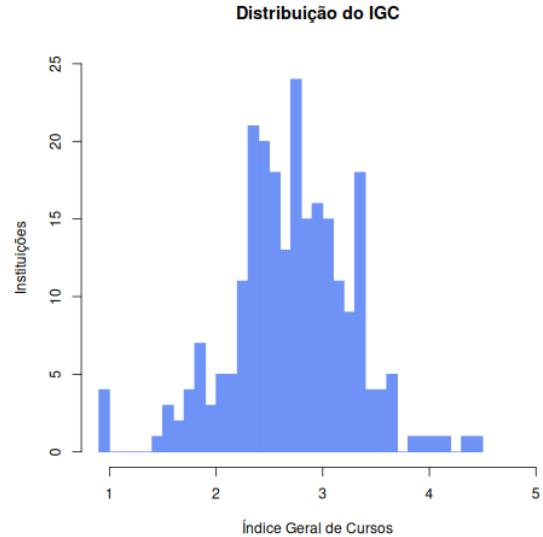


Tabela 7: Distribuição de Frequências do Índice Geral de Cursos (IGC) – SC

Limites	Frequência
[0.923, 1.31)	4
[1.31, 1.7)	6
[1.7, 2.09)	19
[2.09, 2.48)	54
[2.48, 2.86)	70
[2.86, 3.25)	47
[3.25, 3.64)	37
[3.64, 4.03)	3
[4.03, 4.42]	4



3.2 Medidas de Tendência Central e Dispersão

Nesta seção, calculamos, para as variáveis quantitativas, as medidas de tendência central de média e mediana. Para medidas de dispersão, calculamos o desvio padrão, variância, coeficiente de variação (CV), mínimo, máximo e amplitude. Os resultados podem ser observados na Tabela 8.

Como é possível perceber, esses resultados complementam as conclusões obtidas na seção 3.1, com a média e mediana muito próximas indicando estabilidade nos quatro índices. Além disso, esses valores também se aproximam do valor central, como explicado anteriormente, com o CPC e o IGC contendo as menores variâncias devido a ausência de picos em suas extremidades.

Tabela 8: Estatísticas Descritivas dos Indicadores Contínuos (SC)

Indicador	Média	Mediana	Desvio Padrão	Variância	CV (%)	Mínimo	Máximo	Amplitude
Conceito ENADE (Contínuo)	2.478	2.488	0.971	0.943	39.196	0	4.972	4.972
IDD (Contínuo)	2.667	2.631	0.983	0.965	36.838	0	5	5
CPC (Contínuo)	2.885	2.857	0.636	0.404	22.029	0.977	4.603	3.626
IGC (Contínuo)	2.703	2.726	0.565	0.319	20.909	0.926	4.416	3.49

4 Relações Entre Variáveis

Nesta seção são apresentadas as relações entre as variáveis qualitativas e quantitativas escolhidas. Optou-se pela representação em boxplot para relacionar as variáveis qualitativas com os indicadores quantitativos do Conceito Enade, IDD e CPC, sendo exemplificado a relação das variáveis qualitativas e o Conceito Enade.

Na relação de Categoria Administrativa com o Conceito Enade, observa-se valores de mediana mais altos entre os cursos nas categorias de instituições públicas(Federal, Estadual e Municipal) enquanto as categorias privadas (com e sem fim lucrativo) apresentam medianas menores entre os cursos.

Em relação a comparação entre a modalidade de ensino e Conceito Enade, cursos de modalidade presencial mostram-se com mediana maior quando comparados com a modalidade Ensino à Distância, o que pode-se interpretar como um melhor rendimento nos cursos dessa modalidade.

Para a relação entre IGC e Categoria administrativa, optou-se por uma representação em tabela de contingência, descrito na Tabela 9 devido ao baixo número de dados associados especificamente às instituições classificadas como categoria administrativa “Especial” com apenas uma instituição representante na relação, o que não justificaria um boxplot específico. É importante destacar que, como os dados abrangem as entradas de 3 anos, algumas instituições aparecem repetidas com categorias administrativas diferentes entre os anos, o que se explica por uma mudança na administração da instituição. Nesses casos foi adotada a convenção de tratar as instituições repetidas mas com categorias diferentes entre os anos como entradas independentes tomando o valor único de seus IGCs. Já instituições de mesmo nome e categorias iguais entre os anos tiveram a sua faixa de IGC apresentada como uma média entre as entradas nos diferentes anos.

Na tabela de contingência, observa-se claramente a distribuição dos dados com uma maior representatividade das categorias de “Privada com fins lucrativos” e “Privada sem fins lucrativos”, ambas com uma grande parcela de instituições situando-se entre as faixas de IGC 2.20 e 3.05. Outro ponto relevante a ser apontado é quanto aos dados referentes a instituições públicas federais, que, embora com pouca representatividade nos dados (4 instituições), é a única categoria a apresentar uma instituição na faixa mais alta, entre 3.9 a 4.32.

A seguir, são apresentados os gráficos e a tabela de contingência gerados e analisados. O restante dos gráficos podem ser observados no Anexo 6.

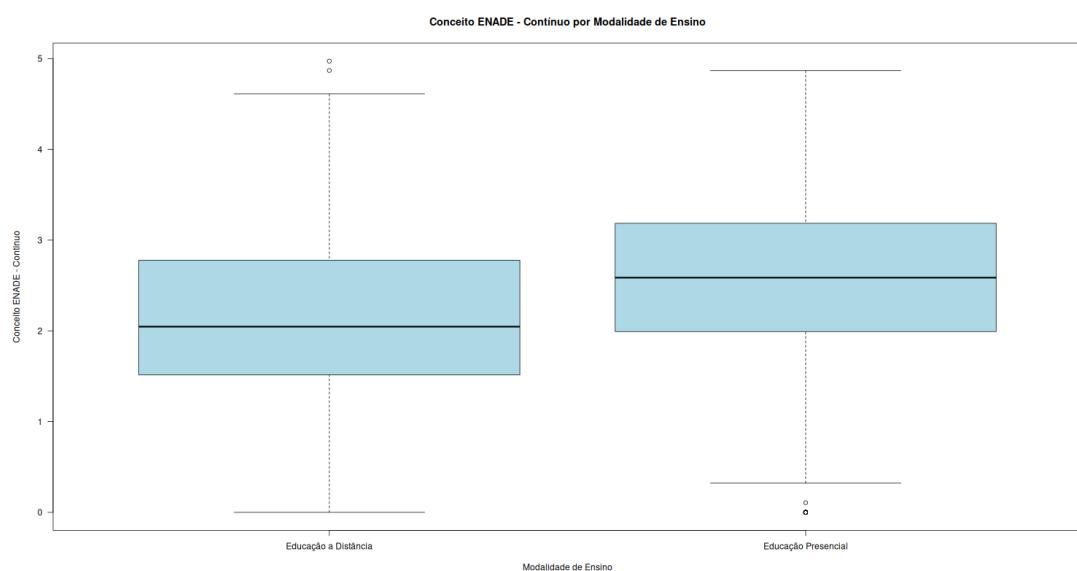
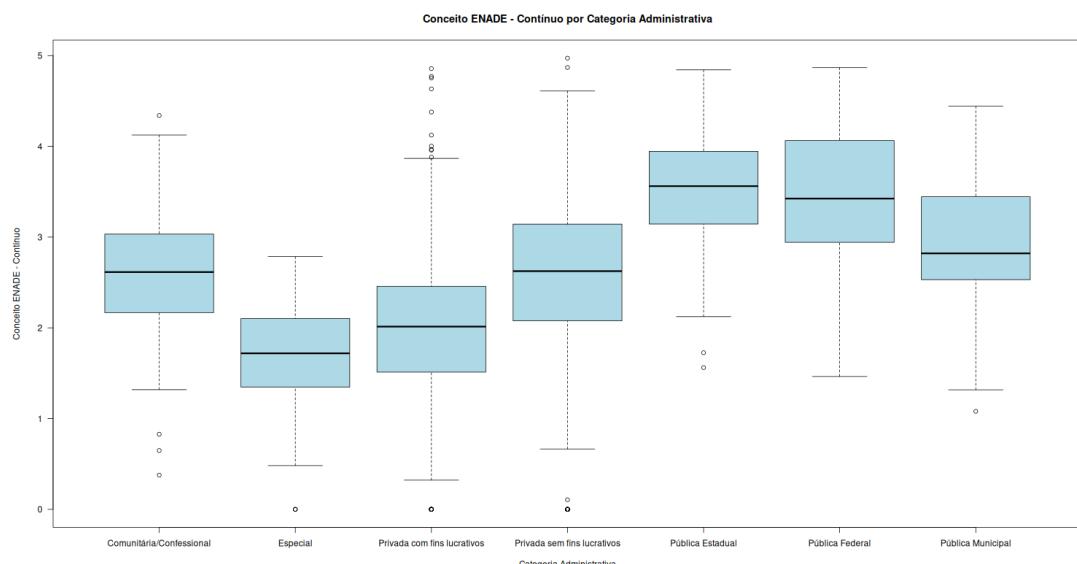


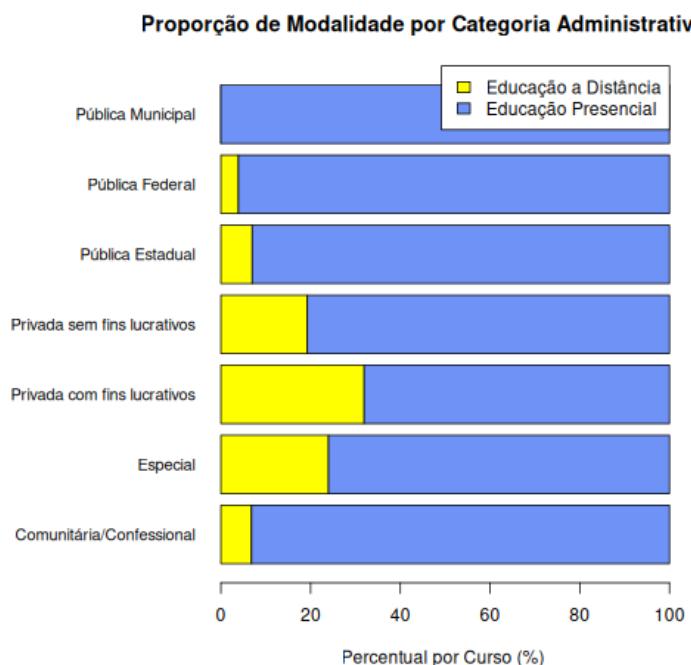
Tabela 9: Frequência por Categoria Administrativa e Intervalo de IGC (SC)

Categoría	[0.923,1.35)	[1.35,1.77)	[1.77,2.2)	[2.2,2.62)	[2.62,3.05)	[3.05,3.47)	[3.47,3.9)	[3.9,4.32]
Comunitária/Confessional	0	0	0	2	5	4	2	0
Especial	0	0	0	1	0	0	0	0
Privada com fins lucrativos	1	2	6	20	16	5	0	0
Privada sem fins lucrativos	0	1	6	12	13	10	1	0
Pública Estadual	0	0	0	0	0	0	1	0
Pública Federal	0	0	0	0	0	2	1	1
Pública Municipal	0	0	0	1	1	1	0	0

Por fim, ao analisar a relação entre as variáveis qualitativas, percebe-se que a modalidade Educação a Distância (EaD) tem ganhado espaço considerável, sobretudo nas instituições privadas com fins lucrativos, o que indica uma adaptação dessas instituições às demandas de flexibilidade e alcance ampliado de estudantes. Esse crescimento da EaD reflete a tendência nacional de diversificação das modalidades de ensino superior e o investimento em tecnologia educacional.

Todavia, ao observar o gráfico abaixo, nota-se que a maior parte das universidades ainda mantém a modalidade presencial como predominante. Esse predomínio evidencia que, apesar da expansão do EaD, a educação presencial continua sendo a principal estratégia para garantir qualidade

acadêmica, interação direta entre alunos e professores e experiência formativa completa. Portanto, embora a EaD esteja em ascensão, ela ainda não substitui o ensino presencial, servindo mais como complemento do que como substituto.



5 Considerações Finais

A análise dos dados do ENADE para Santa Catarina revelou padrões consistentes entre modalidade de ensino, categoria administrativa e os indicadores de desempenho acadêmico. Observou-se que a grande maioria dos cursos é presencial, representando cerca de 80% do total, enquanto cursos na modalidade Educação a Distância correspondem a apenas 20%. Isso indica que, embora o EaD esteja presente, não substituiu o ensino presencial no estado.

No que tange à categoria administrativa, as instituições privadas com e sem fins lucrativos predominam, correspondendo a mais de 80% do total, enquanto públicas, comunitárias e especiais representam menos de 20%. A distribuição do IGC evidencia que a maioria das instituições privadas concentra-se em faixas intermediárias (2.2 a 3.05), enquanto algumas públicas federais alcançam as faixas mais altas (3.9 a 4.32), indicando que, apesar de sua menor quantidade, instituições públicas apresentam melhor desempenho médio.

As variáveis contínuas (Conceito ENADE, IDD, CPC e IGC) mostram comportamento relativamente regular, com tendência central próxima à média, embora o Conceito Enade apresente leve inclinação para valores inferiores. O CPC demonstra desempenho consistente em torno da média, e o IGC indica que a maior parte das instituições possui qualidade média em relação ao país. Boxplots e tabelas de contingência reforçam a influência tanto da modalidade de ensino quanto da categoria administrativa nos resultados, evidenciando desigualdades e oportunidades de melhoria.

Esses achados sugerem que políticas e estratégias voltadas à melhoria da qualidade acadêmica devem considerar tanto a distribuição da modalidade de ensino quanto as diferenças entre categorias administrativas, especialmente para fortalecer o desempenho das instituições privadas e ampliar a equidade entre públicas e privadas.

Referências

- INDICADORES DA EDUCAÇÃO SUPERIOR (CGGI/DAES/INEP), COORDENAÇÃO-GERAL DE. **Nota Técnica CEI/CGGI/DAES nº 19/2024 – Metodologia utilizada no cálculo do CPC referente ao ano de 2023.** [S.I.], 2024. Acessado em: 18 set. 2025. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_19_2024_cei_cggi_daes_inep_metodologia_utilizada_no_calculo_do_cpc_referente_ao_ano_de_2023.pdf>.
- _____. **Nota Técnica CEI/CGGI/DAES nº 6/2024 – Metodologia utilizada no cálculo do Conceito Enade referente ao ano de 2023.** [S.I.], 2024. Acessado em: 18 set. 2025. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2024/nota_tecnica_6_metodologia_calculo_conceito_enade_2023.pdf>.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Áreas de conhecimento e eixos tecnológicos do Enade.** [S.I.: s.n.], 2025. Acessado em: 30 ago. 2025. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>>.
- _____. **Edital nº 37, de 25 de maio de 2023 – Exame Nacional de Desempenho dos Estudantes (Enade) 2023.** [S.I.: s.n.], mai. 2023. Ministério da Educação. Acessado em: 30 ago. 2025. Disponível em: <<https://www.in.gov.br/en/web/dou/-/edital-n-37-de-25-de-maio-de-2023-486214440>>.
- SECRETARIA ESPECIAL DE AVALIAÇÃO INSTITUCIONAL. **CONCEITO ENADE (CE) E CONCEITO PRELIMINAR DE CURSO (CPC) DOS CURSOS DE GRADUAÇÃO DO INMA.** [S.I.]: Universidade Federal de Mato Grosso do Sul (UFMS), 2020. Acessado em: 30 ago. 2025. Disponível em: <<https://seavi.ufms.br/files/2020/05/Relat%C3%B3rio-1-INMA-2020.pdf>>.

6 Anexos - Gráficos

Distribuição por Categoria Administrativa - Outros

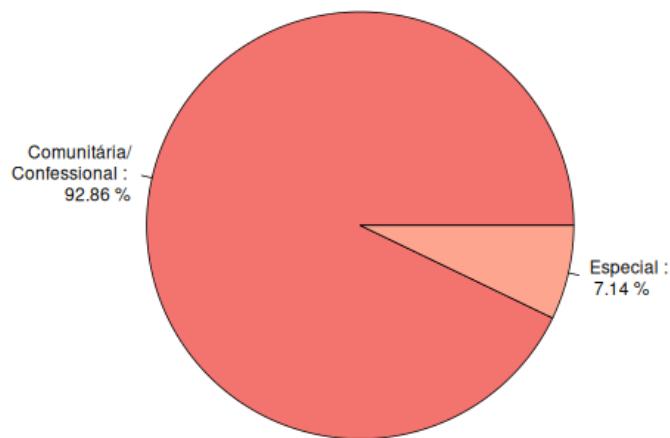


Figura 2: Pizza: Categoria Administrativa SC Outros

Distribuição por Categoria Administrativa - Privadas

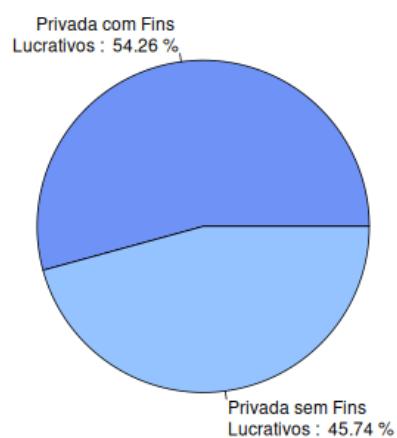


Figura 3: Pizza: Categoria Administrativa SC Privada

Distribuição por Categoria Administrativa - Públia

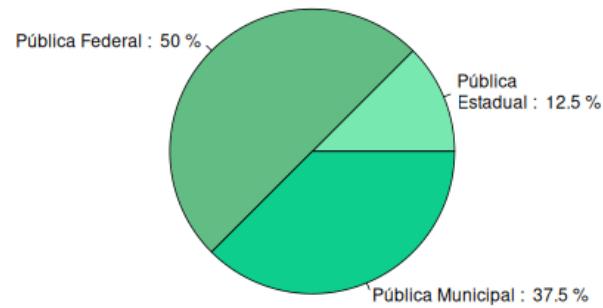


Figura 4: Pizza: Categoria Administrativa SC Públia

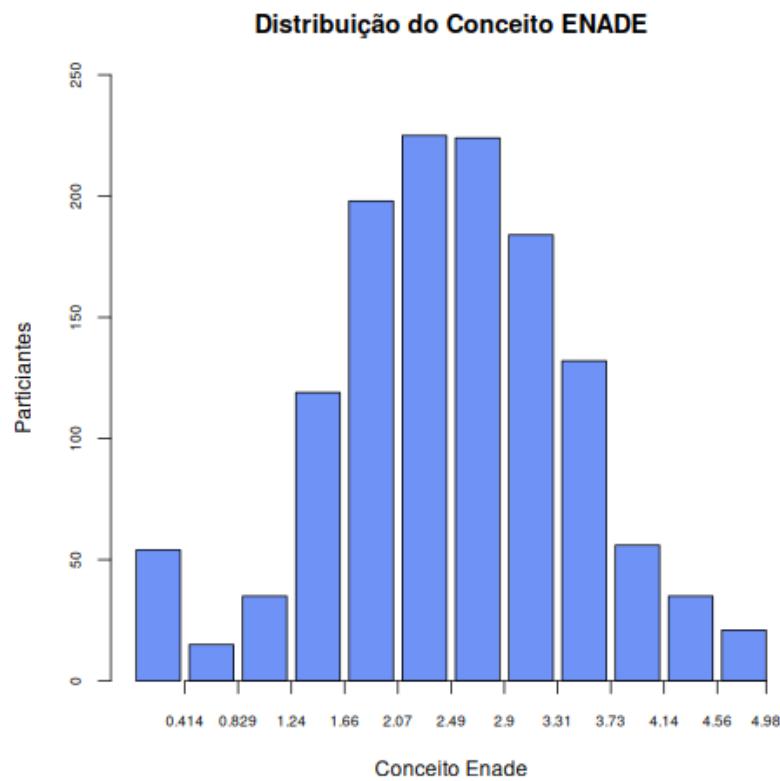


Figura 5: Barras: Conceito ENADE SC

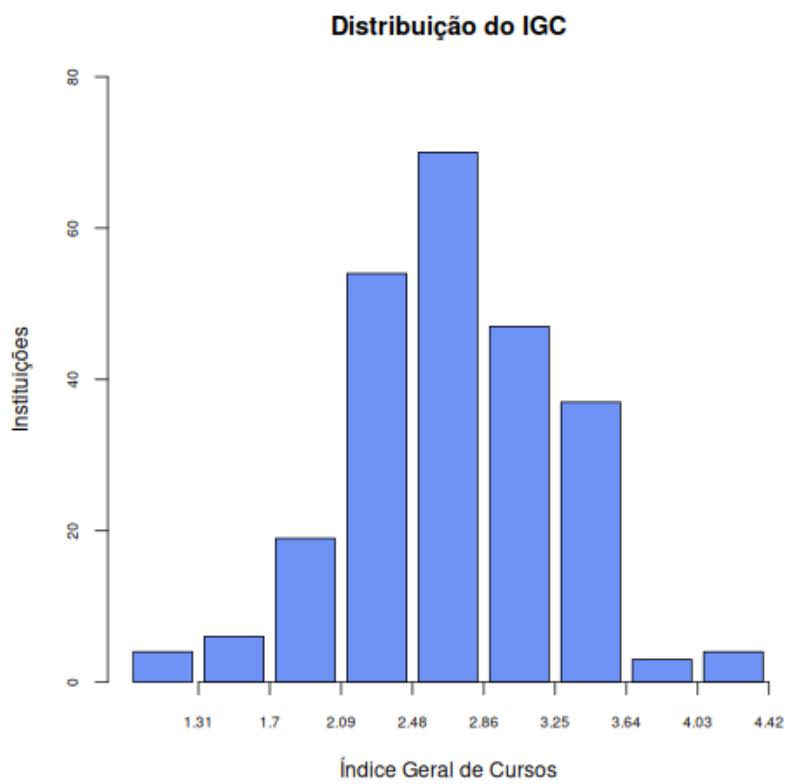


Figura 6: Barras: IGC SC

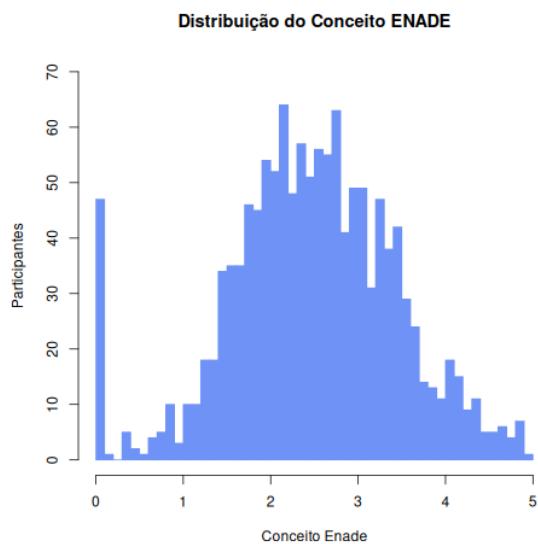


Figura 7: Histograma: Conceito ENADE SC

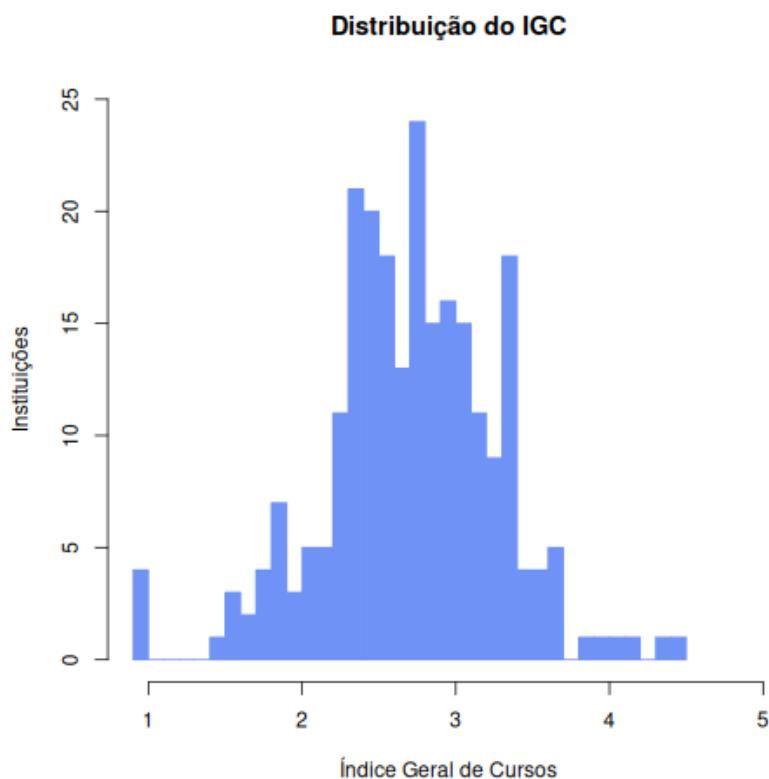


Figura 8: Histograma: IGC SC

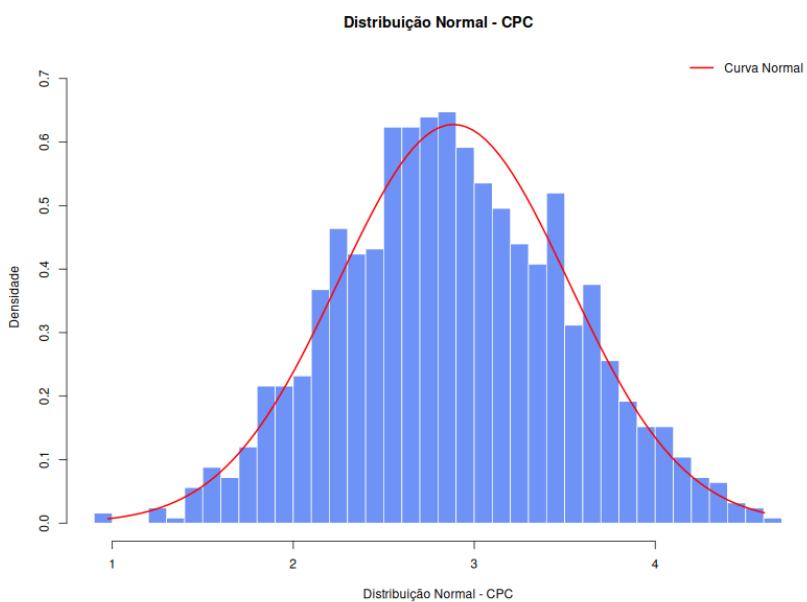


Figura 9: Curva Normal: CPC SC

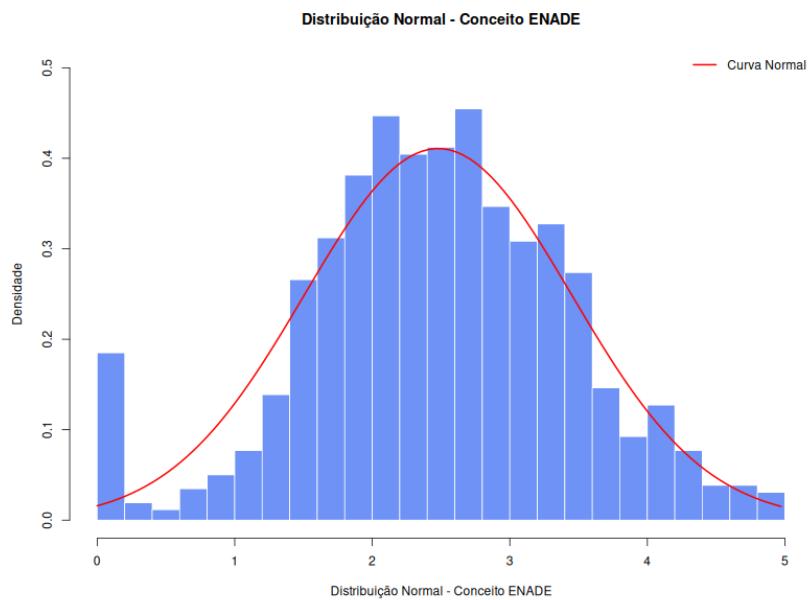


Figura 10: Curva Normal: ENADE SC

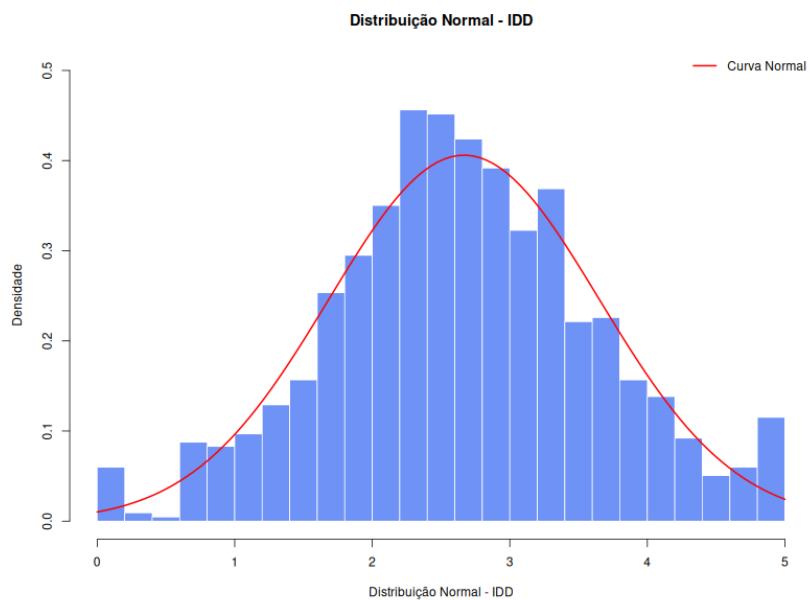


Figura 11: Curva Normal: IDD SC

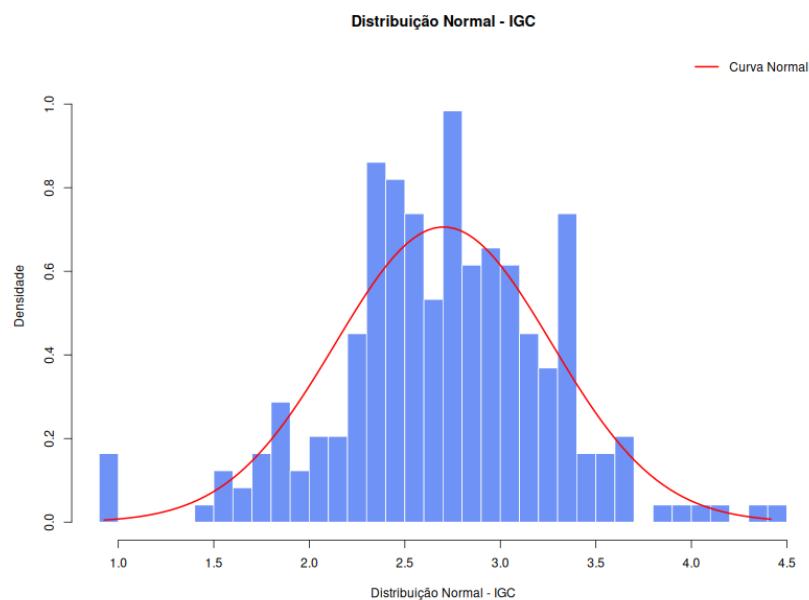


Figura 12: Curva Normal: IGC SC

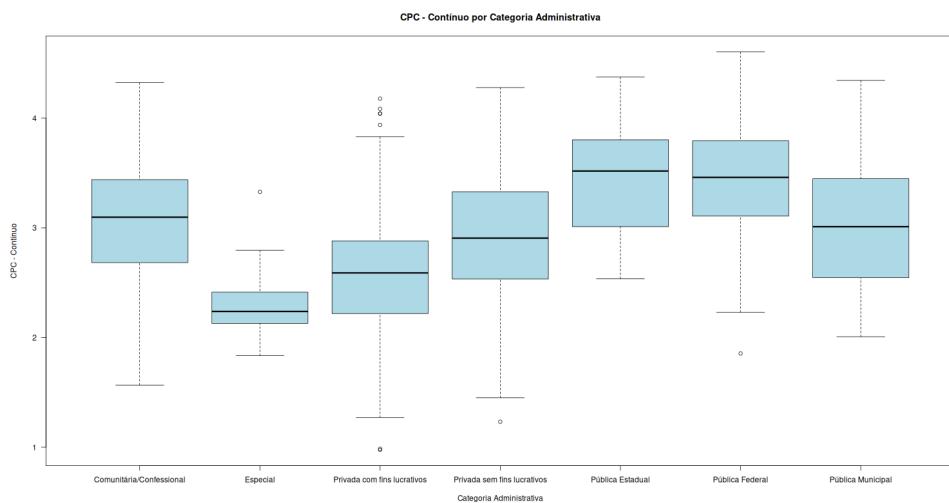


Figura 13: Boxplot: CPC por Categoria Administrativa

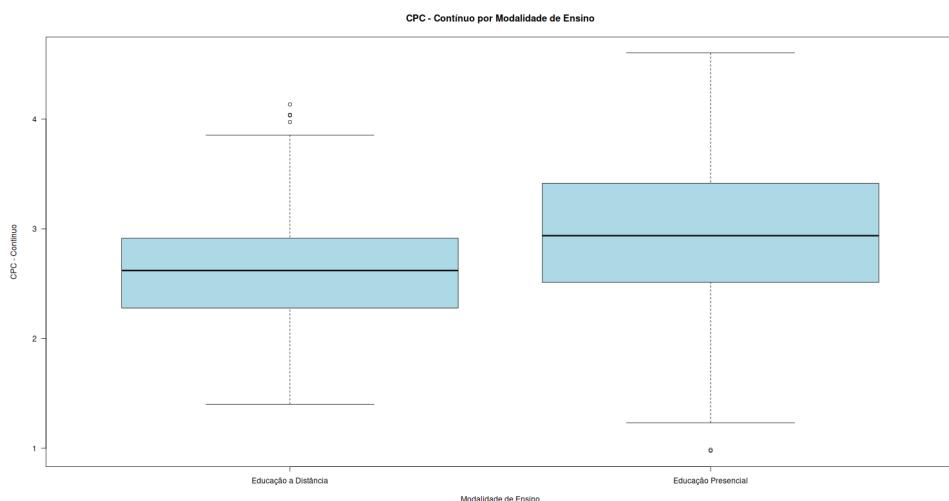


Figura 14: Boxplot: CPC por Modalidade de Ensino

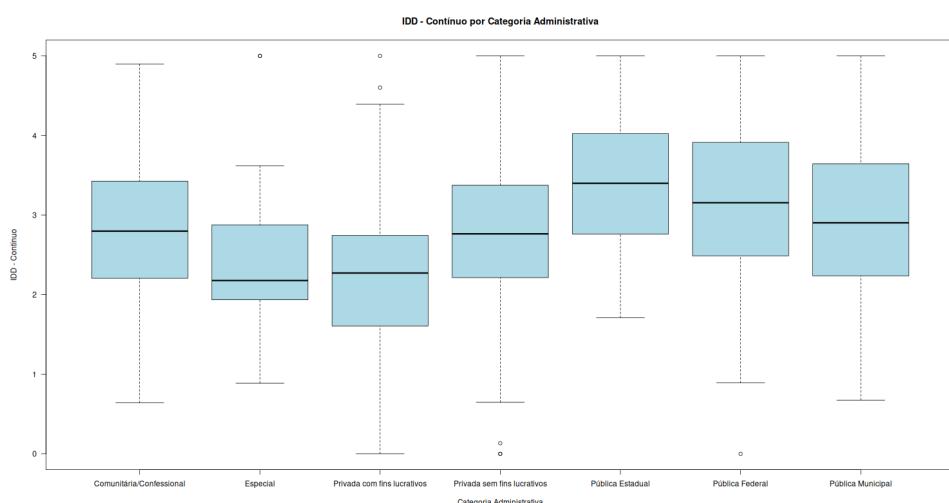


Figura 15: Boxplot: IDD por Categoria Administrativa

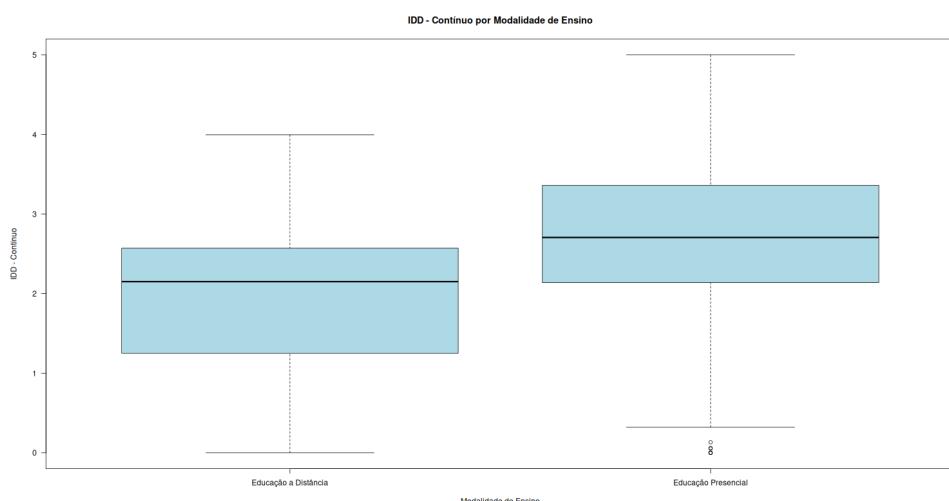


Figura 16: Boxplot: IDD por Modalidade de Ensino

7 Anexos - Códigos Fonte

Todos os códigos, bases de dados, gráficos e tabelas podem ser observados em detalhes no repositório: <https://github.com/leonardosm14/ENADE-Estatisticas>

Listing 7.1: Anexo A: Código Python

```
# Script para converter dados em xlsx para csv e formatar nomes das colunas
import pandas as pd
from os import mkdir
from pathlib import Path
from unicodedata import normalize
from unidecode import unidecode

xlsx_folder = Path("raw_data/")
csv_folder = Path("csv_data/")
if not csv_folder.exists():
    mkdir(csv_folder)

for xlsx_file in xlsx_folder.iterdir():
    print(f"Convertendo {xlsx_file.name} para CSV e formatando colunas.")
    df = pd.read_excel(io=xlsx_file)
    df.columns = [
        unidecode(
            normalize("NFKD", col.strip().replace(' - ', ' ').replace(' ', '_').
            rstrip('*').lower()))
        ) for col in df.columns
    ]
    df.to_csv(path_or_buf=f"{csv_folder}/{xlsx_file.stem}.csv", index=False)
```

Listing 7.2: Anexo B: Código R - Geral

```
# Diretório da base de dados - Talvez precise alterar, dependendo de onde o
# repositório estiver clonado.
setwd("~/Documentos/UFSC/ENADE-Estatisticas")

# Iremos utilizar a seguinte convenção:
# dataA <- ano de 2021 -> então, teremos dataA_CPC, dataA_IDD e dataA_IGC
# dataB <- ano de 2022 -> então, teremos dataB_CPC, dataB_IDD e dataB_IGC
# dataC <- ano de 2023 -> então, teremos dataC_CPC, dataC_IDD e dataC_IGC

# Carregando as bases de dados
dataA_CPC <- read.csv("data/csv_data/CPC_2021.csv")
dataA_IDD <- read.csv("data/csv_data/IDD_2021.csv")
dataA_IGC <- read.csv("data/csv_data/IGC_2021.csv")
```

```

dataB_CPC <- read.csv("data/csv_data/CPC_2022.csv")
dataB_IDD <- read.csv("data/csv_data/IDD_2022.csv")
dataB_IGC <- read.csv("data/csv_data/IGC_2022.csv")

dataC_CPC <- read.csv("data/csv_data/CPC_2023.csv")
dataC_IDD <- read.csv("data/csv_data/IDD_2023.csv")
dataC_IGC <- read.csv("data/csv_data/IGC_2023.csv")

# Função para padronizar colunas com base em um padrão
padroniza_colunas <- function(df, col_padrao) {
  for(col in col_padrao) {
    if(!col %in% colnames(df)) {
      df[[col]] <- NA # Cria coluna ausente com NA
    }
  }
  # Ordena as colunas para ficarem na mesma ordem do padrão
  df <- df[, col_padrao, drop = FALSE]
  return(df)
}

# Dados combinados usando 2023 como padrão
cols_CPC <- colnames(dataC_CPC)
cols_IDD <- colnames(dataC_IDD)
cols_IGC <- colnames(dataC_IGC)

dataA_CPC <- padroniza_colunas(dataA_CPC, cols_CPC)
dataB_CPC <- padroniza_colunas(dataB_CPC, cols_CPC)
# dataC_CPC já está no padrão, não precisa alterar

dataA_IDD <- padroniza_colunas(dataA_IDD, cols_IDD)
dataB_IDD <- padroniza_colunas(dataB_IDD, cols_IDD)

dataA_IGC <- padroniza_colunas(dataA_IGC, cols_IGC)
dataB_IGC <- padroniza_colunas(dataB_IGC, cols_IGC)
# dataC_IGC já está no padrão

# Combina os datasets
data_CPC <- rbind(dataA_CPC, dataB_CPC, dataC_CPC)
data_IDD <- rbind(dataA_IDD, dataB_IDD, dataC_IDD)
data_IGC <- rbind(dataA_IGC, dataB_IGC, dataC_IGC)
colnames(data_CPC) <- gsub("^X_", "", colnames(data_CPC))
colnames(data_IDD) <- gsub("^X_", "", colnames(data_IDD))

```

```

colnames(data_IGC) <- gsub("^X_", "", colnames(data_IGC))

# Para a análise exploratória, vamos utilizar como variáveis:
# Qualitativas: Categoria Administrativa e Modalidade de Ensino;
# Quantitativas: CPC, IDD e IGC -- IDD está presente na base de CPC.

# Tabelas de Frequência - Ignorando variáveis nulas
tab_cat_admin <- table(data_CPC$categoria_administrativa)
tab_cat_admin <- tab_cat_admin[names(tab_cat_admin) != ""]

tab_modalidade <- table(data_CPC$modalidade_de_ensino)
tab_modalidade <- tab_modalidade[names(tab_modalidade) != ""]

write.csv(tab_cat_admin, "src/tabelas/
    tabela_frequencia_categoria_administrativa.csv", row.names = FALSE)
write.csv(tab_modalidade, "src/tabelas/tabela_frequencia_modalidade_ensino.csv
", row.names = FALSE)

# Gráfico de pizza para Categoria Administrativa
png("src/gráficos/pizza_categoria_administrativa.png", width = 800, height =
600)
pie(tab_cat_admin,
    main = "Distribuição por Categoria Administrativa",
    col = rainbow(length(tab_cat_admin)), # cores diferentes para cada fatia
    labels = paste(names(tab_cat_admin), tab_cat_admin, sep = ": ")) # nomes +
    valores
dev.off()

# Gráfico de pizza para Modalidade de Ensino
png("src/gráficos/pizza_modalidade_ensino.png", width = 800, height = 600)
pie(tab_modalidade,
    main = "Distribuição por Modalidade de Ensino",
    col = rainbow(length(tab_modalidade)),
    labels = paste(names(tab_modalidade), tab_modalidade, sep = ": "))
dev.off()

```

Listing 7.3: Anexo C: Código R - SC

```

# Diretório da base de dados - Talvez precise alterar, dependendo de onde o
# repositório estiver clonado.
setwd("~/Documentos/ENADE-Estísticas")

```

```

# Vamos reutilizar os dados agrupados em script_geral.r
source(file="src/script_geral.r")
source(file="src/utils.r")

# Dados filtrados para Santa Catarina relativos a CPC - Instituições se repetem
# para cada curso
data_CPC_SC <- subset(data_CPC, sigla_da_uf == "SC")

# Dados filtrados para Santa Catarina relativos a IDD - Instituições se repetem
# para cada curso
data_IDD_SC <- subset(data_IDD, sigla_da_uf == "SC")

# Dados filtrados para Santa Catarina relativos a IGC - Instituições aparecem
# uma única vez
data_IGC_SC <- subset(data_IGC, sigla_da_uf == "SC")

# ----- TABELAS DE FREQUÊNCIA -----
#
# Modalidade de Ensino por Curso
modalidade_de_ensino_curso <- table(data_CPC_SC$modalidade_de_ensino)
write.csv(modalidade_de_ensino_curso, "src/tabelas/SC/modalidade_ensino_curso.
csv", row.names = FALSE)

# Categoria Administrativa por Instituição - Aparições únicas
instituicoes_unicas <- unique(data_IGC_SC[, c("nome_da_ies", "codigo_da_ies", "categoria_administrativa")])
categoria_administrativa <- table(instituicoes_unicas$categoria_administrativa)
write.csv(as.data.frame(categoria_administrativa), "src/tabelas/SC/
categoria_administrativa_instituicao.csv", row.names = FALSE)

# Conceito ENADE - Contínuo
enade_continuo <- data_CPC_SC$conceito_enade_.continuo.
freq_enade_continuo <- salvar_frequencia_continuo(valores = enade_continuo,
arquivo_saida = "src/tabelas/SC/conceito_enade_continuo.csv")

# Indicador de Diferença entre os Desempenhos Observados e Esperado (IDD) -
Continuo
idd_continuo <- data_IDD_SC$idd_.continuo.
freq_idd_continuo <- salvar_frequencia_continuo(valores = idd_continuo,
arquivo_saida = "src/tabelas/SC/idd_continuo.csv")

# Conceito Preliminar do Curso - CPC - Contínuo
cpc_continuo <- data_CPC_SC$cpc_.continuo.

```

```

freq_cpc_continuo <- salvar_frequencia_continuo(valores = cpc_continuo,
  arquivo_saida = "src/tabelas/SC/conceito_preliminar_curso_continuo.csv")

igc_continuo <- data_IGC_SC$igc_.continuo.
freq_igc_continuo <- salvar_frequencia_continuo(valores = igc_continuo,
  arquivo_saida = "src/tabelas/SC/igc_continuo.csv")

# ----- GRÁFICOS -----
# Modalidade de Ensino por Curso
png("src/gráficos/pizza_modalidade_de_ensino_SC.png")
pie(modalidade_de_ensino_curso,
  main = "Distribuição por Modalidade de Ensino",
  col = c("#FFFF00", "#6e92f5ff"), # cores diferentes para cada fatia
  labels = paste(c("Educação a Distância", "Educação Presencial"), ":" ,
  round((modalidade_de_ensino_curso/1349)*100, 2), "%"))
dev.off()

# Categoria Administrativa por Instituição - Completo: difícil visibilidade
png("src/gráficos/pizza_categoria_administrativa_SC_completo.png")
pie(categoria_administrativa,
  main = "",
  col = c("#f3736fff", "#fda58fff", "#6e92f5ff", "#95c3ffff", "#77e9b0ff",
  "#62bd85ff", "#0ece8eff"), # cores diferentes para cada fatia
  radius = 0.44,
  labels = paste(c("Comunitária/
Confessional", "Especial", "Privada com Fins
Lucrativos", "
Privada sem Fins
Lucrativos", "
Pública Estadual", "Pública Federal", "Pública Municipal"),": ",round((
  categoria_administrativa/116)*100, 2), "%"))
title("Distribuição por Categoria Administrativa", line = -2, cex = 0.5)
dev.off()

# Categoria Administrativa por Instituição - Divisão geral
categorias <- factor(c("Outros", "Outros", "Privada", "Privada", "Pública", "Pú
blica", "Pública"))
lista_categorizada <- split(categoria_administrativa, categorias)
data <- c(sum(lista_categorizada$Outros), sum(lista_categorizada$Privada), sum(
  lista_categorizada$Pública))
png("src/gráficos/pizza_categoria_administrativa_SC_geral.png")
pie(data,

```

```

main = "Distribuição por Categoria Administrativa - Agrupada",
radius = 0.8,
col = c("#fda58fff", "#95c3ffff", "#77e9b0ff"), # cores diferentes para cada
fatia
labels = paste(c("Outros", "Privada", "Pública"), ": ", round((data/116)
*100, 2), "%"))
dev.off()

# Categoria Administrativa por Instituição - Divisão específica

#Privada
png("src/gráficos/pizza_categoria_administrativa_SC_privada.png")
pie(lista_categorisada$Privada,
     main = "Distribuição por Categoria Administrativa - Privadas",
     radius = 0.7,
     col = c("#6e92f5ff", "#95c3ffff"), # cores diferentes para cada fatia
     labels = paste(c("Privada com Fins
Lucrativos", "
Privada sem Fins
Lucrativos"), ": ", round((lista_categorisada$Privada/94)*100, 2), "%"
"))
dev.off()

#Pública
png("src/gráficos/pizza_categoria_administrativa_SC_publica.png")
pie(lista_categorisada$Pública,
     main = "",
     radius = 0.6,
     col = c("#77e9b0ff", "#62bd85ff", "#0ece8eff"), # cores diferentes para cada
fatia
     labels = paste(c("Pública
Estadual", "Pública Federal", "Pública Municipal"), ": ", round((
lista_categorisada$Pública/8)*100, 2), "%"))
title("Distribuição por Categoria Administrativa - Públia", line = -2, cex =
0.5)
dev.off()

#Outros
png("src/gráficos/pizza_categoria_administrativa_SC_outros.png")
pie(lista_categorisada$Outros,
     main = "Distribuição por Categoria Administrativa - Outros",
     radius = 0.9,

```

```

col = c("#f3736fff", "#fda58fff"), # cores diferentes para cada fatia
labels = paste(c("Comunitária/
Confessional", "Especial"), ":" ,
", round((lista_categorisada$Outros/14)*100, 2), "%"))
dev.off()

# Conceito ENADE - Contínuo
png("src/gráficos/barras_conceito_enade_SC.png")
barplot(height = as.vector(freq_enade_continuo$Frequênci),
names = c("", "", "", "", "", "", "", "", "", "", "", ""),
main = "Distribuição do Conceito ENADE",
xlab = "Conceito Enade",
ylab = "Participantes",
ylim = c(0, 250),
cex.axis = 0.7,
col = "#6e92f5ff")
axis(1, at=(c(1.3,2.5,3.7,4.9,6.1,7.3,8.5,9.7,10.9,12.1,13.3,14.4)),
labels = c(0.414,0.829,1.24,1.66,2.07,2.49,2.9,3.31,3.73,4.14,4.56,4.98),
cex.axis = 0.7)
dev.off()

png("src/gráficos/histograma_conceito_enade_SC.png")
hist(enade_continuo , breaks=50 ,
col="#6e92f5ff" ,
main="Distribuição do Conceito ENADE",
include.lowest = TRUE,
ylim = c(0,70),
xlab = "Conceito Enade", ylab = "Participantes",
border = "#6e92f5ff")

# Indicador de Diferença entre os Desempenhos Observados e Esperado (IDD) -
Continuo
png("src/gráficos/barras_IDD_SC.png")
barplot(height = freq_idd_continuo$Frequênci,
names = c("", "", "", "", "", "", "", "", "", "", "", ""),
main = "Distribuição do IDD",
xlab = "Indicador de Diferença entre os Desempenhos Observados e Esperado",
ylab = "Participantes",
ylim = c(0, 250),
cex.axis = 0.7,
col = "#6e92f5ff")
axis(1, at=(c(1.3,2.5,3.7,4.9,6.1,7.3,8.5,9.7,10.9,12.1,13.3,14.4)),

```

```

labels = c(0.417,0.833,1.25,1.67,2.08,2.5,2.92,3.33,3.75,4.17,4.58,5), cex.
axis = 0.7)
dev.off()

# Conceito Preliminar do Curso - CPC - Contínuo
png("src/gráficos/barras_CPC_SC.png")
barplot(height = freq_cpc_continuo$Frequência,
names = c("", "", "", "", "", "", "", "", "", "", "", ""),
main = "Distribuição do CPC",
xlab = "Conceito Preliminar do Curso",
ylab = "Participantes",
ylim = c(0, 250),
cex.axis = 0.7,
col = "#6e92f5ff")
axis(1, at=(c(1.3,2.5,3.7,4.9,6.1,7.3,8.5,9.7,10.9,12.1,13.3,14.4)),
labels = c(1.28,1.58,1.88,2.19,2.49,2.79,3.09,3.39,3.7,4,4.3,4.61), cex.axis
= 0.7)
dev.off()

# Índice Geral de Cursos - Contínuo
png("src/gráficos/barras_IGC_SC.png")
barplot(height = freq_igc_continuo$Frequência,
names = c("", "", "", "", "", "", "", "", ""),
main = "Distribuição do IGC",
xlab = "Índice Geral de Cursos",
ylab = "Instituições",
ylim = c(0, 80),
cex.axis = 0.7,
col = "#6e92f5ff")
axis(1, at=(c(1.3,2.5,3.7,4.9,6.1,7.3,8.5,9.7,10.9)),
labels = c(1.31,1.7,2.09,2.48,2.86,3.25,3.64,4.03,4.42), cex.axis = 0.7)
dev.off()

png("src/gráficos/histograma_IGC_SC.png")
hist(igc_continuo , breaks=30 ,
col="#6e92f5ff" ,
main="Distribuição do IGC",
include.lowest = TRUE,
ylim = c(0,25),
xlim = c(0.9,5),
xlab = "Índice Geral de Cursos", ylab = "Instituições",
border = "#6e92f5ff")

```

```

# ----- MEDIDAS DE TENDÊNCIA CENTRAL ----- #

cat("\n--- Calculando Medidas de Resumo para Santa Catarina ---\n")

# Lista das variáveis quantitativas de SC para analisar
vars_sc_para_analise <- list(
  "Conceito ENADE (Contínuo)" = data_CPC_SC$conceito_enade_.continuo.,
  "IDD (Contínuo)" = data_IDD_SC$idd_.continuo.,
  "CPC (Contínuo)" = data_CPC_SC$cpc_.continuo.,
  "IGC (Contínuo)" = data_IGC_SC$igc_.continuo.
)

# Criando uma tabela única com as principais medidas de resumo
tabela_medidas_resumo_sc <- data.frame(
  Indicador = names(vars_sc_para_analise),
  Media = sapply(vars_sc_para_analise, mean, na.rm = TRUE),
  Mediana = sapply(vars_sc_para_analise, median, na.rm = TRUE),

  # ----- MEDIDAS DE DISPERSÃO ----- #

  Desvio_Padrao = sapply(vars_sc_para_analise, sd, na.rm = TRUE),
  Variancia = sapply(vars_sc_para_analise, var, na.rm = TRUE),
  CV = sapply(vars_sc_para_analise, function(x) sd(x, na.rm = TRUE) / mean(x, na
    .rm = TRUE) * 100),
  Minimo = sapply(vars_sc_para_analise, min, na.rm = TRUE),
  Maximo = sapply(vars_sc_para_analise, max, na.rm = TRUE)
)

# Adicionando a Amplitude (Max - Min)
tabela_medidas_resumo_sc$Amplitude <- tabela_medidas_resumo_sc$Maximo -
  tabela_medidas_resumo_sc$Minimo

# Arredondando os valores para 3 casas decimais para melhor visualização
tabela_medidas_resumo_sc[, -1] <- round(tabela_medidas_resumo_sc[, -1], 3)

# Salvando a tabela em um arquivo CSV na pasta correta
write.csv(tabela_medidas_resumo_sc, "src/tabelas/SC/tabela_medidas_resumo_sc.
  csv", row.names = FALSE)

# ----- GRÁFICOS DE DISTRIBUIÇÃO NORMAL ----- #

criar_curva_normal <- function(dados, titulo, nome_arquivo, cor = "#6e92f5ff")
{

```

```

png(nome_arquivo, width = 800, height = 600)

h <- hist(dados, plot = FALSE, breaks = 30)

plot(h, freq = FALSE, main = titulo,
      xlab = titulo, ylab = "Densidade",
      col = cor, border = "white",
      ylim = c(0, max(h$density) * 1.1))

if(length(dados) > 1 && sd(dados, na.rm = TRUE) > 0) {
  x <- seq(min(dados, na.rm = TRUE), max(dados, na.rm = TRUE), length = 100)
  y <- dnorm(x, mean = mean(dados, na.rm = TRUE), sd = sd(dados, na.rm = TRUE))
}

lines(x, y, col = "red", lwd = 2)

legend("topright", legend = "Curva Normal", col = "red", lwd = 2, bty = "n")
}

dev.off()
}

criar_curva_normal(enade_continuo, "Distribuição Normal - Conceito ENADE",
                   "src/gráficos/curva_normal_enade_SC.png")

criar_curva_normal(idd_continuo, "Distribuição Normal - IDD",
                   "src/gráficos/curva_normal_idd_SC.png")

criar_curva_normal(cpc_continuo, "Distribuição Normal - CPC",
                   "src/gráficos/curva_normal_cpc_SC.png")

criar_curva_normal(igc_continuo, "Distribuição Normal - IGC",
                   "src/gráficos/curva_normal_igc_SC.png")

# ----- RELAÇÕES ENTRE VARIÁVEIS ----- #

# Vetor de nomes formatados para variáveis quantitativas
nomes_quantitativas <- c(
  "conceito_enade_.continuo." = "Conceito ENADE - Contínuo",
  "cpc_.continuo." = "CPC - Contínuo",
  "idd_.continuo." = "IDD - Contínuo",
  "igc_.continuo." = "IGC - Contínuo"
)

```

```

# Vetor de nomes formatados para variáveis qualitativas
nomes_qualitativas <- c(
  "categoria_administrativa" = "Categoria Administrativa",
  "modalidade_de_ensino" = "Modalidade de Ensino"
)

# Variáveis qualitativas
qualitativas <- c("categoria_administrativa", "modalidade_de_ensino")

# Variáveis quantitativas de CPC
quant_cpc <- c("conceito_enade_.continuo.", "cpc_.continuo.")

# Variáveis quantitativas de IDD
quant_idd <- c("idd_.continuo.")

# Relações em data_CPC_SC
for (q in qualitativas) {
  for (var in quant_cpc) {
    plotar_relacao(data_CPC_SC, q, var)
  }
}

# Relações em data_IDD_SC
for (q in qualitativas) {
  for (var in quant_idd) {
    plotar_relacao(data_IDD_SC, q, var)
  }
}

# ---TABELA DE CONTINGÊNCIA IGC x CATEGORIA ADMINISTRATIVA--- #

data_IGC_SC$uni_cat <- paste(data_IGC_SC$nome_da_ies,
  data_IGC_SC$categoria_administrativa, sep = "_")

# Calcular média de IGC por (universidade + categoria)
uni_cat_ids <- unique(data_IGC_SC$uni_cat)
igc_universidades <- data.frame(nome_da_ies = character(),
  categoria_administrativa = character(), igc_medio = numeric(),
  stringsAsFactors = FALSE)

for (id in uni_cat_ids) {
  linhas <- data_IGC_SC[data_IGC_SC$uni_cat == id, ]
  media_igc <- mean(linhas$igc_.continuo., na.rm = TRUE)
}

```

```

uni <- linhas$nome_da_ies[1]
cat_admin <- linhas$categoria_administrativa[1]

igc_universidades <- rbind(igc_universidades, data.frame(nome_da_ies = uni,
    categoria_administrativa = cat_admin, igc_medio = media_igc,
    stringsAsFactors = FALSE))
}

df_igc_bins <- salvar_frequencia_continuo(valores = igc_universidades$igc_medio
    , arquivo_saida = "src/tabelas/SC/igc_continuo_universidades_cat.csv")

k <- nrow(df_igc_bins)
igc_bins <- cut(igc_universidades$igc_medio, breaks = k, include.lowest = TRUE,
    right = FALSE)

tabela_contingencia <- table(igc_universidades$categoria_administrativa,
    igc_bins)

write.csv(as.data.frame(tabela_contingencia), "src/tabelas/SC/
    igc_categoria_administrativa_contingencia.csv", row.names = FALSE)

# ----- GRÁFICOS DE COMPARAÇÃO ENTRE VARIÁVEIS QUALITATIVAS ----- #

# Tabela de contingência correta entre modalidade e categoria administrativa (
# por cursos)
tabela_contingencia_modalidade_categoria <- table(
    data_CPC_SC$modalidade_de_ensino, data_CPC_SC$categoria_administrativa)
write.csv(as.data.frame(tabela_contingencia_modalidade_categoria), "src/tabelas
    /SC/contingencia_modalidade_categoria.csv", row.names = TRUE)

png("src/gráficos/proporcao_modalidade_por_categoria.png")
par(mar = c(5, 10, 4, 2)) # bottom, left, top, right; mais espaço para nomes
# longos
prop_table <- prop.table(tabela_contingencia_modalidade_categoria, margin = 2)
* 100

# Barras horizontais
barplot(prop_table,
    main = "Proporção de Modalidade por Categoria Administrativa (%)",
    xlab = "Percentual por Curso (%)",
    col = c("#FFFF00", "#6e92f5ff"),
    horiz = TRUE,
    las = 1, # rótulos horizontais legíveis

```

```

cex.names = 0.9)
legend("topright", legend = rownames(prop_table), fill = c("#FFFF00", "#6
e92f5ff"))
dev.off()

```

Listing 7.4: Anexo D: Funções utilitárias

```

salvar_frequencia_continuo <- function(valores, arquivo_saida, right = FALSE,
  include.lowest = TRUE) {
  # Número de classes pelo método de Sturges
  k <- nclass.Sturges(valores)
  classes <- cut(valores, breaks = k, include.lowest = include.lowest, right =
    right)
  freq <- table(classes)

  df_freq <- data.frame(
    Limites = names(freq),
    Frequência = as.vector(freq)
  )

  # Salvar CSV
  write.csv(df_freq, arquivo_saida, row.names = FALSE)
  return(df_freq)
}

plotar_relacao <- function(dados, qualitativa, quantitativa) {
  # Nome formatado da qualitativa
  nome_ql <- ifelse(qualitativa %in% names(nomes_qualitativas),
    nomes_qualitativas[[qualitativa]], qualitativa)
  # Nome formatado da quantitativa
  nome_qt <- ifelse(quantitativa %in% names(nomes_quantitativas),
    nomes_quantitativas[[quantitativa]], quantitativa)

  # Diretório onde os gráficos serão salvos
  dir_graficos <- "src/gráficos/"
  # Criar nome do arquivo dinamicamente com base nas variáveis
  arquivo_saida <- paste0(dir_graficos,"Boxplot_de_", gsub(" ", "_", nome_qt),"_
  _por_", gsub(" ", "_", nome_ql),".png")
  png(filename = arquivo_saida, width = 1500, height = 800)
  boxplot(dados[[quantitativa]] ~ dados[[qualitativa]],main = paste(nome_qt, "_
  por", nome_ql),xlab = nome_ql, ylab = nome_qt,col = "lightblue", las = 1)
  dev.off()
}

```