



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

**Inferência 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).

Nesse contexto, o ENADE fornece um conjunto de métricas destinadas à avaliação dos cursos e das instituições de ensino superior, 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). Esses indicadores, juntamente com as variáveis Categoria Administrativa e Modalidade de Ensino dos cursos avaliados, serão detalhados na seção de Materiais e Métodos.

O presente trabalho dá continuidade à análise exploratória realizada no Trabalho 1 da disciplina, ampliando o estudo por meio da aplicação de técnicas de inferência estatística – a partir dos dados de uma amostra, criamos testes de hipóteses para realizar conclusões e fazer generalizações sobre a população. Assim como na análise anterior, a população considerada corresponde aos cursos do estado de Santa Catarina. A partir desse conjunto de dados, será selecionada uma amostra aleatória, permitindo a realização de inferências estatísticas sobre a população com base nos resultados obtidos anteriormente, apresentados na Tabela 1.

A partir dos indicadores apresentados pelo ENADE, serão formuladas e testadas hipóteses relacionadas às médias, proporções, correlações e regressões, com o objetivo de aprofundar a compreensão acerca das variáveis quantitativas e qualitativas previamente examinadas.

Tabela 1: 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

1.1 Objetivos

O objetivo principal do trabalho é tomar os indicadores apresentados anteriormente e que também foram utilizados no trabalho anterior e testar hipóteses levantadas sobre eles, sendo essas hipóteses aplicadas sobre a média e proporção dos dados de indicadores, além de hipóteses sobre a correlação e regressão entre duas dessas variáveis.

Com base nos dados dos indicadores, foram definidas as hipóteses a serem analisadas no presente trabalho:

1. Hipótese sobre a média:

H_0 : A média do CPC em cursos com ensino presencial em Santa Catarina é igual a média do CPC em cursos com ensino a distância.

2. Hipótese sobre a proporção:

H_0 : A proporção de instituições federais em Santa Catarina é de 25%.

3. Hipótese sobre a Correlação:

H_0 : Não há correlação entre o IDD e o Conceito Enade.

4. Hipóteses sobre a Regressão:

H_0 : O IDD não tem influência significativa sobre o Conceito Enade.

2 Materiais e Métodos

Nesta seção são apresentadas as variáveis quantitativas e qualitativas selecionadas para as análises estatísticas, as quais foram previamente exploradas no trabalho anterior e agora serão utilizadas na etapa de inferência estatística. A base de dados considerada compreende os cursos participantes do ENADE no ciclo trienal de 2021 a 2023, pertencentes ao estado de Santa Catarina, que constitui a população de interesse deste estudo.

Para a realização das inferências estatísticas, será determinado o tamanho de amostra necessário para cada tipo de variável, conforme as expressões apresentadas nas Equações 1a e 1b, aplicadas respectivamente às variáveis qualitativas e quantitativas:

$$\left. \begin{array}{l} (a) \quad n_0 = \frac{z^2 \cdot p \cdot (1-p)}{d^2} \\ (b) \quad n_0 = \frac{z^2 \cdot \sigma^2}{d^2} \end{array} \right\} \quad n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (1)$$

Nas equações, n_0 representa o tamanho amostral inicial e n o tamanho amostral corrigido para populações finitas. O parâmetro z corresponde ao valor crítico da distribuição normal padrão associado ao nível de confiança adotado. Para obtermos um nível de confiança de 95%, iremos considerar $z = 1,96$. A variância populacional σ^2 é conhecida através da Tabela 1. O parâmetro p indica a proporção esperada de ocorrência de um determinado evento. Para nossa amostragem, consideraremos $p = 0,5$, a fim de maximizar a variância $p \cdot (1 - p)$. Já a variável d expressa o erro amostral tolerável, frequentemente fixado em 5%. E, por fim, N representa o tamanho da população.

2.1 Variáveis Quantitativas

1. Conceito Enade Contínuo (CE) - Variável Quantitativa 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, 2024).

Conforme do anexo 1, que sintetiza as fórmulas supracitadas, o tamanho da amostra aleatória para essa variável será $n = 699$.

2. Indicador de Diferença entre os Desempenhos Observados e Esperado Contínuo (IDD) - Variável Quantitativa Contínua:

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.

Conforme o script do anexo 1, o tamanho da amostra aleatória para essa variável será $n = 707$.

3. 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, 2024). Essa medida é relativa ao resultado médio da área de avaliação em todo o país, gerando uma faixa contínua definida de 1 a 5.

Conforme o script do anexo 1, o tamanho da amostra aleatória para essa variável será $n = 426$.

2.2 Variáveis Qualitativas

1. Categoria Administrativa - Variável Nominal

Categoria Administrativa 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. Conforme o script do anexo 1, o tamanho da amostra aleatória para essa variável será $n = 150$.

2. Modalidade de Ensino - Variável Nominal

A Modalidade de Ensino 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. Conforme o script do anexo 1, o tamanho da amostra aleatória para essa variável será $n = 300$.

2.2.1 Base de Dados

A base de dados escolhida para o trabalho foi a mesma já apresentada no primeiro trabalho: os dados do Conceito Enade de 2021, 2022 e 2023 disponível no site do ENADE em formato XLSX, contendo além do indicador de Conceito Enade, o CPC e o IDD.

Novamente, os dados foram tomados por um intervalo suficiente para abranger um ciclo trienal completo e filtrados com um script em Python, para formatação das tabelas e conversão do arquivo XLSX para CSV, e um script em R, para conter somente as entradas referentes ao estado de Santa Catarina.

3 Resultados e Discussões

4 Considerações Finais

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>.

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>>.

5 Anexos - Códigos Fonte