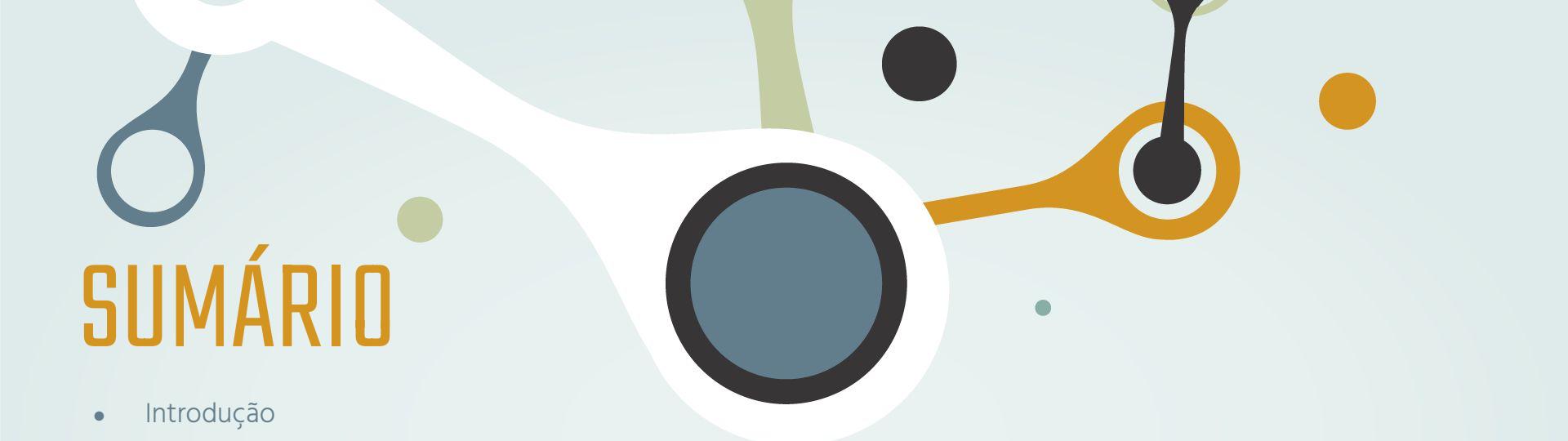


# ANÁLISE E INFERÊNCIA ESTATÍSTICA: Exame Nacional de Desempenho dos Estudantes (ENADE)

**Disciplina:** Probabilidade e Estatística (INE5405)

**Docente:** Dr<sup>a</sup>. Simone Silmara Werner

**Discentes:** Leonardo de S. Marques, Rafael V. Ribeiro, Ruan A. Ferreira, Thayse E. Teixeira



# SUMÁRIO

- Introdução
- Fonte dos dados
- Variáveis utilizadas
- Resultados do Trabalho 1
- Análise descritiva
- Hipóteses
- Amostragem
- Análises e Testes
- Resultados obtidos
- Conclusões

# 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 avaliação segue um ciclo trienal:

Ano	Cursos Avaliados
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
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.
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.

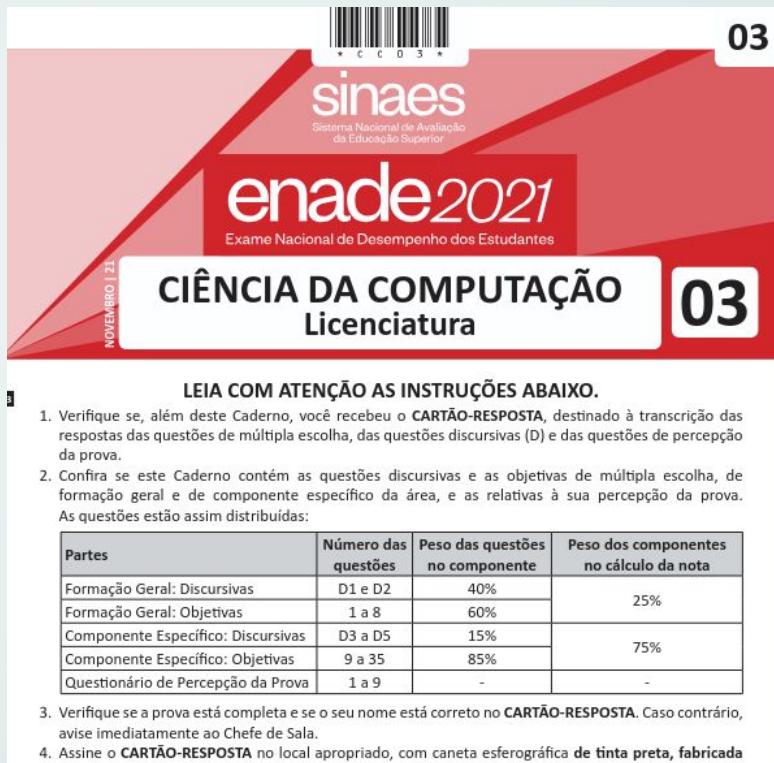
Fonte: Adaptado de INEP (2025)

# INTRODUÇÃO

A prova é composta por **40 questões**:

- 10 de Formação Geral (FG)
- 30 de Conhecimento Específico (CE).

Além disso, os estudantes também responderam 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).



# FONTE DE DADOS

A fim de obter uma maior cobertura de dados, iremos analisar um ciclo trienal completo, em relação às últimas provas aplicadas: **2021, 2022 e 2023.**

Devido o grande volume de dados, os dados foram filtrados para o Estado de Santa Catarina. Para fins do trabalho, esta será considerada nossa **população**.

```
.data
└── CPC_2021.xlsx
    ├── CPC_2022.xlsx
    ├── CPC_2023.xlsx
    └── IDD_2021.xlsx
        ├── IDD_2022.xlsx
        └── IDD_2023.xlsx
    └── IGC_2021.xlsx
        ├── IGC_2022.xlsx
        └── IGC_2023.xlsx
```

o ENADE utiliza as provas e os questionários para calcular índices 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).



# VARIÁVEIS UTILIZADAS

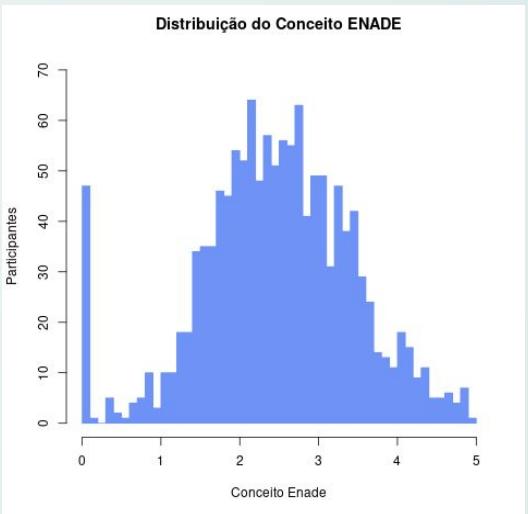
As variáveis utilizadas foram extraídas da base de dados de 2021, 2022 e 2023, disponível pelo ENADE.

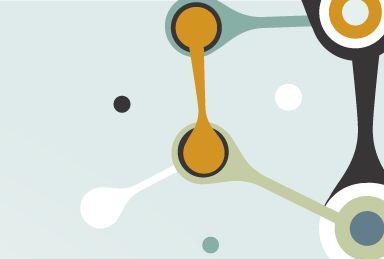
- **Conceito Enade (CE)**: quantitativa contínua [0, 5];
- **Conceito Preliminar de Curso (CPC)**: quantitativa contínua [0, 5];
- **Indicador de Diferença entre os Desempenhos Observados e Esperado (IDD)**: quantitativa contínua [0, 5];
- **Índice Geral de Cursos (IGC)**: quantitativa contínua [0, 5];
- **Categoría Administrativa**: qualitativa nominal (Pública Federal, Privada com fins lucrativos...);
- **Modalidade de Ensino**: qualitativa nominal (Presencial, Ensino a Distância...).

# RESULTADOS DO TRABALHO 1

A partir dos dados da população de Santa Catarina, obtivemos os seguintes resultados para as medidas de tendência central e dispersão:

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





# HIPÓTESES

Para este trabalho, nosso objetivo é avaliar as seguintes hipóteses:

1. **Hipótese sobre a Média:** 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:** A proporção de instituições federais em Santa Catarina é de 25%.
3. **Hipótese sobre a Correlação:** Não há correlação entre o IDD e o Conceito Enade
4. **Hipóteses sobre a Regressão:** O IDD não tem influência significativa sobre o Conceito Enade

Mas, antes, vamos realizar o processo de análise das variáveis e criar uma amostra.



# AMOSTRAGEM

Para realizarmos inferências estatísticas, criamos **amostras aleatórias** a partir da nossa população. Para isso, utilizamos as seguintes fórmulas:

**Variáveis qualitativas:**

$$n_0 = \frac{z^2 \cdot p \cdot (1 - p)}{d^2}$$

**Variáveis quantitativas:**

$$n_0 = \frac{z^2 \cdot \sigma^2}{d^2}$$

$$n = \left. \begin{array}{l} n_0 \\ \hline 1 + \frac{n_0}{N} \end{array} \right\}$$

Utilizamos:

**$z = 1.96$**  - para uma confiança de 95%,

**$p = 0.5$**  - para maximizar  $p(1-p)$ ,

**$d = 0.05$**  - erro máximo tolerável,

**$\sigma^2$**  é a variância da população, conhecida do Trabalho 1, assim como o tamanho da população **N**.

# AMOSTRAGEM

Variável	$\sigma^2$	N	n <sub>0</sub>	n
Conceito ENADE (Contínuo)	0.943	1349	1449.66	<b>699</b>
IDD (Contínuo)	0.965	1349	1483.46	<b>707</b>
CPC (Contínuo)	0.404	1349	620.73	<b>426</b>
Categoria Administrativa	—	244	384.16	<b>150</b>
Modalidade de Ensino	—	1349	384.16	<b>300</b>

IGC e Categoria Administrativa são fornecidos por instituição de ensino (244), e não por curso, diferentemente das demais medidas (1349).

# ANÁLISE DESCRIPTIVA

Em um primeiro momento essa análise, feita sobre as amostras obtidas, foca nas estimativas das variáveis qualitativas, onde é possível fazer algumas observações:

Variável	Média	Mediana	Desvio Padrão
Conceito Enade	2.4753	2.4753	0.9464
CPC	2.8335	2.8249	0.6381
IDD	2.6826	2.661	1.0202

- Quanto aos desvios, fica claro que o CPC concentra mais seus valores ao redor da média, alcançando um desvio menor. Isso reflete o fato de as outras duas variáveis possuírem uma pequena concentração de valores nas extremidades dos seus intervalos.
- O Conceito Enade e o CPC alcançam média e mediana levemente abaixo do centro de seus intervalos (2,5 e 3, respectivamente);
- O IDD apresenta esse deslocamento para cima de seu centro (2,5);

# ANÁLISE DESCRIPTIVA

Variável	Intervalo de Confiança (Média)
Conceito Enade	(2.405, 2.5456)
CPC	(2.7727, 2.8942)
IDD	(2.6076, 2.7582)

Variável	Intervalo de Confiança (Proporção (%))
Modalidade de Ensino (Presencial)	(72.5957, 82.071)
Categoria Administrativa (Públicas)	(2.6748, 10.6585)

Já ao analisar os intervalos de 95% de confiança das estimativas, fica perceptível que:

- Os intervalos das variáveis quantitativas possuem uma amplitude semelhante, indicando que seus desvios tiveram pouco impacto na estimativa. Isso demonstra como a amostragem levando em conta parâmetros como a variância da população auxiliam a alcançar melhores resultados;
- Os intervalos das estimativas de proporção são relativamente mais extensos.



# HIPÓTESE SOBRE A MÉDIA

Para realizar o teste de hipótese, precisamos explicitar nossas hipóteses:

1. **Hipótese nula ( $H_0$ ):** A média do **CPC** em cursos com ensino presencial em Santa Catarina é **igual** a média do CPC em cursos com ensino a distância;  
(i.e.,  $\mu_{\text{Presencial}} = \mu_{\text{EaD}}$  ).
2. **Hipótese alternativa ( $H_1$ ):** A média do **CPC** em cursos com ensino presencial em Santa Catarina é **maior** que a média do CPC em cursos com ensino a distância;  
(i.e.,  $\mu_{\text{Presencial}} > \mu_{\text{EaD}}$  ).

Para realizar o teste, utilizamos **n = 426** como tamanho da amostra. Além disso, como queremos verificar se  $\mu_{\text{Presencial}} > \mu_{\text{EaD}}$ , aplica-se um **teste unilateral à direita**.

# HIPÓTESE SOBRE A MÉDIA

Aplica-se um **Teste-t de Welch**, uma vez que  $\sigma^2_{\text{Presencial}} \neq \sigma^2_{\text{EaD}}$ .

No R:

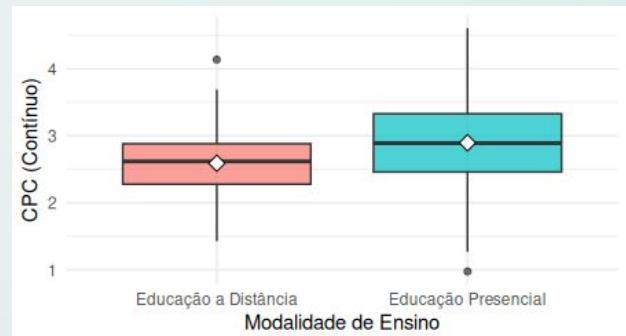
```
teste_t <- t.test(presencial, ead,  
                    alternative = "greater", # teste unilateral à direita  
                    var.equal = FALSE) # t-Welch (variâncias diferentes)
```

## Resultado:

Estatística	Valor	Descrição
t	4.6738	Estatística de teste
Graus de liberdade (df)	162.24	Graus de liberdade aproximados (Welch)
p-valor	3.09e-06	Probabilidade de erro tipo I
Hipótese alternativa	$\mu_{\text{Presencial}} > \mu_{\text{EaD}}$	Teste unilateral à direita
IC 95% da diferença	$[0.1967, +\infty)$	Intervalo de confiança unilateral
Média (Presencial)	2.894211	Média amostral dos cursos presenciais
Média (EaD)	2.589764	Média amostral dos cursos a distância

Uma vez que utilizamos  $\gamma = 95\%$ , assumimos um nível de significância  $\alpha = 5\%$ .

Uma vez que p-valor <  $\alpha$ , **rejeita-se a hipótese nula**. Conclui-se, portanto, que os cursos com ensino presencial apresentam CPC médio significativamente superior aos cursos com ensino a distância



# HIPÓTESE SOBRE A PROPORÇÃO

O objetivo desta análise foi verificar se a proporção de instituições federais em SC é de 25%, conforme afirma a hipótese nula. A variável analisada foi a categoria administrativa, sendo considerado “sucesso” cada instituição classificada como Pública Federal.

1. **Hipótese nula ( $H_0$ ):** A proporção de instituições federais em Santa Catarina é **igual** a 25% (i.e.,  $p = 0, 25$ ).
2. **Hipótese alternativa ( $H_1$ ):** A proporção de instituições federais em Santa Catarina é **diferente** de 25% (i.e.,  $p \neq 0, 25$ ).

Para essa variável qualitativa, utilizamos uma amostra de tamanho  **$n = 150$**  e aplicamos um **teste de proporção bilateral**.



# HIPÓTESE SOBRE A PROPORÇÃO

Para o teste de hipótese, considerou-se como “sucesso” cada instituição classificada como “Pública Federal”.

No R:

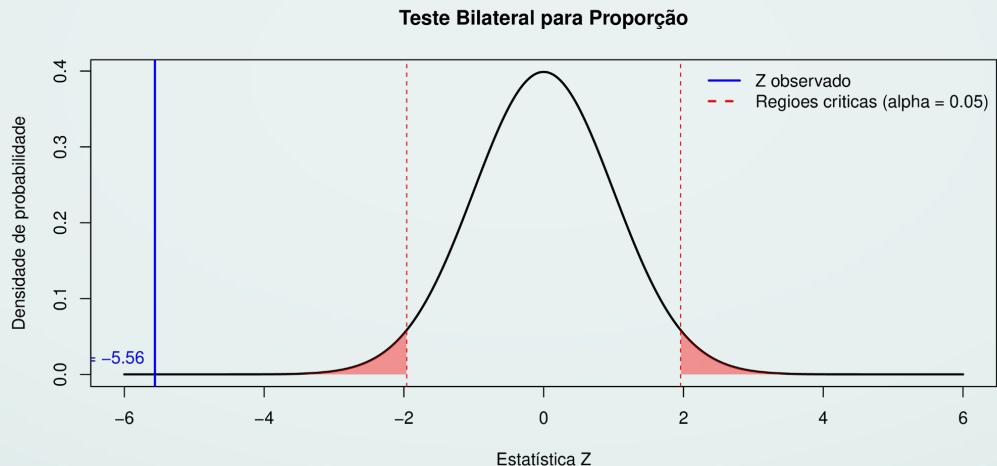
```
teste_prop <- prop.test(x = sucesso,  
                         n = total,  
                         p = 0.25,  
                         alternative = "two.sided", # teste bilateral  
                         correct = FALSE)          # sem correção de continuidade
```

## Resultado:

Estatística	Valor	Descrição
Tamanho da amostra	150	Número total de instituições na amostra
Número de instituições federais	8	Instituições federais observadas na amostra
Proporção observada	<b>0.0533</b>	Proporção amostral de instituições federais
Qui-quadrado	30.942	Estatística do teste qui-quadrado
Graus de liberdade (df)	1	Graus de liberdade do teste
p-valor	2.66e-08	Probabilidade de erro tipo I
Hipótese alternativa	$p \neq 0.25$	Teste bilateral para proporção
IC 95% da proporção	[0.0273, 0.1017]	Intervalo de confiança de 95% para a proporção

# HIPÓTESE SOBRE A PROPORÇÃO

Dado que o p-valor é muito inferior ao nível de significância de 5%, **rejeita-se a hipótese nula**, concluindo que há evidências estatísticas de que a proporção de instituições federais em Santa Catarina é significativamente diferente de 25%.



# HIPÓTESE DE CORRELAÇÃO

O teste realizado para conferir a correlação entre as variáveis IDD e Conceito Enade considerou as seguintes hipóteses:

1. **Hipótese nula ( $H_0$ ):** Não há correlação entre o IDD e o Conceito Enade
2. **Hipótese Alternativa ( $H_1$ ):** Existe correlação entre IDD e Conceito Enade

Para testar a Hipótese, primeiro foram tomadas as tabelas de IDD e CPC filtradas para o estado de Santa Catarina e criadas amostras independentes de semente = 123 e nCE = 698 e nIDD = 706.



# HIPÓTESE DE CORRELAÇÃO

O teste foi realizado usando Correlação Linear de Pearson sobre uma tabela contendo as 388 entradas relacionadas de IDD e Conceito Enade:

No R:

```
cor.test(data_corre_reg$conceito_enade, data_corre_reg$idd, method = "pearson")
```

## Resultados:

Estatística	Valor	Descrição
Coeficiente de Correlação de Pearson (r)	0,6945254	Força e a direção da relação linear entre as variáveis.
Valor t calculado	18,966	Estatística de teste associada ao coeficiente de correlação.
Graus de liberdade (gl)	386	Determinado por $n-2$ , onde $n$ é o tamanho da amostra.
p-valor	$< 2,2 \times 10^{-16}$	Probabilidade de observar esse valor de correlação se não houver relação linear na população.
Intervalo de confiança (95%)	[0,6391626 ; 0,7427272]	Faixa onde se espera que esteja o valor verdadeiro da correlação populacional com 95% de confiança.

Observou-se uma forte correlação positiva do coeficiente de correlação ao estar próximo de  $r = 1$ , além de estar dentro do intervalo de confiança calculado.

O p-valor se mostrou muito inferior ao nível de significância adotado (0.05), o que leva a **rejeição da hipótese nula  $H_0$** , ou seja, existe sim uma correlação significativa entre os valores dos indicadores de IDD e o Conceito Enade.

# HIPÓTESE DE REGRESSÃO

Quanto ao teste de hipótese com referência a regressão, consideramos novamente a relação entre as variáveis IDD e Conceito Enade com as seguintes hipóteses:

1. **Hipótese nula ( $H_0$ ):** O IDD não tem influência significativa sobre o Conceito Enade
2. **Hipótese Alternativa ( $H_1$ ):** O IDD tem influência significativa sobre o Conceito Enade

Para testar a Hipótese, fora definido o modelo de regressão linear simples como segue:

$$\text{Conceito Enade} = \beta_0 + \beta_1 \times (\text{IDD}) + \varepsilon$$

A hipótese nula será aceita caso o valor de  $\beta_1$  seja igual a 0, ou seja, o valor do IDD não tem influência sobre o valor do Conceito Enade.



# HIPÓTESE DE REGRESSÃO

No R:

**# Modelo de regressão linear (ENADE dependente de IDD)**

```
modelo <- lm(conceito_enade ~ idd, data = data_corre_reg)
```

**# Sumário com teste t dos coeficientes**

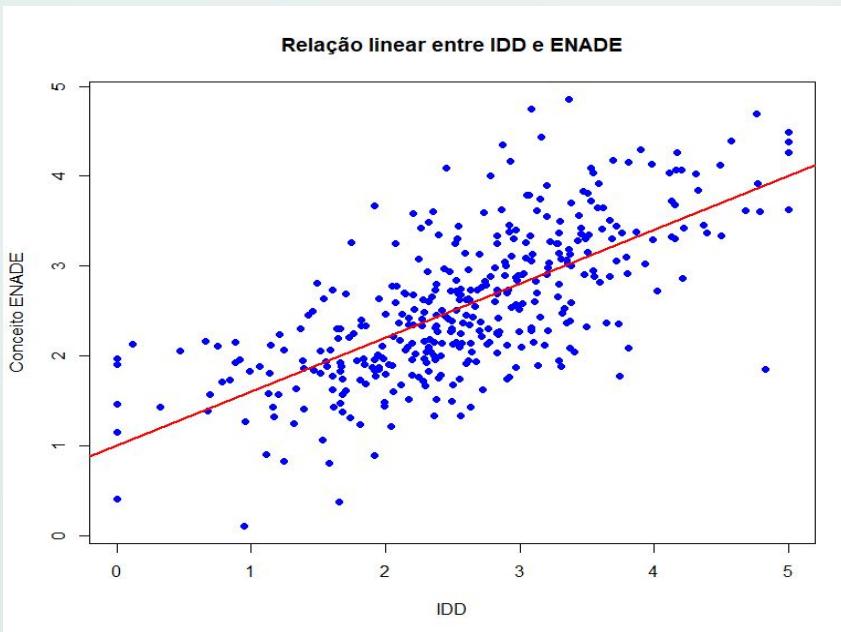
```
summary(modelo)
```

## Resultados:

Estatística	Valor	Descrição
Intercepto ( $\beta_0$ )	1,00395	Valor esperado do Conceito ENADE quando o IDD é igual a 0.
Coeficiente angular ( $\beta_1$ )	0,59984	Variação média esperada no Conceito ENADE para cada unidade adicional de IDD.
Erro padrão residual	0,5829	Mede a dispersão dos resíduos (diferença entre valores observados e ajustados).
p-valor ( $\beta_1$ )	$< 2,2 \times 10^{-16}$	Probabilidade de observar esse valor de correlação se não houver relação linear na população.
Coeficiente de determinação ajustado ( $R^2$ ajustado)	0,481	Proporção da variação do Conceito ENADE explicada linearmente pelo IDD, ajustada para o número de observações.

# HIPÓTESE DE REGRESSÃO

O valor do Coeficiente de Determinação  $R^2$  do modelo ajustado(0,481) indica que o IDD é um forte preditor do Conceito ENADE, explicando cerca de 48,1% de sua variação, porém, 51,9% ainda não é explicado pelo modelo.



O coeficiente angular positivo (0,59984) mostra que, à medida que o IDD aumenta, o Conceito ENADE também tende a aumentar proporcionalmente.

O p-valor extremamente pequeno ( $<2,2 \times 10^{-16}$ ) confirma que essa relação é estatisticamente significativa e como possui valor menor que os 0,05 de nível de significância, **rejeita-se a hipótese nula H0**.

# CONCLUSÕES

O trabalho avançou na análise exploratória aplicando inferência estatística para testar hipóteses sobre indicadores de qualidade do ENADE em Santa Catarina (média do CPC, proporção de federais e relação IDD/Conceito Enade).

## Resultados dos Testes ( $\alpha = 5\%$ )

- **H1 (Média):** Rejeita-se  $H_0$ . Cursos presenciais possuem CPC médio significativamente superior aos da modalidade EaD ( $p\text{-valor} \approx 3,09 \times 10^{-6}$ ).
- **H2 (Proporção):** Rejeita-se  $H_0$ . A proporção de instituições federais ( $\approx 5,3\%$ ) é significativamente diferente (e menor) dos 25% hipotetizados ( $p\text{-valor} \approx 2,66 \times 10^{-8}$ ).
- **H3 (Correlação):** Rejeita-se  $H_0$ . Há uma forte correlação positiva ( $r \approx 0,694$ ) entre o IDD e o Conceito Enade.
- **H4 (Regressão):** Rejeita-se  $H_0$ . O IDD tem influência estatisticamente significativa sobre o Conceito Enade

# CONCLUSÕES

## Limitações e Trabalhos Futuros

- O IDD, embora significativo, explica menos da metade da variabilidade do Conceito Enade ( $R^2 \approx 0,481$ ).
- 51,9% da variação não é explicada pelo IDD neste modelo.
- Sugere-se a inclusão de outras variáveis (ex: infraestrutura, qualificação docente) em estudos futuros para um modelo preditivo mais completo.

## Conclusão Estatística

- Os resultados são **fortes evidências** contra as hipóteses nulas, não provas absolutas.
- É crucial diferenciar **significância estatística** de **causalidade** e relevância prática.

# REFERÊNCIAS

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Áreas de conhecimento e eixos tecnológicos do Enade.** [S.l.: 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>>.

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>>.

INDICADORES DA EDUCAÇÃO SUPERIOR (CGGI/DAES/INEP), Coordenação-Geral de. **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](https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2024/nota_tecnica_6_metodologia_calculo_conceito_enade_2023.pdf)>.



# Dados e Scripts:



<https://github.com/leonardosm14/ENADE-Estatisticas>