

Análise Descritiva dos Indicadores de Qualidade da Educação Superior

Brenno Soares Aguiar, Davi Alves Mugayar, Ian Carlos Lima Tavares e Joel Antônio Rezende

1. Introdução

O presente trabalho realiza uma análise estatística descritiva do conjunto de dados (<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>), com cerca de 9.000 observações e 10 variáveis dos Indicadores de Qualidade da Educação Superior Brasileira de 2023, disponibilizados pelo Inep, focando nos resultados do ENADE e IDD.

O objetivo é compreender a qualidade da educação superior no Brasil, verificando as seguintes teses:

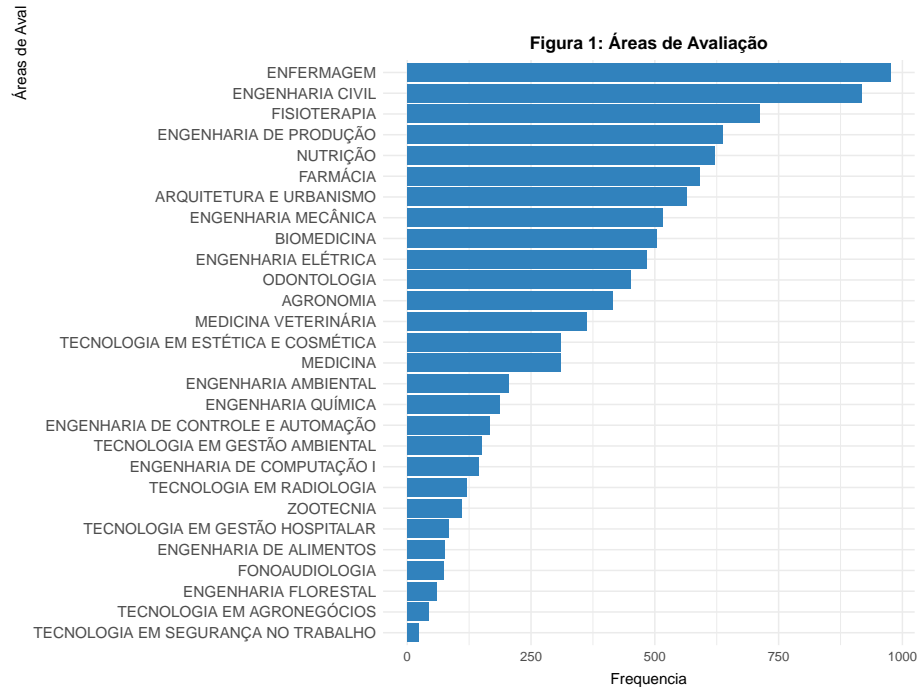
- (1) Existe uma estrutura concentrada e desequilibrada da Educação Superior Brasileira em um número limitado de categorias.
- (2) Há forte correlação entre Formação Geral (FG) e Componente Específico (CE)
- (3) O número de participantes não afeta a qualidade de ensino avaliada
- (4) Fatores como modalidade de ensino e categoria administrativa podem influenciar ou ter relação com a qualidade de ensino avaliada
- (5) A qualidade de ensino, conceito Enade, pode ser influenciada pelo IDD

Para verificar isso, foi necessário compreender a distribuição e frequência das variáveis para reconhecer tendências e possíveis concentração de valores e, posteriormente, investigar relações estatísticas entre os indicadores e variáveis para entender como se associam ou influenciam uns aos outros.

2. Análises Descritivas da Variáveis

2.1. Área de Avaliação

A variável “Área de Avaliação” é de natureza qualitativa que representa os cursos de graduação ou de educação superior.



Conforme evidenciado no Gráfico de Barras (Figura 1), há uma clara concentração nas áreas de saúde e ciências sociais, sendo a ENFERMAGEM a área de avaliação mais frequente, correspondendo aproximadamente 9.95% das áreas do conjunto de dados.

Por outro lado, cursos como Tecnologia em Segurança no Trabalho e Engenharia Florestal apresentam menores frequências, evidenciando que a demanda do mercado de trabalho e a tradição acadêmica podem favorecer áreas da saúde e engenharia que costumam apresentar alta demanda e tradição social historicamente.

2.2. Categoria Administrativa

A Categoria Administrativa é uma variável qualitativa, que representa o tipo de administração que a Instituição de Ensino Superior (IES) pode assumir.

Table 1: Tabela 1: Categorias Administrativas

Categoria Administrativa	Frequência	Frequência Relativa (%)
Privada com fins lucrativos	5049	51.46
Privada sem fins lucrativos	1788	18.22
Pública Federal	1492	15.21
Comunitária/Confessional	920	9.38
Pública Estadual	431	4.39
Pública Municipal	67	0.68
Especial	65	0.66
Total	9812	100.00

A análise da Categoria Administrativa revela uma forte predominância das instituições privadas. A categoria mais frequente é Privada com fins lucrativos, que representa 51.46% do total de registros. Em seguida, aparece a categoria Privada sem fins lucrativos com 18.22%. As categorias de natureza pública (federal, estadual e municipal) somam a menor parte, indicando que a amostra do ENADE está majoritariamente concentrada em instituições privadas, totalizando 69.68% de representação.

2.3. Modalidade de Ensino

A modalidade de ensino é uma variável qualitativa que representa a forma de ensino, seja presencialmente ou de forma on-line (a distância).

Table 2: Tabela 2: Modalidades de Ensino

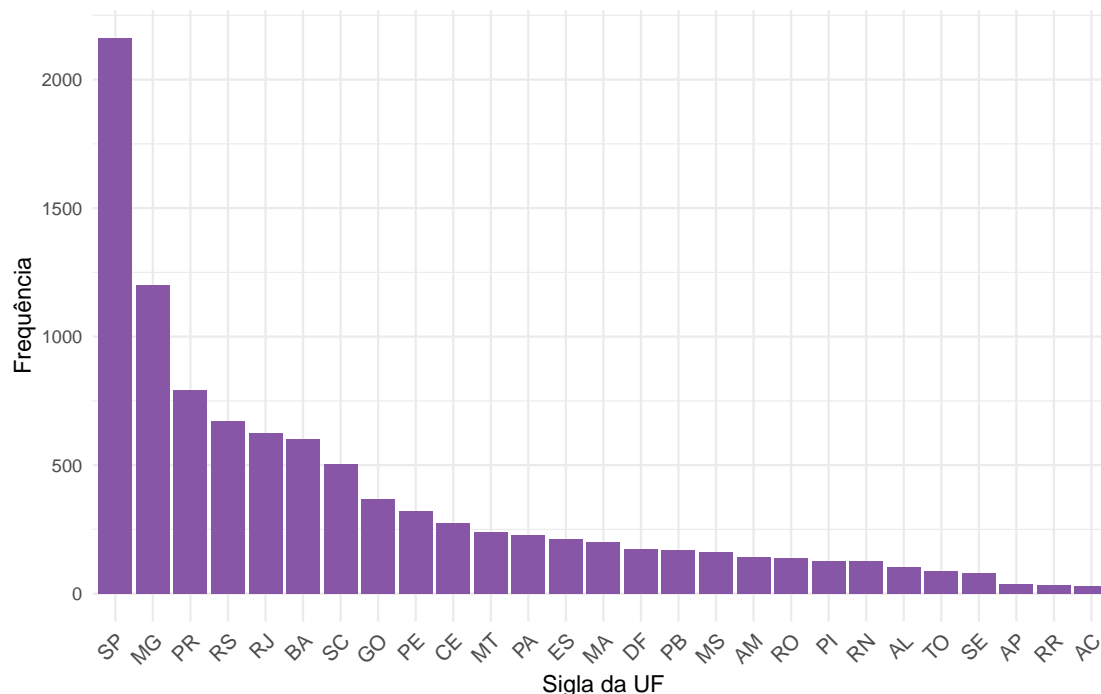
Modalidade de Ensino	Frequencia Absoluta (n)	Frequencia Relativa (%)
Educação Presencial	9120	92.95
Educação a Distância	692	7.05
Total	9812	100.00

Analizando a figura 3, percebe-se que a grande maioria dos cursos são realizados na modalidade Educação Presencial, totalizando 9120 observações. Em termos relativos, essa modalidade responde por 92.95% do conjunto de dados. O contraste entre as categorias é significativo, de modo que a diferença entre as duas modalidades é de 8428 cursos, o que se traduz em uma disparidade de 85.9 pontos percentuais, indicando uma clara dominância e concentração na modalidade Educação Presencial.

2.4. Sigla da UF

A variável UF é uma variável qualitativa que representa Unidade Federativa (UF), que é uma abreviação para os 26 estados brasileiros e o Distrito Federal.

Figura 2: Unidades Federativas

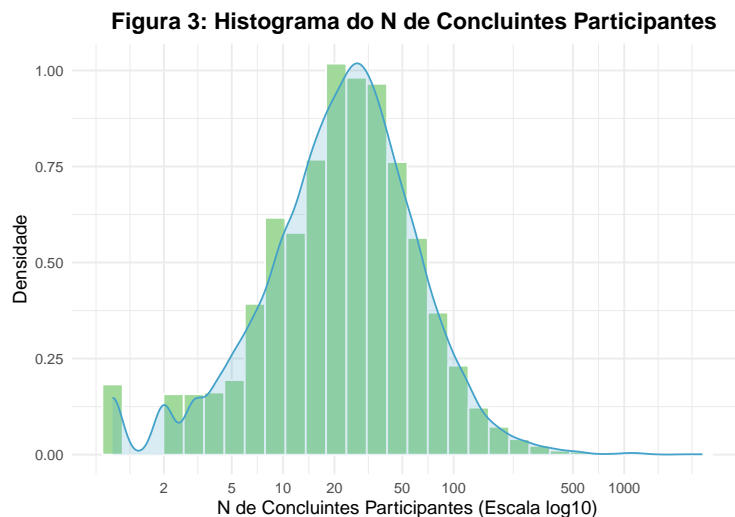


A análise da distribuição de frequência da variável Sigla da UF mostra que UF mais frequente é SP, com 2162 registros, o que representa aproximadamente 22.03% do conjunto. Em seguida, encontra-se MG com 1201 registros.

Essa concentração de instituições em poucas UFs, como SP e MG, aponta para uma disparidade regional da presença de cursos concentrados em determinadas UFs.

2.5. Número de Concluintes Participantes

A variável Número de Concluintes Participantes é uma variável quantitativa discreta que pode assumir diferentes valores devido a variedade na quantidade de estudantes que estavam concluindo o curso e que efetivamente participaram da prova do ENADE.



Primeiramente, nota-se que 50% dos cursos avaliados tiveram a mediana 23 concluintes. Sob outra ótica, a diferença significativa entre a Mediana e a Média 36.16 confirma que poucas instituições possuem um número muito elevado de concluintes - chegando ao máximo de 2874, enquanto a grande maioria concentra-se em turmas menores -.

2.6. Nota Padronizada FG

A Nota Padronizada FG é uma variável quantitativa contínua (com valores de 0 a 5) que reflete o desempenho médio dos estudantes de uma instituição no componente de Formação Geral do ENADE.

O Histograma (figura 4) apresenta uma forma que se aproxima de uma distribuição normal (simétrica), com a maior concentração de frequência nas classes centrais. Essa concentração máxima de observações está localizada em torno das notas de 3.25 a 3.75, indicando que a maior parte das instituições obteve uma performance mediana.

O Boxplot (Figura 5) complementa esta visualização. A linha central da caixa, que representa a Mediana (Q2), posiciona-se em 2.3891, sugerindo uma distribuição relativamente simétrica dos dados.

Ademais, a Média é de 2.466, demonstrando que o valor da Média é ligeiramente inferior ao da Mediana e, por isso, existe uma leve assimetria negativa (à esquerda) visualizada no Histograma.

Figura 4: Histograma da Nota FG

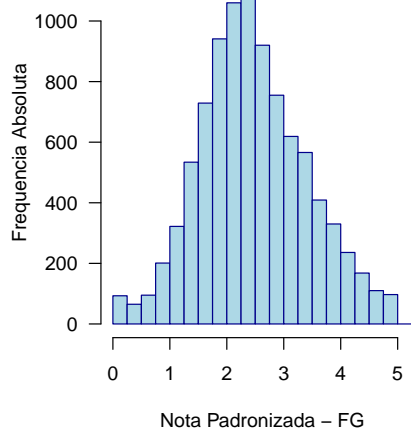
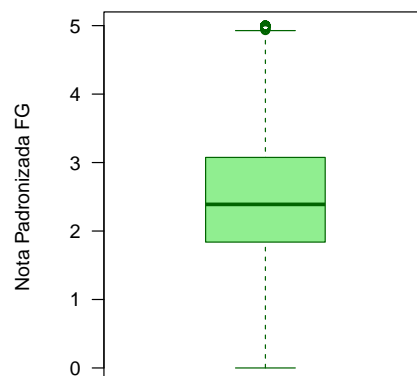


Figura 5: Boxplot da Nota FG



2.7. Nota Padronizada CE

A Nota Padronizada CE é uma variável quantitativa contínua (com valores de 0 a 5) que reflete o desempenho médio dos estudantes de uma instituição no componente de Formação Específica do ENADE.

Figura 6: Histograma da Nota CE

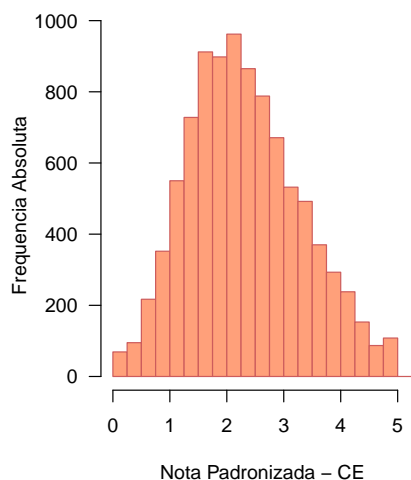
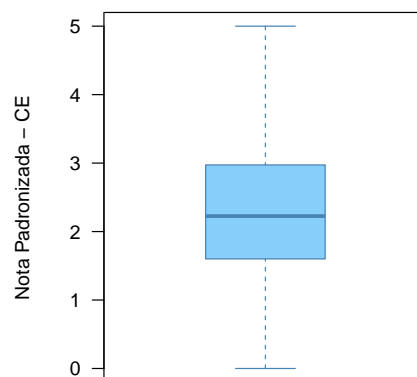


Figura 7: Boxplot da Nota CE



Em primeiro plano, o Histograma (figura 6) mostra uma curva aproximadamente simétrica, com o pico de frequência (moda) bem concentrada nas classes entre 1.25 e 1.75 (notas entre 3.15 e 3.65). Isso demonstra que a maioria das instituições apresenta um desempenho intermediário no componente específico do ENADE.

O Boxplot (figura 7) corrobora a simetria, visto que a Mediana (Q2) de 2.2262 está centralizada dentro da caixa (Intervalo Interquartil). O Boxplot também não exibe outliers, o que sugere uma ausência de notas extremamente discrepantes nos extremos.

Em relação às Medidas Separatrizes, o 90º Percentil (P90), com valor de 3.7028, mostra que apenas 10% das instituições conseguiram uma Nota Padronizada - CE igual ou superior a este valor, estabelecendo um limite de alto desempenho neste componente.

2.8. Conceito Enade

O Conceito Enade é uma variável quantitativa contínua (com valores de 0 a 5) que sintetiza o desempenho dos estudantes em Formação Geral e Componente Específico, sendo a nota final para classificar a qualidade de um curso.

Figura 8: Histograma Conceito Enade

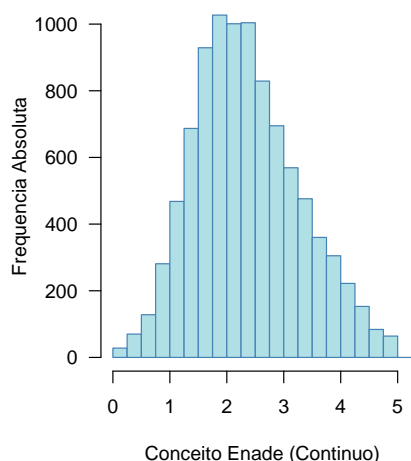
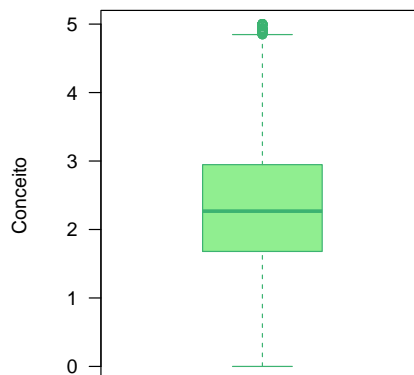


Figura 9: Boxplot do Conceito Enade



A distribuição do Conceito Enade no histograma (figura 8) é o reflexo da média ponderada dos componentes de Formação Geral e Específica, e, por isso, sua forma é similar às variáveis anteriores. Ele exhibe que a moda está concentrada em torno dos conceitos de 3.35 a 3.85, indicando que a maioria dos cursos avaliados pelo ENADE alcança conceitos médios a bons.

O Boxplot (figura 9) mostra uma situação relevante. A Média, que é de 2.3556, é marginalmente menor que a Mediana 2.2681, o que sugere uma leve assimetria negativa (ou à esquerda), indicando que os cursos com notas mais baixas puxam a média ligeiramente para baixo.

Por fim, o 90º Percentil (P90) atinge 3.6582, mostrando novamente que apenas 10% dos cursos alcançaram um Conceito Enade igual ou superior a este valor, representando o topo da qualidade de ensino avaliada.

2.9. Nota Bruta - IDD

A Nota Bruta IDD é uma variável qualitativa contínua que representa a comparação entre a nota dos concluintes no ENADE e a nota dos ingressantes, estimada por indicadores como o Enem ou outros dados.

Figura 10: Histograma do IDD Bruto

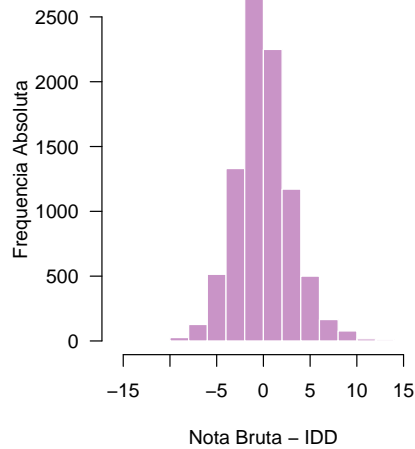
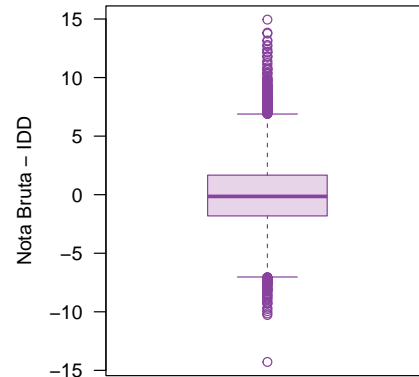


Figura 11: Box Plot do IDD Bruto



A análise do histograma (figura 9) revela uma distribuição que se aproxima da normalidade, mas com uma leve assimetria à esquerda (negativa). A Média das notas é de 0, e a Mediana (Q2) é ligeiramente maior, em -0.1485627, indicando que a metade dos cursos tem uma nota abaixo deste valor e a outra metade acima.

O Box Plot (Figura 10) confirma a concentração dos dados, onde 50% das notas estão contidas no Intervalo Interquartil (IQR) entre Q1 (-1.8110063) e Q3 (1.6679733). Existem vários outliers (pontos extremos), representando cursos com notas de IDD muito distantes do corpo principal, ou seja, cursos com desempenhos muito superiores e muito inferiores em relação à maioria.

2.10. IDD Contínuo

A variável IDD Contínuo é a transformação da nota bruta do IDD em uma nota numérica exata do indicador, em escala contínua (0 a 5), que busca manter uma precisão numérica real.

Figura 11: Histograma IDD Contínuo

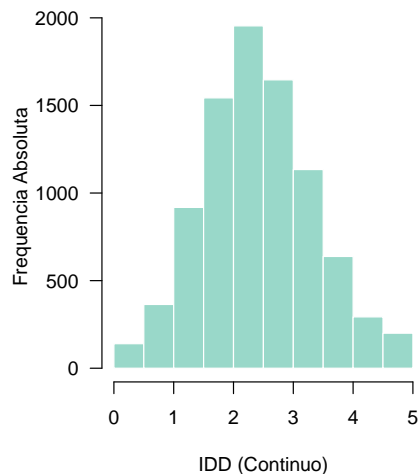
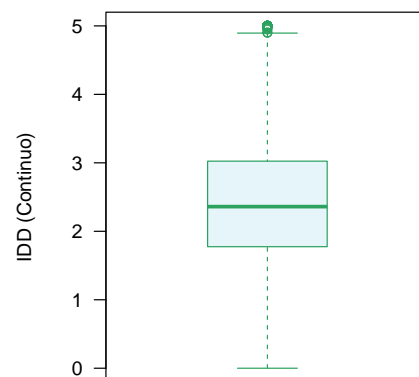


Figura 12: Box Plot do IDD (Contínuo)



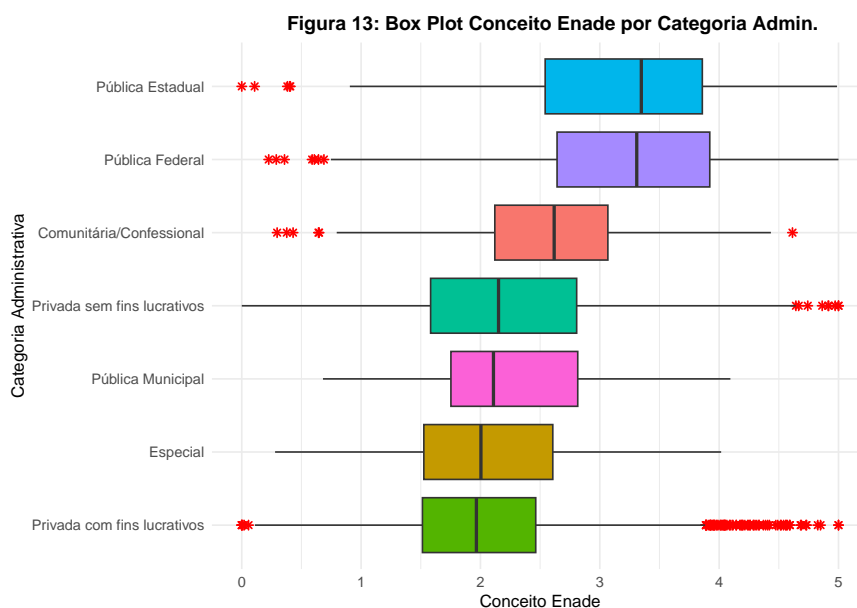
A distribuição do IDD (Contínuo) se apresenta de forma aproximadamente normal, centralizada e simétrica, conforme observado no Histograma (Figura 11). A Média (2.416) e a Mediana (2.3601294) são valores muito próximos, o que reforça essa simetria.

O IDD (Contínuo) varia de 0 a 5, e a maior parte dos dados está concentrada entre o Q1 (1.7751004) e o Q3 (3.0241881), totalizando um Intervalo Interquartil (IQR) de 1.2490878. O Box Plot (Figura 12) mostra que a distribuição é bem equilibrada, com poucos outliers localizados nas caudas.

3. Análises de Variáveis Em Conjunto

3.1. Análise Em Conjunto de Categoria Administrativa e Conceito Enade

Pretende-se verificar se a categoria administrativa pode influenciar valores maiores ou menores do conceito enade e, se sim, quais delas possuem valores superiores.



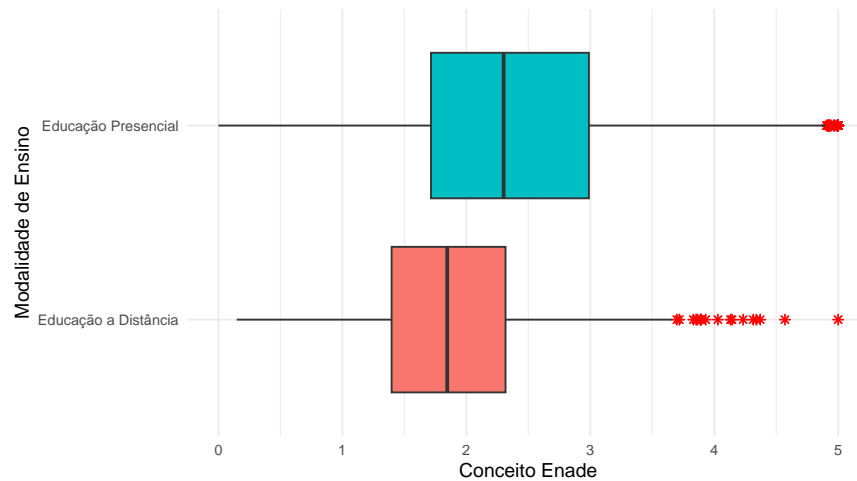
A figura 13 demonstra claramente que a categoria com a melhor performance em termos de mediana é Pública Estadual, alcançando um Conceito Enade de 3.3478. Em contraste, a categoria Privada com fins lucrativos registra o desempenho mais baixo, com mediana de 1.9659.

Ademais, nota-se que as categorias Pública Federal e Pública Estadual não apenas têm as maiores medianas, mas também apresentam distribuições mais concentradas e com quartis superiores, o que sugere maior qualidade e menor dispersão nos resultados. Em resumo, a afiliação administrativa está relacionada e associada valores maiores ou menores do conceito enade.

3.2. Análise Em Conjunto de Modalidade de Ensino e Conceito Enade

Pretende-se verificar se a modalidade de ensino também pode se correlacionar com o conceito enade e, se sim, qual delas possui maior correlação.

Figura 14: Conceito Enade por Modalidade de Ensino



O Box Plot (Figura 14) mostram que a Educação Presencial apresenta o melhor desempenho, com uma mediana de 2.3009. A modalidade Educação a Distância, por sua vez, registra a menor mediana, em 1.8461.

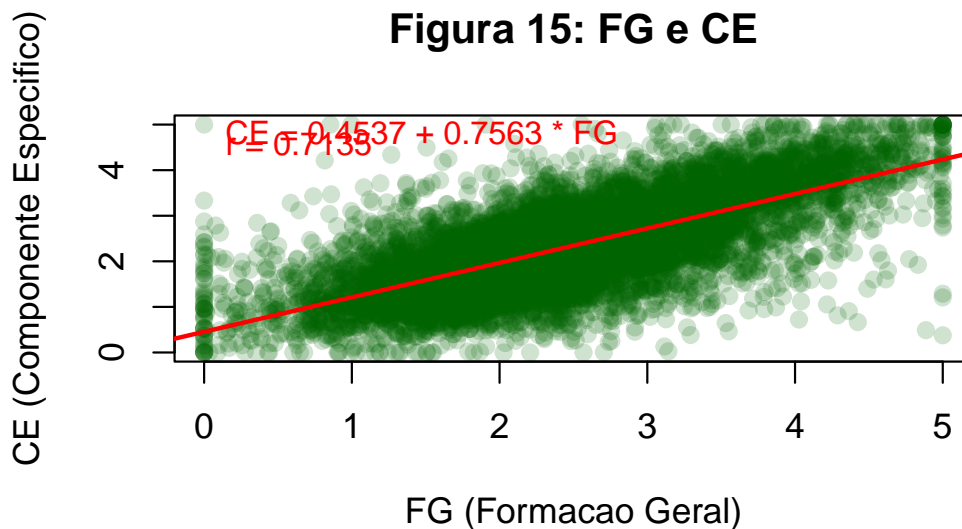
Ademais, percebe-se que, embora a modalidade com melhor desempenho possua uma distribuição mais favorável (caixa mais alta), a dispersão dos resultados (tamanho da caixa) entre as modalidades não é drasticamente diferente, sugerindo que a modalidade é um fator de influência, mas o desempenho final é determinado por outros fatores internos da IES.

4. Análises de Correlação e Regressão Linear Simples

4.1. Nota Padronizada CE x Nota Padronizada FG

Espera-se encontrar uma correlação positiva, forte ou moderada, entre as notas de FG e CE, já que, em geral, estudantes que demonstram bom desempenho em áreas de Formação Geral tendem a aplicar essas habilidades de forma eficaz em seu Componente Específico.

Figura 15: FG e CE



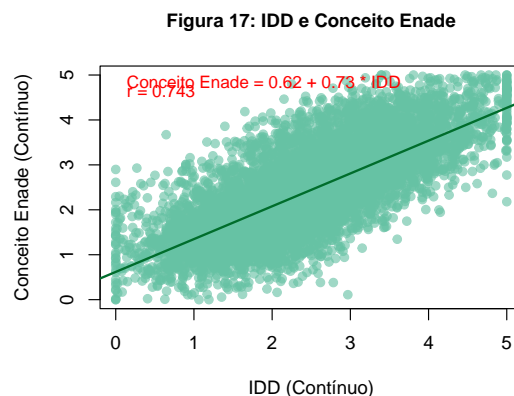
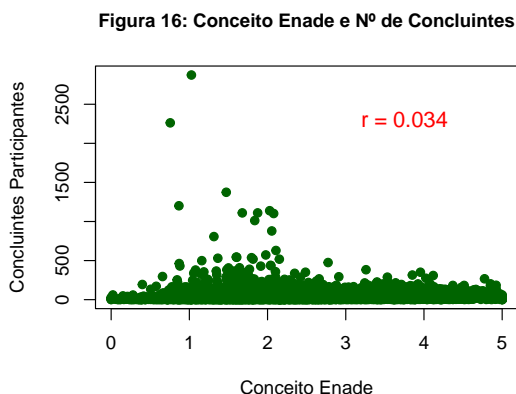
Ao observar o diagrama de dispersão (figura 15), nota-se imediatamente uma associação linear extremamente forte e positiva, onde a maioria dos pontos se agrupa firmemente em torno da linha de regressão, traçando uma trajetória ascendente clara.

O Coeficiente de Correlação de Pearson de 0.7135072 confirma que o sucesso no componente de habilidades gerais (FG) é um excelente preditor do desempenho do componente específico do curso (CE), de maneira que a cada aumento de 1 ponto na nota do componente de Formação Geral (FG), estima-se um aumento média de 0.7563 pontos na nota do Componente Específico (CE)

4.2. Conceito Enade x N° de Concluintes Participantes e IDD

Nesta parte, foi realizada duas análises com o conceito enade, sendo a primeira uma análise de correlação do conceito enade com número de concluintes participantes, em que pretende-se verificar que o número de participantes não influencia e não está relacionado fortemente com a qualidade do curso.

Já a segunda, foi realizada com o conceito enade e IDD, esperando que a correlação do IDD com o conceito enade, como uma variável preditora da qualidade do curso (conceito enade).



O Coeficiente de Correlação de Pearson (r) entre o Conceito Enade (Contínuo) e o N° de Concluintes Participantes é de 0.034, indicando uma correlação extremamente fraca, já que está muito próxima de zero. Ademais, o diagrama de dispersão da figura 16 mostra uma nuvem de pontos espalhadas, sem um padrão claro de crescimento ou decréscimo, sugerindo que não há nenhuma relação entre as duas variáveis.

Por outro lado, percebe-se uma associação positiva muito forte no diagrama de dispersão da figura 17, tendo em vista que o coeficiente de correlação de Pearson (r) é de 0.743458, indicando que os cursos com maior IDD tendem a alcançar um Conceito Enade mais elevado, de maneira que para cada aumento de 1 ponto no IDD Contínuo, o Conceito Enade aumenta em 0.73 pontos.

5. Conclusão

A partir das análises realizadas, notou-se que existe uma estrutura concentrada em um número limitado de categorias, como educação presencial, região sudeste e instituições privadas com fins lucrativos. Ademais, percebe-se que existe forte correlação entre Formação Geral e Componente Específico, de modo que estudantes com desempenho bom em FG tendem a aplicar essas habilidades em seu CE.

Sob outra ótica, foi comprovado que o número de participantes não afeta a qualidade de ensino avaliada ao contrário do IDD. Além disso, foi observado que fatores como modalidade de ensino e categoria administrativa possuem relação com a qualidade de ensino, como as categorias públicas com qualidade maiores e distribuições mais concentradas.