

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Variáveis Aleatórias

IVAN CUNHA VIEIRA

VITÓRIA, ES
2024

1 Introdução

O Sistema de Avaliação da Educação Básica (Saeb) é um sistema que tem, por objetivo, de acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), desencadear um processo de avaliação periódico de informações, permitindo diagnosticar e monitorar o quadro educacional brasileiro, observando diversos fatores, como o desempenho dos alunos por área de conhecimento, formação e adequação profissional dos professores e condições de trabalho, infraestrutura das escolas e níveis socioeconômicos dos alunos. Dessa forma, a observação e análise dos dados coletados pelo Saeb procura, portanto:

- Avaliar a qualidade e eficiência da educação no país;
- Produzir indicadores educacionais para o Brasil;
- Elaborar, monitorar e aprimorar políticas públicas com base nas evidências observadas.

A avaliação é feita bianualmente, com turmas de 2º ano, 5º ano e 9º ano do ensino fundamental (EF), além de turmas da 3ª série e 4ª série do ensino médio (EM).

2 Contextualização

O conteúdo analisado no presente trabalho foi retirado da planilha *TS_ESCOLA* do conjunto de microdados coletado pelo Saeb e disponibilizado pelo Inep. Esta planilha abriga os resultados dos valores médios de proficiência nas provas de matemática e língua portuguesa dos alunos das escolas e turmas participantes da coleta de dados dos 5º anos, 9º anos e séries finais do ensino médio (integrado e tradicional). Apresenta também a taxa de participação dos alunos na realização das provas, além da taxa dos alunos participantes presente em cada nível da escala de proficiência do Saeb para tais áreas de conhecimento. Juntamente a isso, a planilha também traz as regiões e estados das escolas, assim como se elas se encontram em área urbana ou rural e em capitais ou no interior dos estados, os respectivos níveis socioeconômicos apresentados pelos alunos e se são escolas públicas ou privadas.

Para a publicação dos resultados do Saeb, são considerados alguns critérios, sendo eles:

- Registro de, no mínimo, dez estudantes presentes no momento da aplicação dos instrumentos;
- Taxa de participação de, pelo menos, oitenta por cento dos estudantes matriculado.

Para a análise feita neste trabalho, foram objetos de estudo as médias de proficiência dos anos iniciais e finais do ensino fundamental e as séries finais do ensino médio, relacionando seus dados aos níveis socioeconômicos das escolas que realizaram os testes às suas devidas regiões do Brasil. Após uma verificação das variáveis, constatou-se que não há a presença de escolas privadas nos dados, descartando assim a necessidade de se estudar essa variável.

Os níveis socioeconômicos são divididos entre 8 níveis, indo do 1 ao 8, e têm, como objetivo, contextualizar resultados obtidos em avaliações e exames aplicados pelo Inep, possibilitando o conhecimento sobre a realidade social das escolas e redes de ensino do país. Esse conhecimento viabiliza o cumprimento do que é almejado pelo Saeb através desses indicadores.

Sobre as médias, elas são definidas dos níveis 1 ou 0 (quando os testes contam com itens avaliativos para atendimento especial), até os níveis 8, 9 ou 10, variando entre as escolaridades avaliadas, sendo que os primeiros níveis dizem respeito aos conceitos mais básicos da respectiva área de conhecida naquela escolaridade.

3 Análises

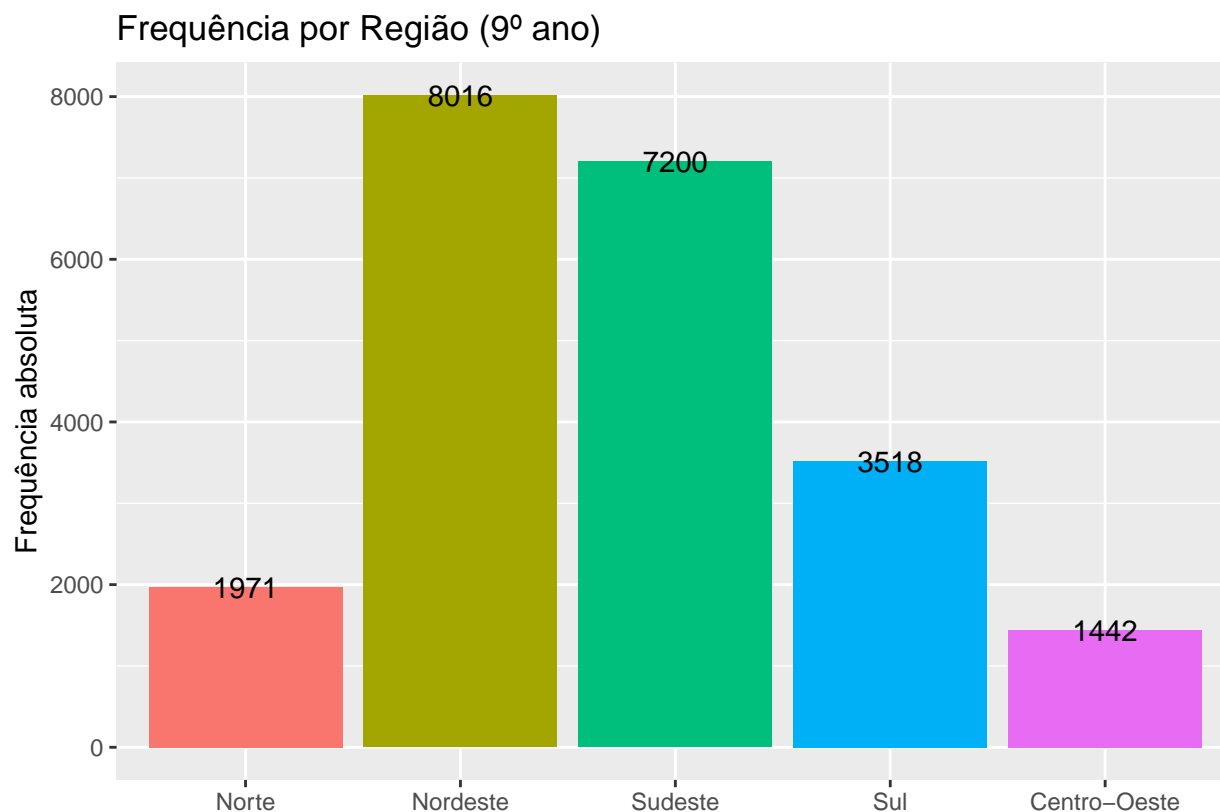
Partindo para a análise do banco de dados de fato, é importante observar, inicialmente, quais são os tipos de variáveis e fazer um estudo separado de cada uma pelos seus tipos.

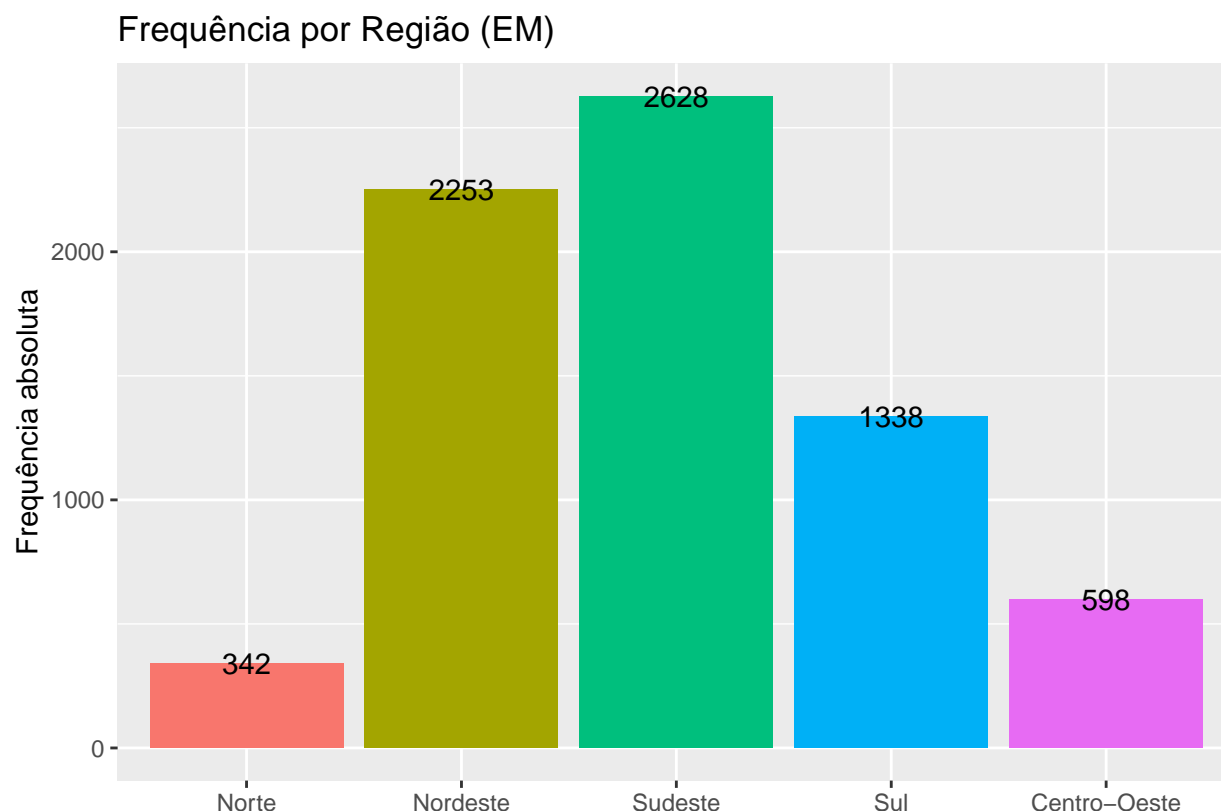
3.1 Variáveis Qualitativas

Entre as variáveis escolhidas para estudo neste trabalho, são qualitativas as que trazem informações sobre:

- Região (Nominal);
- Nível Socioeconômico (Ordinal).

Para esses tipos de variáveis, é interessante representá-las por gráficos de barras para procurar entender seus comportamentos. Começando com as regiões, obtiveram-se os seguintes gráficos:

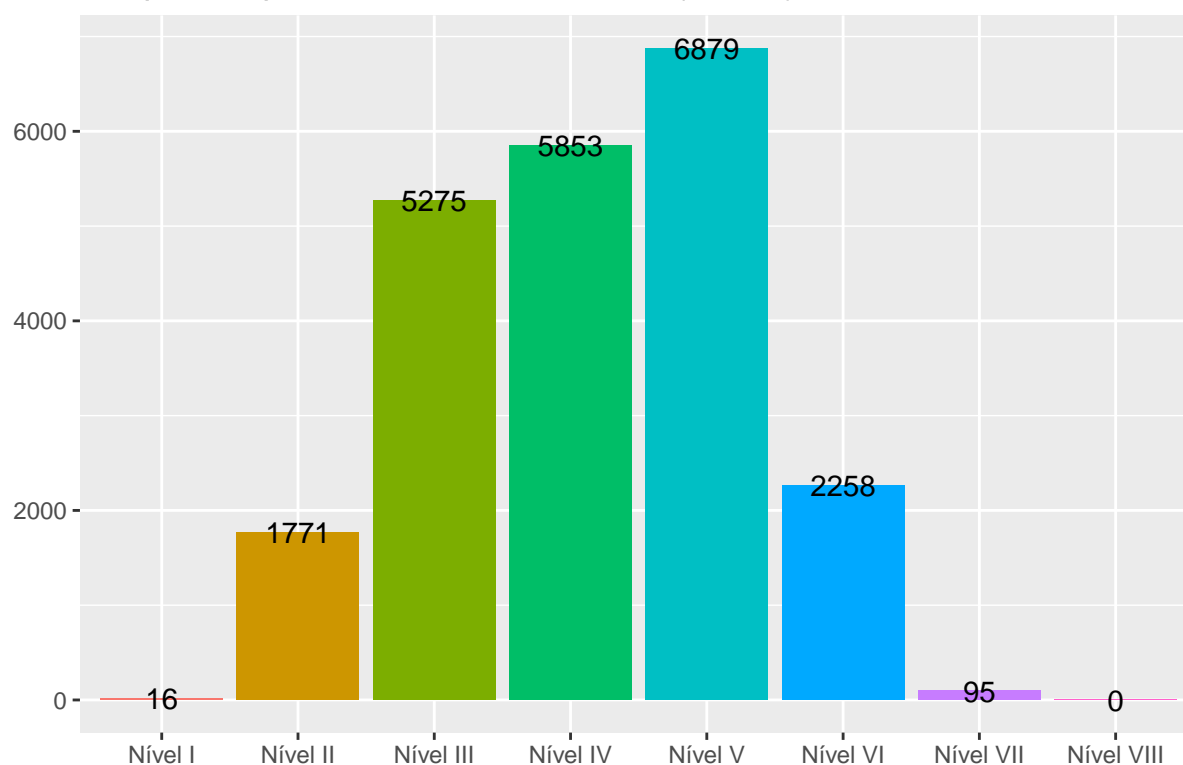




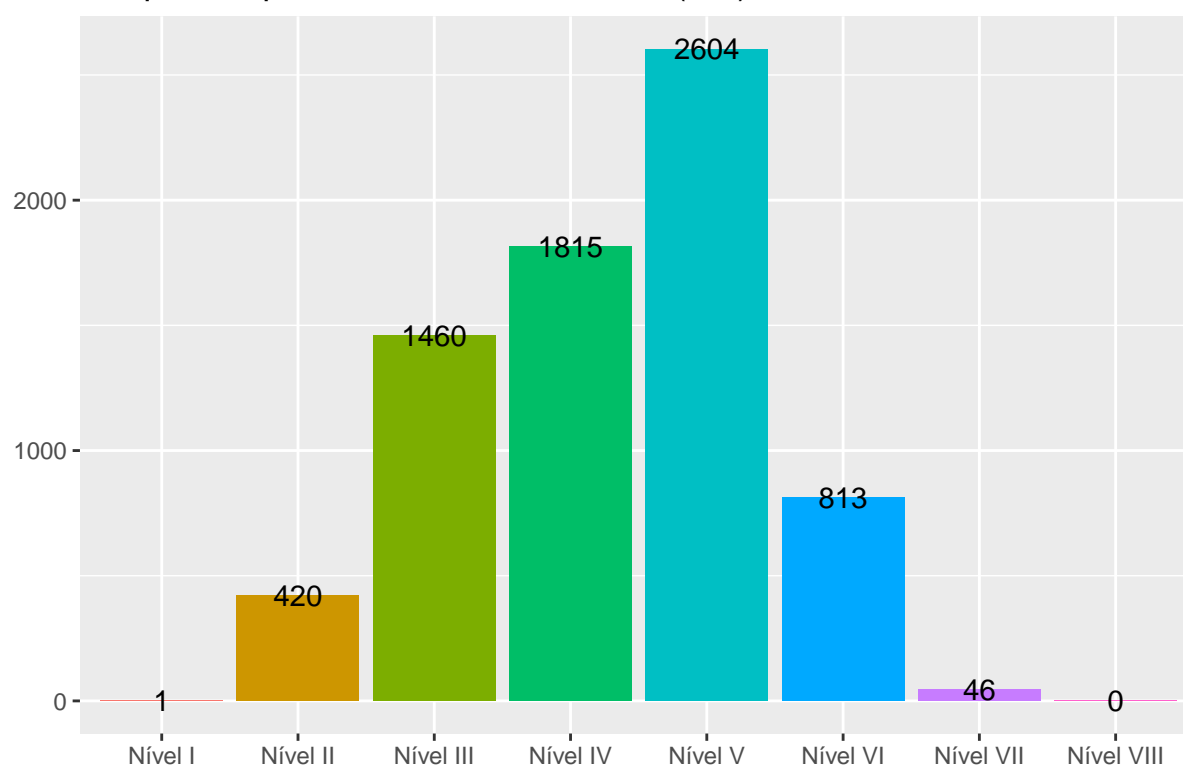
A partir dos gráficos acima, é possível perceber que, tanto para o ensino médio, quanto para os 9º anos, os testes feitos pelo Saeb se localizaram, principalmente, nas regiões Nordeste e Sudeste. Totalizaram juntas, dessa forma, 4881 testes para o ensino médio (68,7% do total) e 15216 testes para os 9º anos (68,2% do total). Dessa forma, entende-se que mais de metade dos resultados obtidos vieram dos testes feitos nessas regiões. Esses dados são equivalentes à relação populacional geral dos estados. Abaixo, segue tabela com o percentual entre Nordeste e Sudeste com relação ao total de dados.

	Regiao_EF	Regiao_EM
Norte_Sudeste	15216.00000	4881.00000
Total	22147.00000	7159.00000
Percentual	68.70457	68.17991

Frequência por Nível Socioeconômico (9º ano)



Frequência por Nível Socioeconômico (EM)



No estudo dos níveis socioeconômicos segue-se, novamente, a mesma tendência para os 9º anos e para o ensino médio, de tal forma que ambos apresentam como medida de tendência

central (suas modas) o nível socioeconômico V, de tal forma que este se apresenta em cerca de um terço dos casos em ambos os gráficos. Abaixo, tabela com as relações entre as relações de frequência do Nível I com a frequência total.

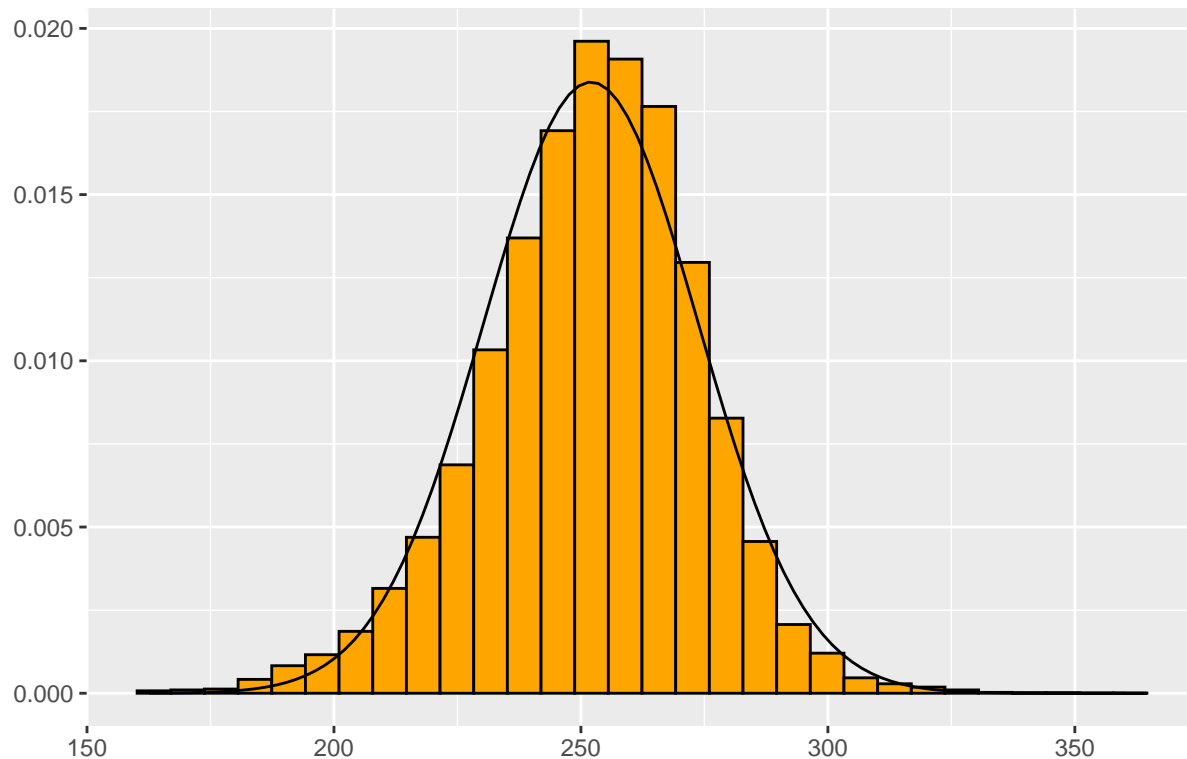
	Nv_EF	Nv_EM
Níve I	6879.00000	2604.0000
Total	22147.00000	7159.0000
Percentual	31.06064	36.3738

3.2 Variáveis Quantitativas

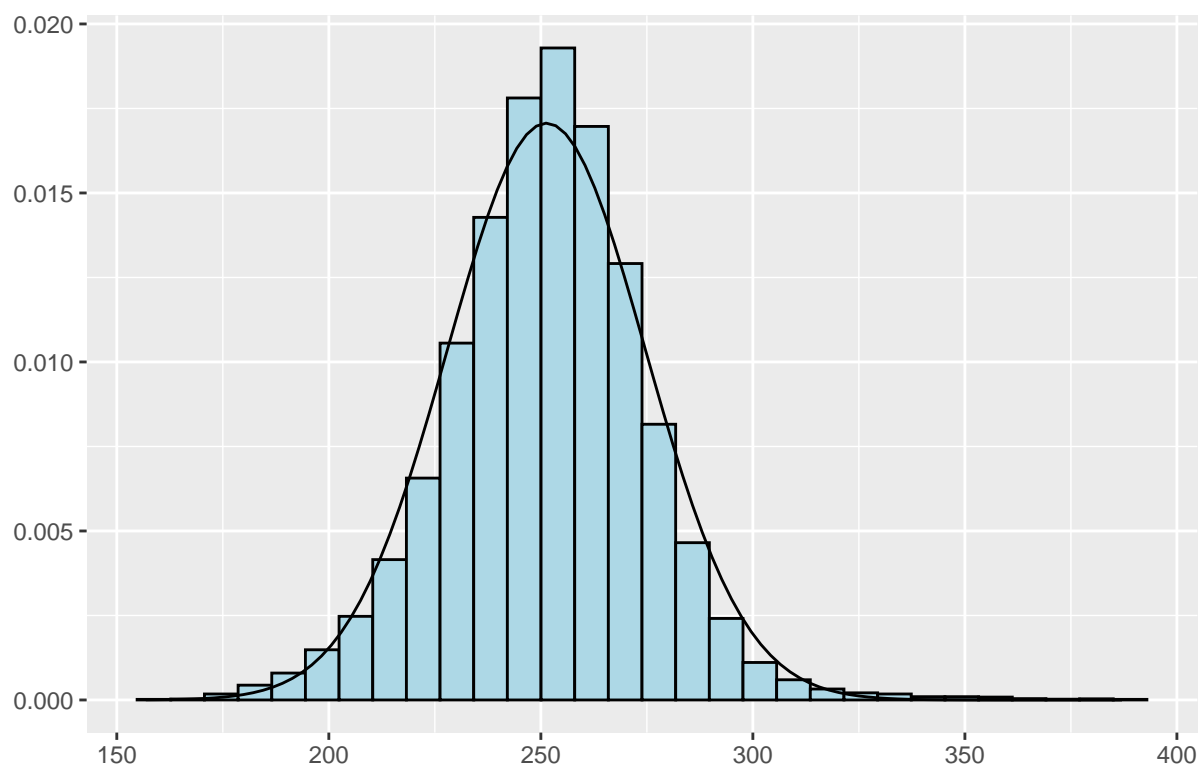
As variáveis quantitativas estudadas foram justamente as médias de proficiência dos alunos em matemática e em língua portuguesa. procurou-se, primeiramente, entender a distribuição dessas notas e, posteriormente, entender a distribuição dessas notas nas escalas de proficiência do Saeb.

Como o número de amostras é muito grande, é esperado que os dados tendam à normalidade. Dessa forma, uma boa maneira de observar esses comportamentos é através de histogramas. Abaixo, seguem os gráficos relativos às distribuições de médias para as duas áreas de conhecimento e para os seus respectivos anos ou séries:

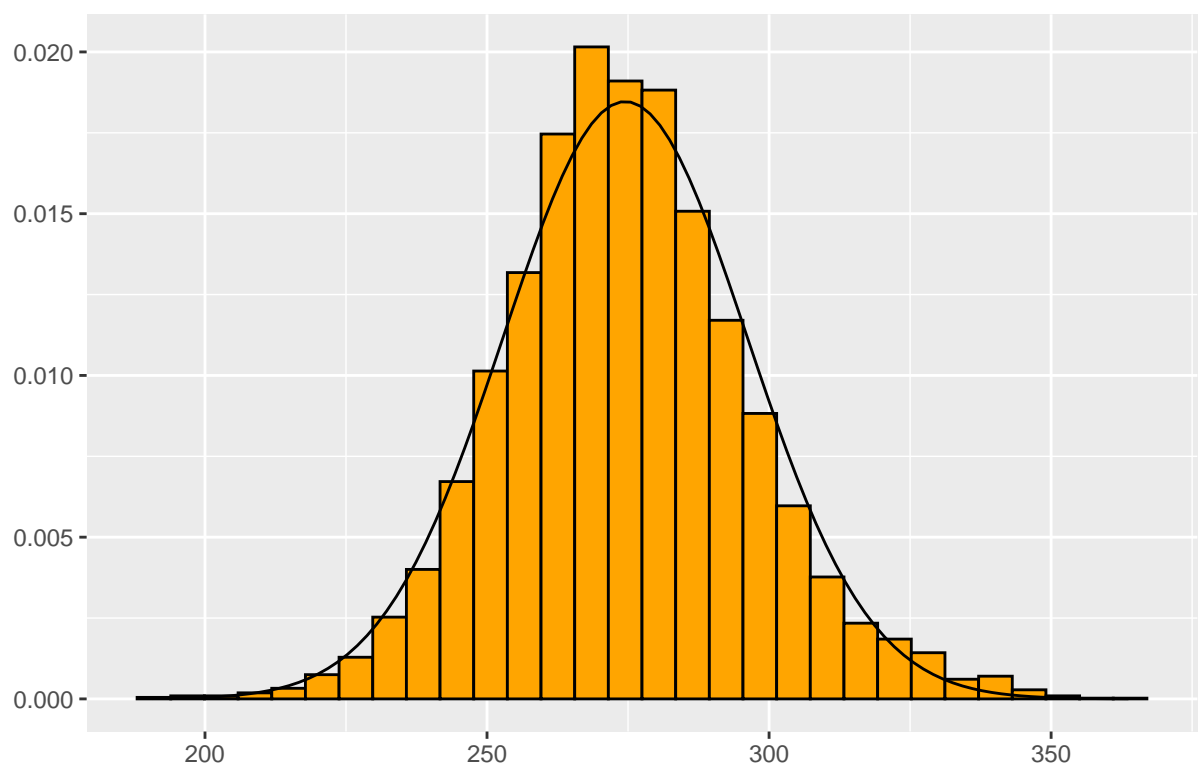
Notas 9º ano em L.P.

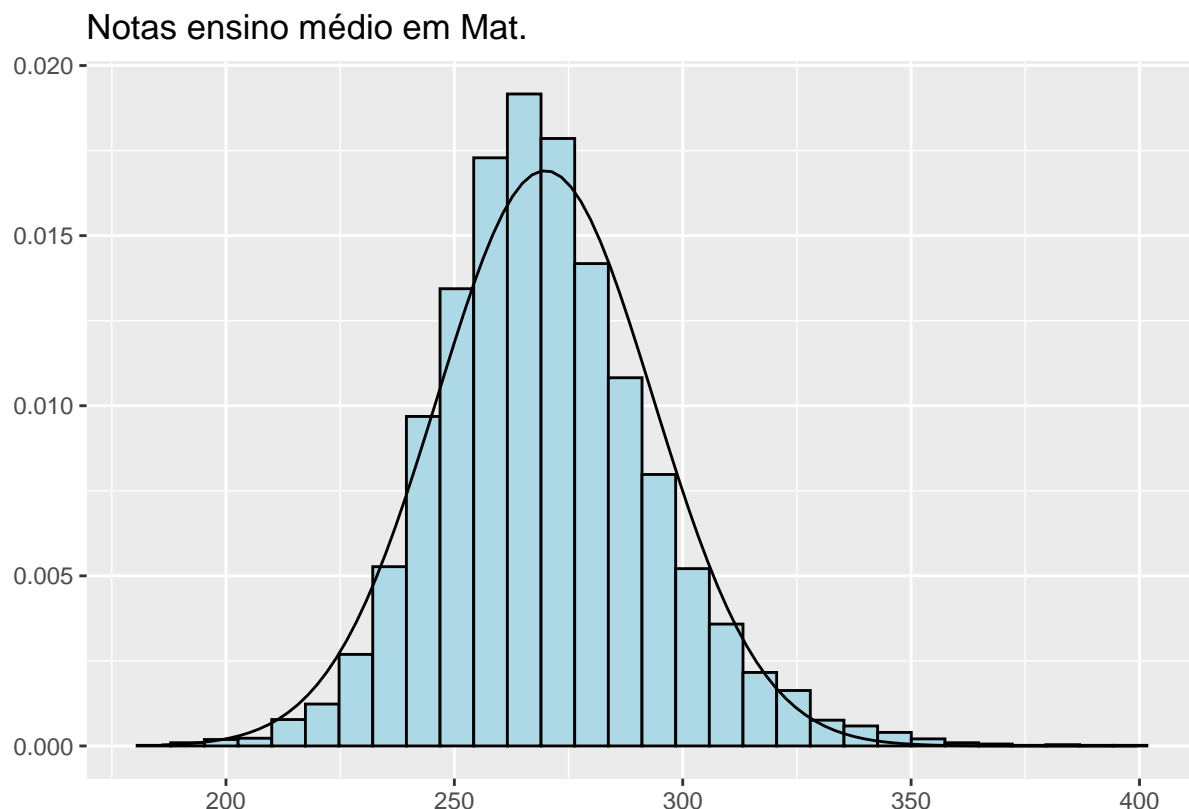


Notas 9º ano em Mat.



Notas ensino médio em L.P.





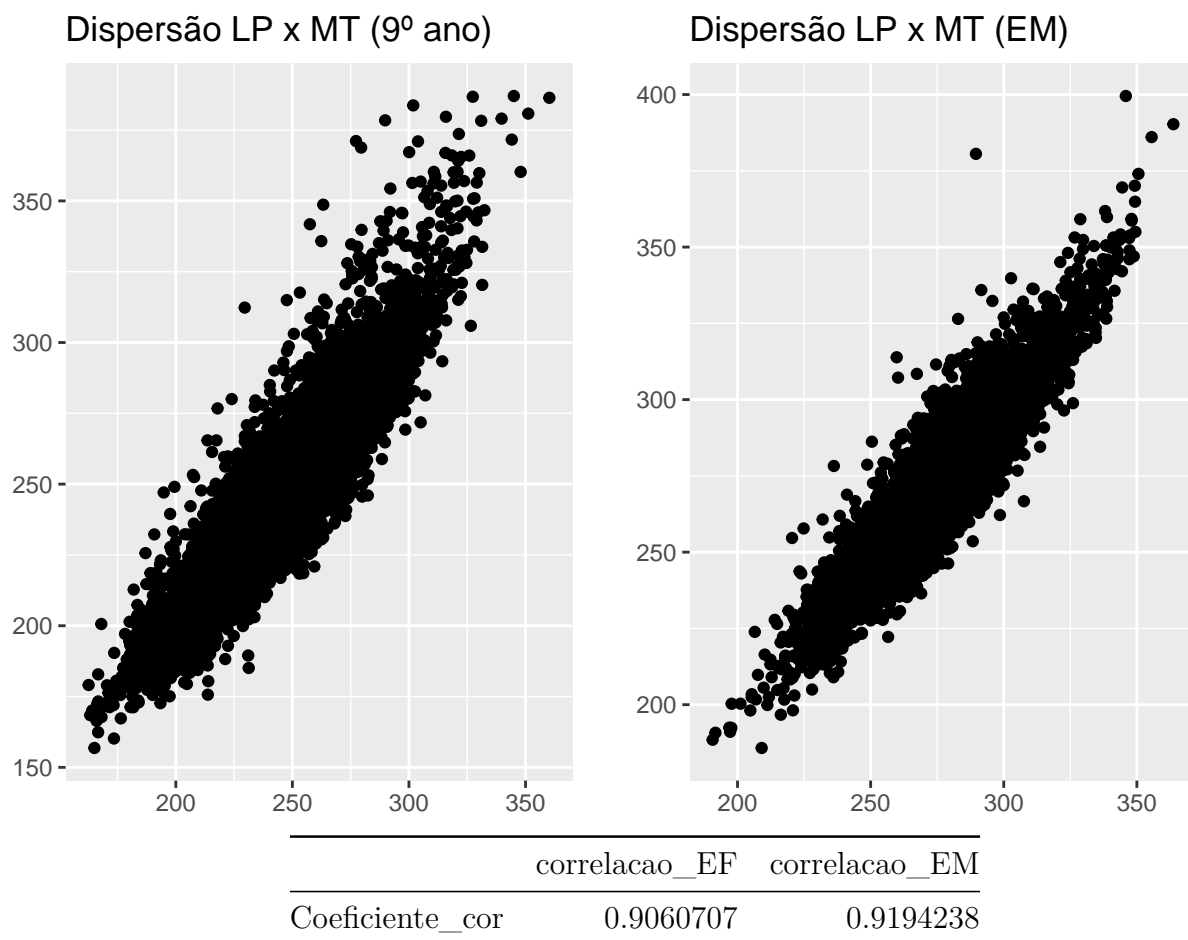
Juntamente às curvas normais, com médias e desvios padrões baseados nas notas de cada variável, se torna perceptível o comportamento semelhante à normalidade, apresentado tanto pelas notas dos 9º anos quanto pelas notas do ensino médio. Abaixo, a tabela indicando os valores de média e desvio padrão das curvas normais desenhadas nos gráficos:

	medidas_9EF_LP	medidas_9EF_MT	medidas_EM_LP	medidas_EM_MT
Média	252.02452	251.31090	274.49664	269.87883
Desvio Padrão	21.69724	23.37602	21.60278	23.59051

Neste ponto, após compreender como as variáveis funcionam separadamente, é interessante procurar como elas funcionam em conjunto, através de correlações entre as variáveis e devidas repetições em seus comportamentos.

3.3 Relação entre as médias

Como mostra a tabela abaixo, o cálculo de coeficiente de correlação (teste de Pearson) resulta em valores acima de 0,9. De acordo com Hinkle, Wiersma & Jurs (2003), é sugerido que valores maiores ou iguais a 0,9 indicam que há evidências estatísticas suficientes para dizer que a correlação entre as duas variáveis é muito forte. Através do gráfico de dispersão também fica evidente o tipo de correlação entre as variáveis, de tal forma que, conforme um cresce, o outro cresce juntamente, demonstrando assim que as variáveis possuem influência significativa uma para com a outra.



3.4 Relação entre as médias e as variáveis qualitativas

Uma forma interessante de verificar esse tipo de correlação é a partir de gráficos, como os gráficos de *boxplot*. A partir desse tipo de gráfico, é possível separar as médias (variáveis quantitativas) pelos grupos definidos nas variáveis qualitativas.

3.4.1 Médias do nono ano

Os gráficos abaixo apresentam visualmente a relação entre as médias dos nonos anos com os níveis socioeconômicos e com as regiões.

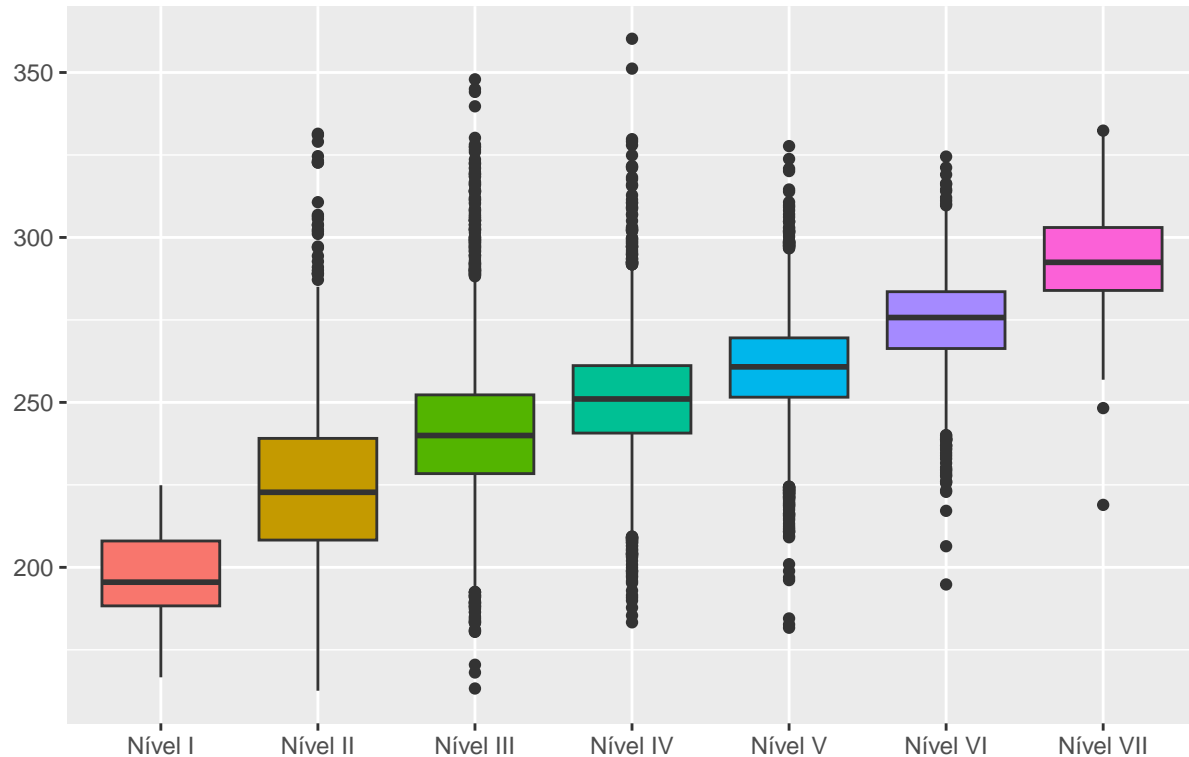
Os dados apresentados abaixo demonstram crescentes nas médias de acordo com o crescimento dos níveis socioeconômicos. Esse tipo de relação era esperada, uma vez que maiores níveis socioeconômicos têm a tendência de proporcionar melhores condições de vida e oportunidades, como maior acesso à informação, tempo disponível para se dedicar ao estudo e até mesmo apoio familiar. Infere-se, dessa forma, que os dados possuem uma correlação positiva.

Sobre a relação entre as médias e as regiões, não foi demonstrada, visualmente, algum tipo de relação, pois as regiões possuem médias próximas. Esse tipo de comportamento é esperado,

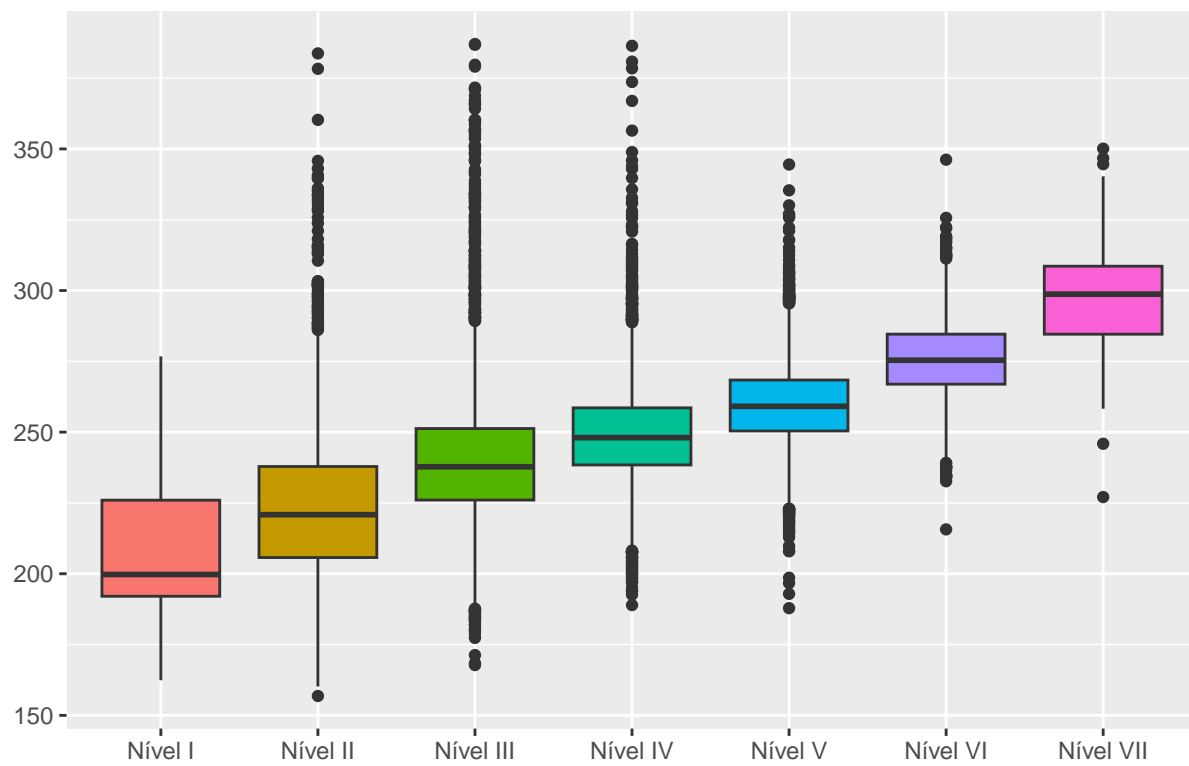
pois as diferentes regiões do Brasil não devem possuir relação entre si, uma vez que são independentes umas das outras. No entanto, percebe-se que alguns estados possuem médias maiores. Como um dos objetivos do Saeb é justamente produzir indicadores da educação no país, esse tipo de informação pode servir para outras análises descobrirem as causas dessas diferenças.

As observações feitas acima dizem respeito tanto aos dados dos 9º anos quanto aos dados do ensino médio. Seguem os gráficos descritos:

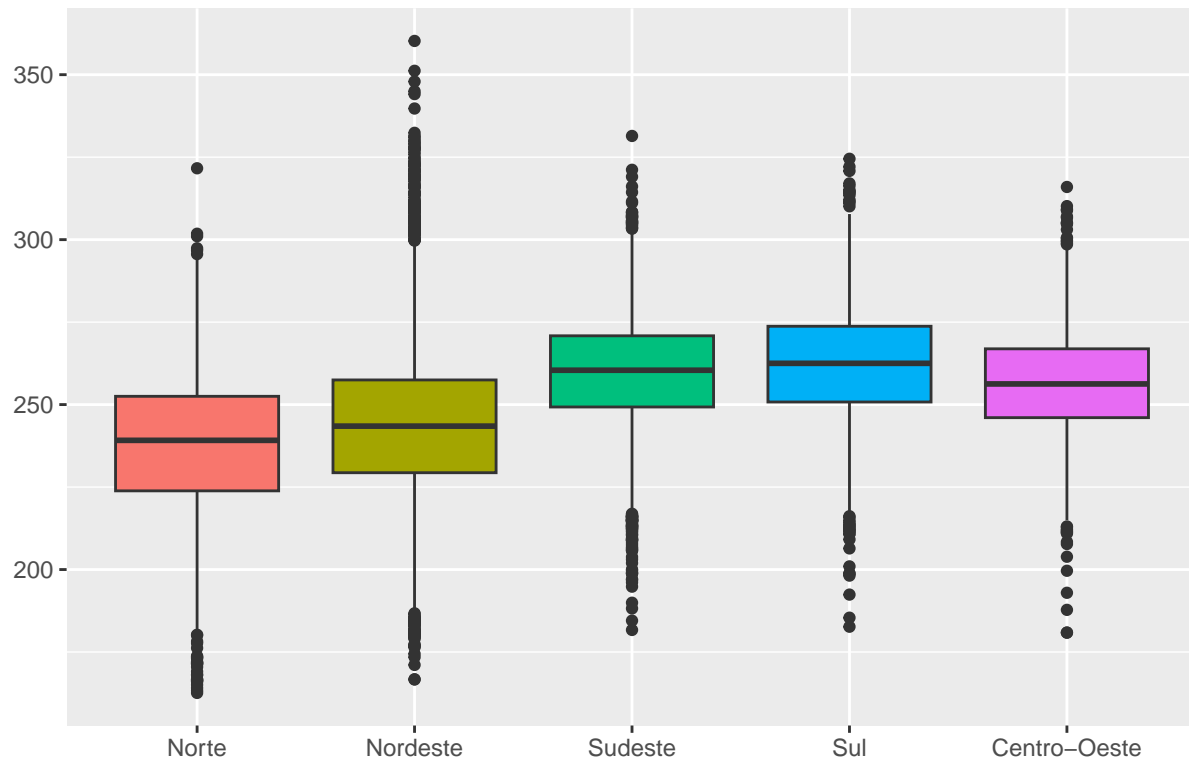
Médias x Níveis Socioeconômicos (LP)

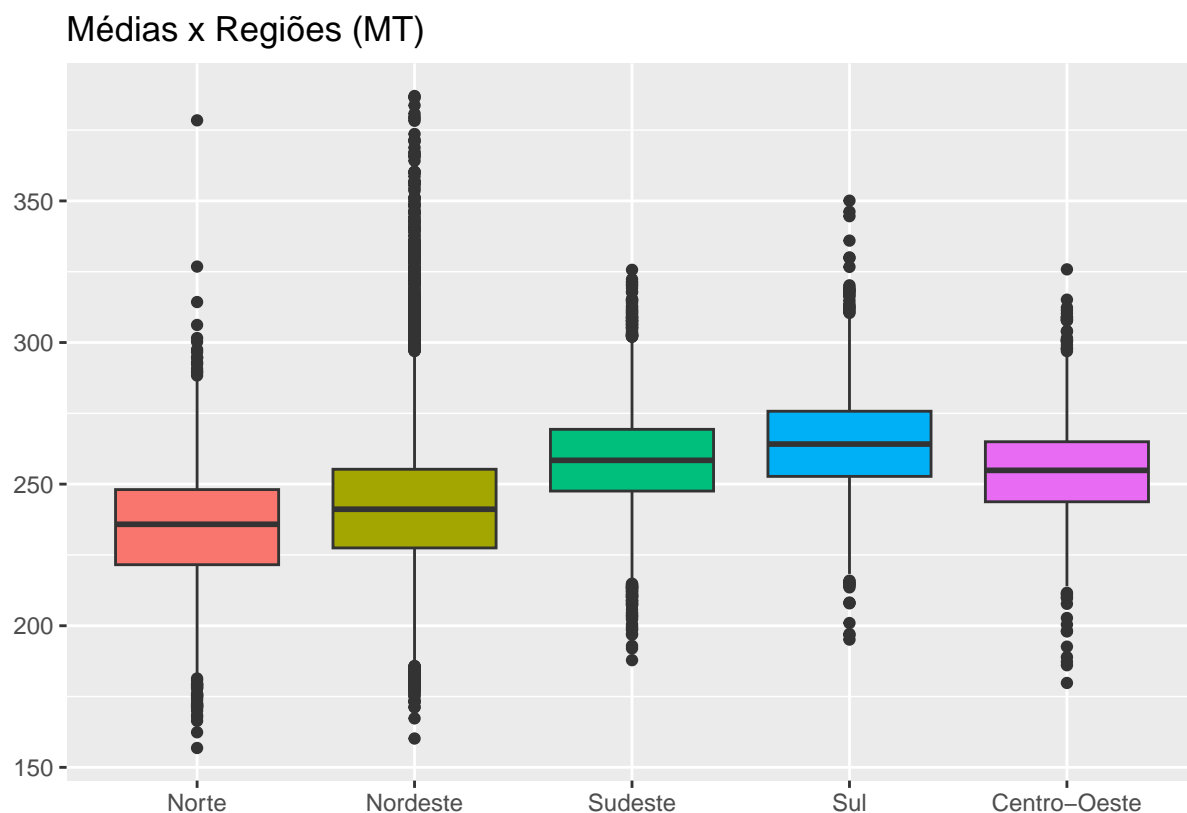


Médias x Níveis Socioeconômicos (MT)



Médias x Regiões (LP)

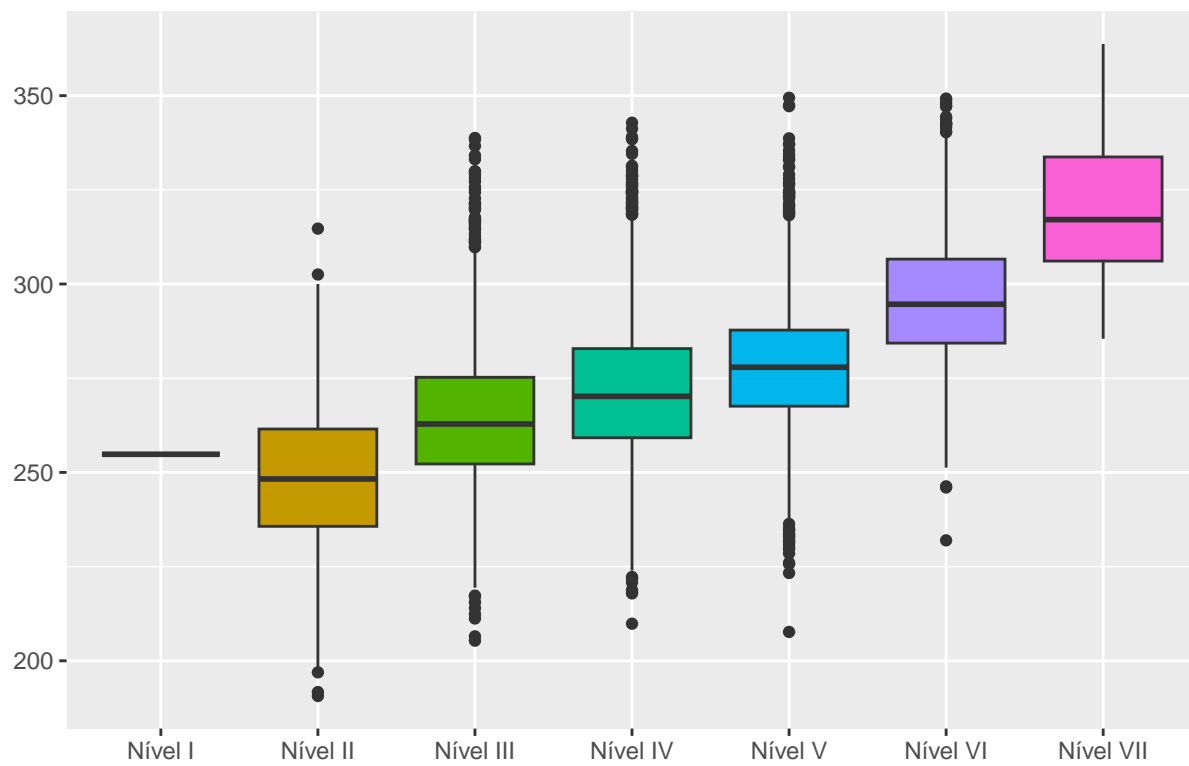




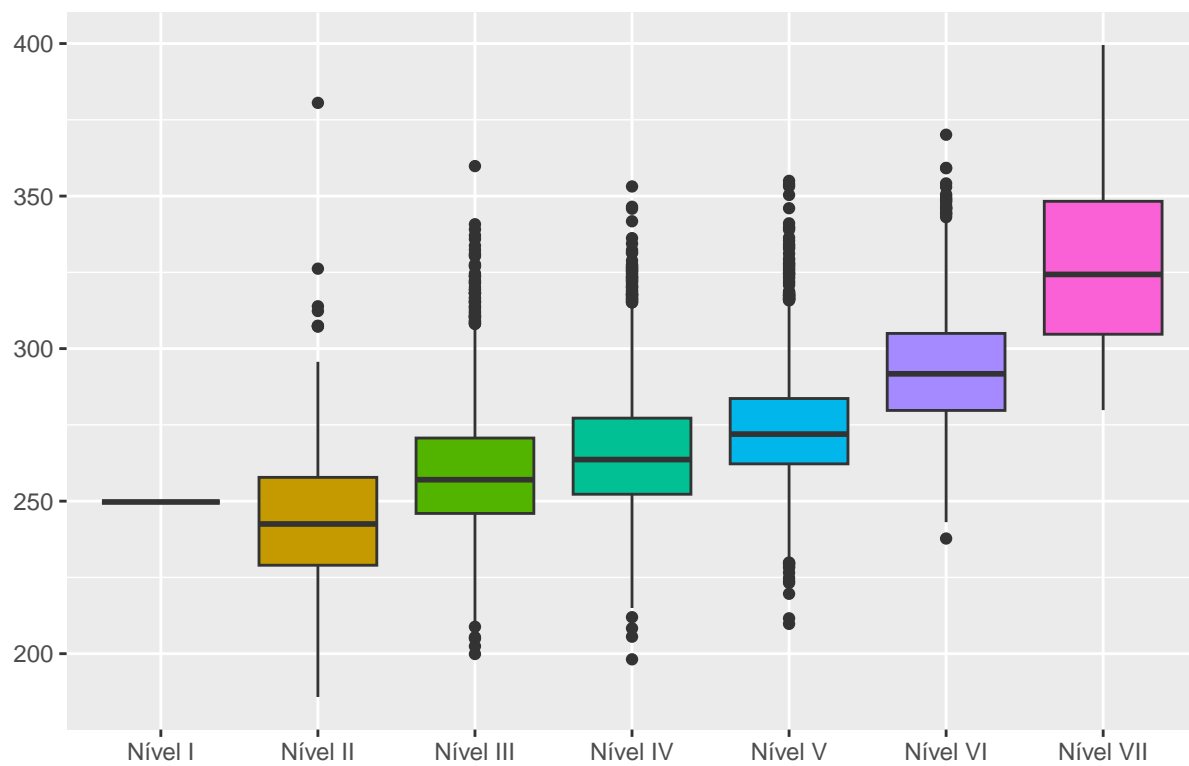
3.4.2 Médias do EM

As observações feitas anteriormente sobre os dados e gráficos dos 9º anos servem igualmente para o ensino médio, o que também é esperado, pois, independente da escolaridade, todas as variáveis apresentaram comportamentos semelhantes. Abaixo, os gráficos referidos:

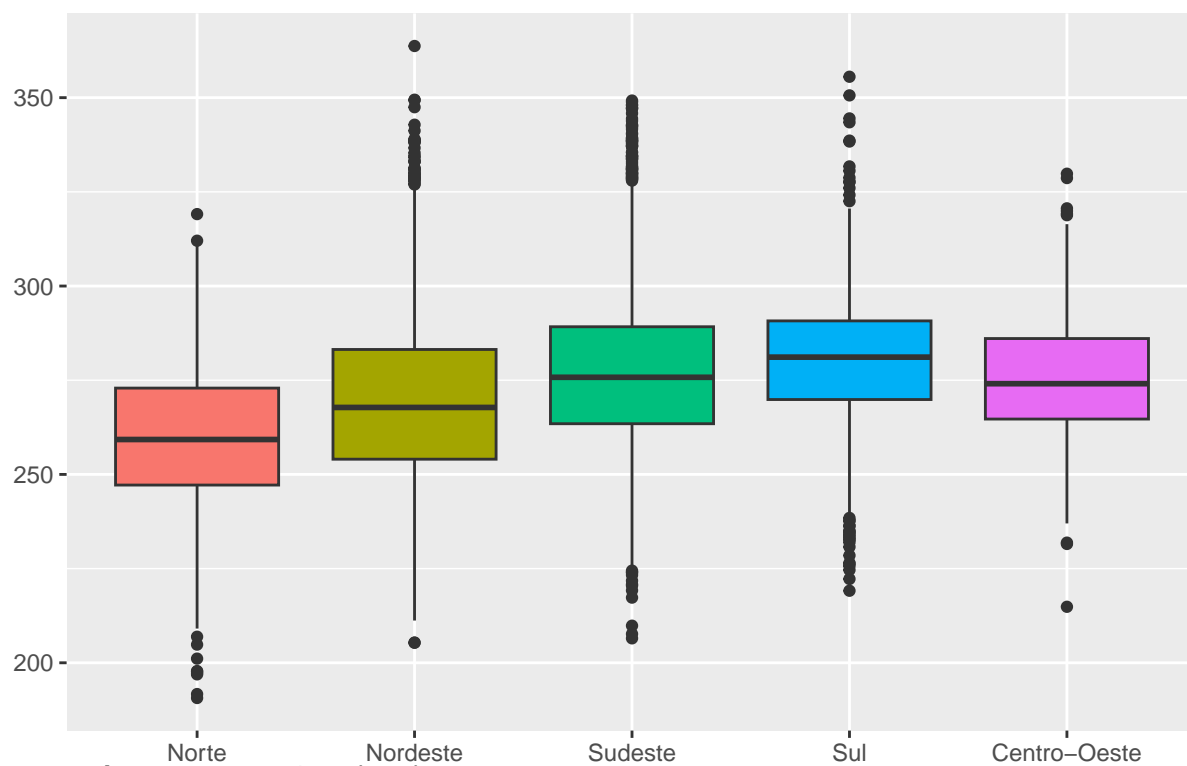
Médias x Níveis Socioeconômicos (LP)



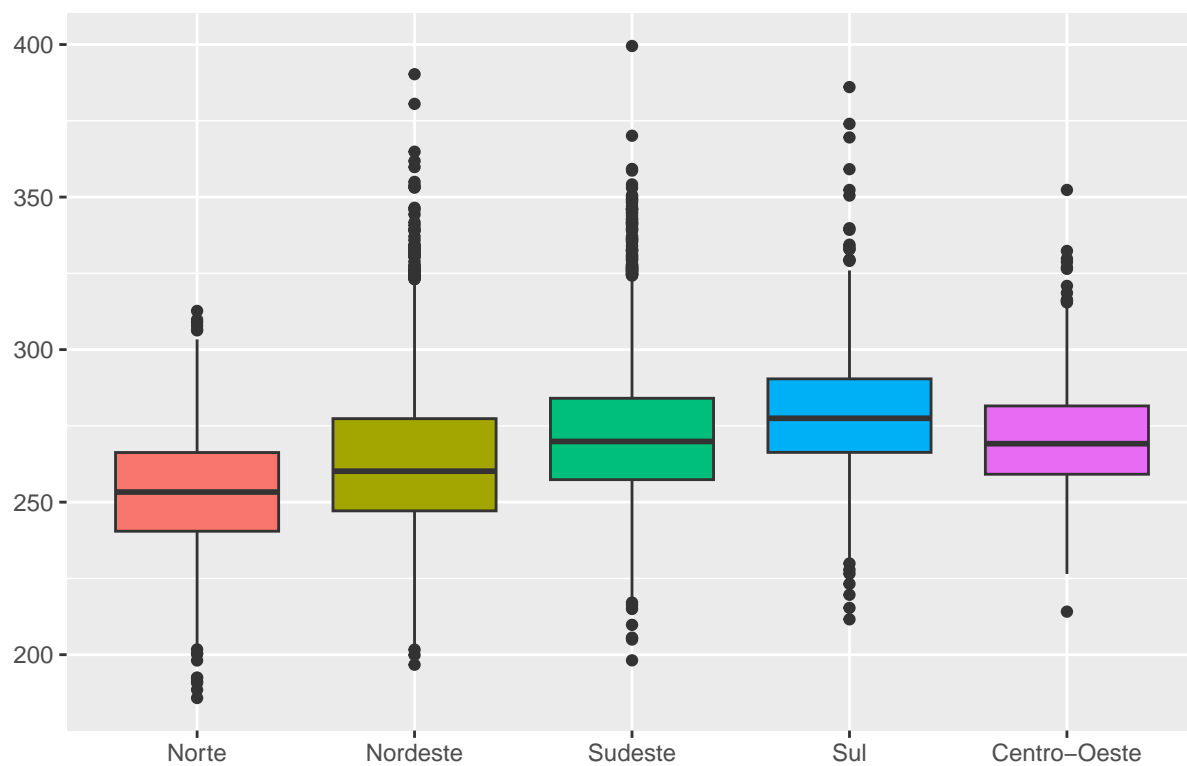
Médias x Níveis Socioeconômicos (MT)



Médias x Regiões (LP)



Médias x Regiões (MT)



4 Conclusões

Os dados analisados trouxeram evidências sobre as relações entre os indicadores socioeconômicos e as médias alcançadas pelos estudantes em ambas as áreas, de forma que as variáveis progrediam igualmente juntas. Além do mais, também foi identificada relações entre as médias em si, tal que ambas também progrediam igualmente. O comportamento das variáveis isoladamente também trouxe informações importantes para suas correlações, pois assim foi identificado que tanto o ensino médio, quanto o ensino fundamental, se comportam de maneira semelhante. Isso também foi possível de se interpretar através dos gráficos e tabelas, que sempre indicavam, visualmente, essas semelhanças.

A partir dos resultados e análises apresentados, percebe-se que os objetivos do Saeb, de monitoramento, avaliação e produção de indicadores educacionais e socioeconômicos, são bem atendidos, uma vez que, mesmo analisando uma pequena parcela das variáveis disponibilizadas pelo banco de dados, é possível retirar informações relevantes sobre o comportamento de cada uma e, ainda por cima, encontrar correlações entre elas e de quais formas elas se influenciam.

Faz-se necessário, portanto, que mais análises aprofundadas sejam feitas a partir destes dados, e que elas sirvam para o melhoramento da educação e das políticas públicas que circulam toda a área educacional.

5 Referências

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados do Saeb 2020. Brasília: Inep, 2021. Disponível em: (<https://www.gov.br/inep/pt-br/aceso-ainformacao/dados-abertos/microdados/saeb>)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Nota Técnica nº 020/2014/CGCQTI/DEED - Indicador de adequação da formação do docente da educação básica. Brasília, 2014. Disponível em: (http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_formacao_legal/nota_tecnica_indicador_docente_formacao_1)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Nota Técnica nº 4/2020/CGCQTI/DEED - Indicador de adequação da formação do docente da educação básica. Brasília, 2020. Disponível em: (http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_formacao_legal/nota_tecnica_indicador_docente_formacao_1)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Nota Informativa do Ideb 2021. Disponível em: (https://download.inep.gov.br/educacao_basica/portal_ideb/planilhas_para_download/2021/nota_informativa_ideb_2021.pdf)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Escalas de proficiência do SAEB. Brasília, DF: INEP, 2020. Disponível em: (https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/escalas_de_proficiencia_do_saeb.pdf)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Saeb 2021: Indicador de Nível Socioeconômico do Saeb 2021: nota técnica. Brasília, DF: Inep, 2023. Disponível em: (https://download.inep.gov.br/areas_de_atuacao/Indicadores_de_nivel_Nota_tecnica_2021.pdf)

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Memorando Técnico - RELATÓRIO DA AMOSTRAGEM DO SAEB 2021. Brasília, 2022. Disponível em: (<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/>) saeb

CABRAL, Umberlândia. De 2010 a 2022, população brasileira cresce 6,5% e chega a 203,1 milhões. **IBGE**, 28/06/2023. Censo 2022. Disponível em: (<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37237-de-2010-a-2022-populacao-brasileira-cresce-6-5-e-chega-a-203-1-milhoes>)

HINKLE, D. E., WIERNSMA W & JURS S. G. Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin, 2003.