

Análise descritiva dos dados

A análise dos dados do MAIS.edu está atualmente baseada em 15 tipos diferentes de indicadores considerados importantes pelos auditores do TCE-PB e estes indicadores foram monitorados ao longo de 6 anos (2000, 2007, 2008, 2009, 2010 e 2011).

É importante citar que não existem dados para certos indicadores em alguns anos, e isto ocorre porque estes não foram coletados ou não foram disponibilizados para nossa base de dados. Descartamos da nossa análise estes anos para certos indicadores. Desta forma, procuramos analisar os indicadores anualmente e verificar se estes são provenientes de populações que seguem distribuições normais, para isto aplicamos o teste de adequação à normalidade Shapiro-Wilk a fim de definir quais são as medidas de tendências centrais mais adequadas a serem usadas. Além disto procuramos correlacionar os indicadores dois a dois por ano, como veremos mais adiante.

Na tabela 1 abaixo, encontramos um sumário que correlaciona os índices dos indicadores aos seus respectivos nomes. Os índices são usados para minimizar o espaço ocupado.

Índice do Indicador	Nome do Indicador
INDICADOR_62	Participação da despesa com pessoal e encargos sociais na função educação(%)
INDICADOR_89	IDEB - 5ºano do ensino fundamental
INDICADOR_90	IDEB - 9ºano do ensino fundamental
INDICADOR_329	Taxa de analfabetismo(%)
INDICADOR_333	Taxa de atendimento escolar(%)
INDICADOR_73	Taxa abandono escolar- ensino fundamental(%)
INDICADOR_74	Taxa de abandono escolar - ensino médio(%)
INDICADOR_80	Taxa de aprovação- ensino fundamental(%)
INDICADOR_81	Taxa de aprovação - ensino médio(%)
INDICADOR_176	Percentual docentes com formação superior(%)
INDICADOR_177	Percentual de docentes temporários(%)
INDICADOR_202	Índice de precariedade infraestrutura(%)
INDICADOR_184	Razão aluno por docente(Qtd. Alunos por Docente)
INDICADOR_7	Despesa educação por aluno(em Reais)
INDICADOR_201	Índice de eficiência da educação básica(%)

Tabela 1: sumário dos índices e nomes dos indicadores

Teste de normalidade dos valores dos indicadores anualmente

Na análise de normalidade dos dados, aplicamos o teste de adequação a normalidade Shapiro-Wilk, pois é um teste poderoso e largamente utilizado. É importante destacar que estes testes foram aplicados com nível de significância de 5%.

Na tabela 2, as linhas representam os anos e as colunas representam os números dos indicadores sumariados, que estão descritos na tabela 1. Os valores nesta tabela demonstram se para cada ano, os indicadores seguem distribuição normal (N), não seguem distribuição normal (NN) e não obtivemos dados para determinado indicador em certo ano (NA).

	329	62	89	90	333	73	74	80	81	176	177	202	184	7	201
2000	N	NA	NA	NA	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2007	NA	N	N	NN	NA	NN	NN	NN	N	NN	NA	NN	NN	NN	NA
2008	NA	NN	NA	NA	NA	NN	NN	N	N	NN	NA	NN	NN	NN	NA
2009	NA	NN	N	N	NA	NN	NN	NN	NN	NN	NA	NN	N	NN	NN
2010	N	NN	NA	NA	NN	NN	N	N	NN	NN	NA	NN	N	NN	NA
2011	NA	N	NN	N	NA	NN	NN	N	NN	NN	N	NN	NN	NN	NN

Tabela 2: tipo de distribuição dos valores dos indicadores por ano

Estes valores encontrados na tabela 2 são importantes, pois na nossa análise no software MAIS edu é importante identificar valores atípicos e para tal faz-se necessária a escolha de um índice de tendência central mais adequado e de um método para identificação de valores atípicos. Assim, conhecer o formato dos dados em termos de simetria é importante, assim como as medidas de tendências centrais para estes indicadores.

Os valores destacados em vermelho na tabela 2 indicam os testes de Shapiro em que *p-values* ficaram abaixo do nível de significância (5%), ou seja, rejeitam a hipótese de que os dados vêm de populações que seguem uma distribuição normal. Dos 58 indicadores para os quais existem dados coletados, aproximadamente 69% desviam-se da distribuição normal e quanto aos 31% restantes podemos afirmar que seguem esta distribuição. A análise dos histogramas e gráficos quantil-quantil indicaram de fato que nem todos os indicadores provem de distribuições normais, porém não foram encontradas caudas muito longas e por consequência a média e a mediana dos indicadores são muito próximas.

Medidas de tendência central

Para nossa análise de medidas de tendência central, utilizamos o pressuposto de que se os dados seguem distribuição normal, a média é uma medida representativa de tendência central para a amostra, caso contrário a mediana se adequa melhor que a média, já que o viés pode distorcer estes resultados.

A tabela 3 demonstra os valores representativos de tendência central para cada indicador anualmente. Os valores destacados em vermelho são valores de medianas daquelas amostras, os valores em preto são valores de médias e os NAs são indicadores dos quais não obtivemos dados para aquele ano.

[illegible]

2007	NA	67.3	3.39	2.7	NA	10.5	16.2	74.4	77.03	70.7	NA	36.1	19.3	1719.2	NA
2008	NA	66.1	NA	NA	NA	10.4	15.7	75.1	76.1	67.2	NA	40.3	17.6	1959.3	NA
2009	NA	70.6	3.63	2.87	NA	9.7	17.1	75.5	76.4	68.5	NA	38.9	17.8	2092.02	0.72
2010	31.3	69.9	NA	NA	92.2	9.8	16.9	75.03	76.7	66.7	NA	38.1	17.4	2356.8	NA
2011	NA	68.6	3.9	2.92	NA	8.7	15	78.4	78.9	77.8	64.3	35.1	16.7	2674.9	0.6

Tabela 3: medidas de tendência central dos indicadores por ano

Correlações entre os indicadores por ano

Esta análise foi feita correlacionando, dois a dois, os indicadores para cada ano, e os valores nas tabelas abaixo representam o valor das correlações para o teste de correlação de Kendall. Os valores de correlação estão dentro do intervalo [-1, 1]. Valores próximos de 1 indicam correlação forte, quanto maior um indicador maior será o outro indicador. E para o valores próximos -1, quanto maior o valor de um indicador menor será o outro indicador.

Cada tabela abaixo corresponde a um ano e estão sendo exibidos apenas os valores que apresentaram correlações acima ou igual a 0,3 e abaixo ou igual a -0,3. Os indicadores que não alcançaram estes valores foram omitidos, e para o ano 2000 nenhum indicador obteve estes valores.

Indicador X	Indicador Y	Valor da Correlação
IDEB - 9ºano do ensino fundamental	Taxa abandono escolar- ensino fundamental(%)	-0,537398
IDEB - 9ºano do ensino fundamental	Taxa de aprovação- ensino fundamental(%)	0,583837
Taxa abandono escolar- ensino fundamental(%)	Taxa de aprovação- ensino fundamental(%)	-0,56687
Taxa de abandono escolar - ensino médio(%)	Taxa de aprovação - ensino médio(%)	-0,626645
Razão aluno por docente(Qtd. Alunos por Docente)	Despesa educação por aluno(em Reais)	-0,381456

Tabela 4: valores de correlação para indicadores do ano 2007

Indicador X	Indicador Y	Valor da Correlação
Taxa abandono escolar- ensino fundamental(%)	Taxa de aprovação- ensino fundamental(%)	-0,594979
Taxa de abandono escolar - ensino médio(%)	Taxa de aprovação - ensino médio(%)	-0,684044
Razão aluno por docente(Qtd. Alunos por Docente)	Despesa educação por aluno(em Reais)	-0.39697

Tabela 5: valores de correlação para indicadores do ano 2008

Indicador X	Indicador Y	Valor da Correlação
IDEB - 9ºano do ensino fundamental	Taxa abandono escolar- ensino fundamental(%)	-0,48999
IDEB - 9ºano do ensino fundamental	Taxa de aprovação- ensino fundamental(%)	0,5532

Taxa abandono escolar- ensino fundamental(%)	Taxa de aprovação- ensino fundamental(%)	-0,575605
Taxa de abandono escolar - ensino médio(%)	Taxa de aprovação - ensino médio(%)	-0,684044
Razão aluno por docente(Qtd. Alunos por Docente)	Despesa educação por aluno(em Reais)	-0,344807
Despesa educação por aluno(em Reais)	Índice de eficiência da educação básica(%)	-0,397397

Tabela 6: valores de correlação para indicadores do ano 2009

Indicador X	Indicador Y	Valor da Correlação
Taxa abandono escolar- ensino fundamental(%)	Taxa de aprovação- ensino fundamental(%)	-0,561508
Taxa de abandono escolar - ensino médio(%)	Taxa de aprovação - ensino médio(%)	-0,67338
Razão aluno por docente(Qtd. Alunos por Docente)	Despesa educação por aluno(em Reais)	-0,386224

Tabela 7: valores de correlação para indicadores do ano 2010

Indicador X	Indicador Y	Valor da Correlação
IDEB - 5ºano do ensino fundamental	Taxa de aprovação- ensino fundamental(%)	0,32106
IDEB - 9ºano do ensino fundamental	Taxa abandono escolar- ensino fundamental(%)	-0,484991
IDEB - 9ºano do ensino fundamental	Taxa de aprovação- ensino fundamental(%)	0,508859
Taxa abandono escolar- ensino fundamental(%)	Taxa de aprovação- ensino fundamental(%)	-0,557196
Taxa de abandono escolar - ensino médio(%)	Taxa de aprovação - ensino médio(%)	-0,587864
Razão aluno por docente(Qtd. Alunos por Docente)	Despesa educação por aluno(em Reais)	-0,395758
Despesa educação por aluno(em Reais)	Índice de eficiência da educação básica(%)	-0,382906

Tabela 8: valores de correlação para indicadores do ano 2011

Observando as tabelas podemos ver que praticamente os mesmos indicadores obtiveram correlações mais altas. Estes indicadores correspondem a nota do IDEB (Indicador 90), taxas de abandono (Indicadores 73 e 74) e taxas de aprovação (Indicadores 80 e 81). Podemos notar que a nota do IDEB é proporcional a taxa de aprovação pois observa-se correlação positiva e alta e inversamente proporcional a taxa de abandono pois foi identificada correlação negativa e alta. As taxas de aprovações e de abandonos dos respectivos níveis de ensino (médio e fundamental) são inversamente proporcionais, pois obtiveram uma correlação negativa e relativamente alta. Estas correlações fazem sentido, porque altas taxas de abandono demonstram falta de interesse dos alunos e consequentemente taxas de aprovação e notas inferiores.

Quanto as correlações mais baixas entre os indicadores acima, verifica-se que os indicadores quantidade de alunos por docentes (Indicador 184) e despesa educação por aluno (Indicador 7) têm correlações negativas nos anos 2007, 2008, 2009, 2010 e 2011, o que pode ser simplesmente explicado pelo fato de quanto mais alunos por professores, menor é a despesa educação por aluno, já que precisa-se pagar a uma quantidade menor de professores para ensiná-los.