

Análise descritiva dos dados

A análise dos dados do MAIS.edu está 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 minizar 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 médioi(%)
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 confiança de 95%.

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 representam os *p-values* dos testes aplicados e as células que tem como valores *NA* são indicadores dos quais não obtivemos dados para o referido ano.

	329	62	89	90	333	73	74	80	81	176	177	202	184	7	201
2000	0.39	NA	NA	NA	0.55	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2007	NA	0.07	0.47	0.046	NA	0.000 0000 12	0.013	0.029 6	0.77	0.000 0006 97	NA	0.000 03	0.003	5.89 * 10 ⁻¹ 6	NA
2008	NA	0.00 009 5	NA	NA	NA	0.000 002	0.000 033	0.13	0.08	0.000 0000 17	NA	0.000 0000 1	0.01	1.1 * 10 ⁻¹ 6	NA
2009	NA	0.00 9	0.4	0.1	NA	0.000 0000 007	0.000 3	0.006 6	0.000 17	0.000 0097	NA	0.000 0000 39	0.5	9.4 * 10 ⁻¹ 9	0.000 0005
2010	0.14	0.00 06	NA	NA	0.00 004 9	0.000 0012	0.14	0.69	0.046	0.000 038	NA	0.000 0000 26	0.35	0.2 * 10 ⁻¹ 5	NA
2011	NA	0.19	0.000 42	0.31	NA	0.000 003	0.000 01	0.82	0.001	0.000 004	0.1	0.000 7	0.029	1.2 * 10 ⁻¹ 6	0.000 0000 198

Tabela 2: p-values dos valores dos indicadores por ano

Estes valores encontrados na tabela 2 são importantes, pois na nossa análise desejamos verificar se estamos nos deparando com dados paramétricos e qual medida de tendência central usa-se para estes indicadores.

Os valores destacados em vermelho na tabela 2 representam *p-values* 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 anualmente analisados, aproximadamente 69% não seguem distribuição normal e quanto aos 31% restantes podemos afirmar que seguem esta distribuição.

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 *NA*s são indicadores dos quais não obtivemos dados para aquele ano.

	329	62	89	90	333	73	74	80	81	176	177	202	184	7	201
2000	39.9	NA	NA	NA	83.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
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. Estes testes foram feitos com a seguinte hipótese nula: “O valor da correlação é igual a zero” e o nível de confiança deste teste é de 95%. 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 indicador menor será o outro indicador.

Cada tabela abaixo corresponde a um ano e estão sendo exibidos apenas os valores que apresentaram correlações moderadas e altas (acima 0,5 e abaixo de -0,5), os indicadores que não alcançaram esse valor foram omitidos e para o ano 2000 nenhum indicador obteve esse valor.

Indicador X	Indicador Y	Valor da Correlação
INDICADOR_90	INDICADOR_73	-0,537398
INDICADOR_90	INDICADOR_80	0,583837
INDICADOR_73	INDICADOR_80	-0,56687
INDICADOR_74	INDICADOR_81	-0,626645

Tabela 4: valores de correlação para indicadores do ano 2007

Indicador X	Indicador Y	Valor da Correlação
INDICADOR_73	INDICADOR_80	-0,594979
INDICADOR_74	INDICADOR_81	-0,684044

Tabela 5: valores de correlação para indicadores do ano 2008

Indicador X	Indicador Y	Valor da Correlação
INDICADOR_90	INDICADOR_80	0,5532
INDICADOR_73	INDICADOR_80	-0,575605
INDICADOR_74	INDICADOR_81	-0,684044

Tabela 6: valores de correlação para indicadores do ano 2009

Indicador X	Indicador Y	Valor da Correlação
-------------	-------------	---------------------

INDICADOR_73	INDICADOR_80	-0,561508
INDICADOR_74	INDICADOR_81	-0,67338

Tabela 7: valores de correlação para indicadores do ano 2010

Indicador X	Indicador Y	Valor da Correlação
INDICADOR_90	INDICADOR_80	0,508859
INDICADOR_73	INDICADOR_80	-0,557196
INDICADOR_74	INDICADOR_81	-0,587864

Tabela 8: valores de correlação para indicadores do ano 2011

Observando as tabelas podemos ver que praticamente os mesmos indicadores obtiveram correlações altas. Esses 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. A taxa de aprovação e de abandono do respectivo nível de ensino (médio e fundamental) são inversamente proporcionais pois obtiveram uma correlação negativa e alta. Essas correlações fazem sentido porque altas taxas de abandono demonstram falta de interesse dos alunos e consequentemente taxas de aprovação e notas inferiores.