Discrepância entre Ensino Público e Ensino Privado pelo ENEM

Joyce de Sant' Anna Brum Projeto de Cálculo Numérico

Sumário

1. Sumário	2
2. Introdução	3
3. Metodologia	4
4. Resultados	7
4.1 - Relação nível x nota	7
4.2 - Relação taxa x nota	9
4.3 - 1ª proposta	11
4.4 - 2ª proposta	14
4.5 - 3 ^a proposta	17
4.6 - 4 ^a proposta	20
5. Discussão e Conclusões	23
6. Referencias	26

Introdução

É fato que, no Brasil, os colégios públicos não são bem vistos, nem por quesito de ensino de boa qualidade e nem por infraestrutura, devido a falta de gerenciamento e investimentos que as escolas passam.

A consequencia imediata disto é o agravamento da discrepância entre o nível educacional proporcionado pelas escolas públicas e o proporcionado pelas escolas privadas. Isto faz com que crianças provenientes de famílias com renda considerada baixa não consigam sustentar seus estudos em uma escola privada, fazendo com que, sua única opção de estudo, seja uma pública.

Enquanto que crianças provenientes de famílias com renda considerada boa raramente optarão por estudar em uma escola pública, visto que seus responsáveis têm dinheiro para pagar uma escola particular para as mesmas.

A consciência de pertencer a uma escola de nível educacional consideravelmente inferior aos das escolas particulares faz com que muitos jovens percam a esperança de alcançar determinados objetivos ou tenham uma visão de futuro limitada ao final do Ensino Médio, pois a perspectiva de cursar uma faculdade pública ou de conseguir condições econômicas para uma particular parece inalcançável.

A cada ano jovens de todo o Brasil, que estão cursando o seu Ensino Médio, participam do ENEM (Exame Nacional do Ensino Médio) que tem por objetivo intermediar a entrada do aluno em uma universidade pública ou mesmo a aquisição de bolsas em universidades particulares.

O problema em questão é avaliar de que forma o nível socioeconômico de uma determinada escola e sua taxa de aprovação podem influenciar no rendimento da mesma no ENEM, a fim de avaliar de fato a discrepância existente entre as escolas que possuem um bom nível socioeconômico (majoritariamente escolas particulares) e as que possuem um nível ruim (majoritariamente escolas públicas), utilizando como apoio também a taxa de aprovação das mesmas.

A taxa de aprovação de uma determinada escola indica o percentual de alunos que foram capazes de passar da atual série em que se encontravam para a próxima. A soma das porcentagens de abandono, reprovação e aprovação devem resultar em 100%, ou seja, a taxa de aprovação informa o percentual de alunos que não repetiu ou não abandonou a escola.

De que forma estariam estes três dados relacionados é a pergunta sobre a qual gira em torno este projeto.

Metodologia

• Estrategia:

Para alcançar este objetivo (comparar as escolas com base em seu nível e taxa de aprovação) foi utilizado o método de Mínimos Quadrados que tenta traçar uma determinada função para melhor estimar o comportamento da nota no ENEM das escolas baseado em seus níveis socioeconômicos e em suas taxas de aprovação.

Foram utilizadas diversas propostas de funções que poderiam estimar este comportamento a fim de comparar (utilizando o erro) no final qual conseguiu se aproximar mais dos pontos dados.

Determinado os coeficientes de uma determinada função, observa-se o esboço de seu gráfico 3d (a fim de ver sua distribuição espacial) e pode-se estimar a nota de uma determinada escola utilizando esta função, dando como parâmetro seu nível socioeconômico e sua taxa de aprovação.

Foram 4 propostas de funções usando o nível e a taxa e 2 usando apenas um dos dois.

Considere que os pontos (nível, taxa, nota/10) são (x, y, z) para fins de simplificação da notação.

Propostas:

1.
$$z = a*y + b$$

2.
$$z = a*x + b$$

3.
$$z = a*x + b*y + c$$

4.
$$z = a*(x)^2 + b*y + c$$

5.
$$z = a*x + b*(y)^2 + c$$

6.
$$z = a*(x)^2 + b*(y)^2 + c$$

1.
$$A*X = B$$
, onde

$$\mathbf{\hat{A}} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & 1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

2.
$$A*X = B$$
, onde

$$\hat{\mathbf{A}} \ = \ \begin{bmatrix} y_1 & 1 \\ y_2 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{bmatrix} \quad \mathbf{X} \ = \ \begin{bmatrix} a \\ b \\ \end{bmatrix} \quad \mathbf{B} \ = \ \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

3.
$$A*X = B$$
, onde

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

4. A*X = B, onde

$$\mathbf{\hat{A}} = \begin{bmatrix} x1^2 & y1 & 1 \\ x2^2 & y2 & 1 \\ \vdots & \vdots & \vdots \\ xn^2 & yn & 1 \end{bmatrix} \mathbf{X} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \mathbf{B} = \begin{bmatrix} z1 \\ z2 \\ \vdots \\ zn \end{bmatrix}$$

5. A*X = B, onde

$$\mathbf{\hat{A}} = \begin{bmatrix} x_1 & y_1^2 & 1 \\ x_2 & y_2^2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n^2 & 1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \mathbf{B} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

6. A*X = B, onde

$$\mathbf{\hat{A}} = \begin{bmatrix} x1^2 & y1^2 & 1 \\ x2^2 & y2^2 & 1 \\ \vdots & \vdots & \vdots \\ xn^2 & yn^2 & 1 \end{bmatrix} \mathbf{X} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \mathbf{B} = \begin{bmatrix} z1 \\ z2 \\ \vdots \\ zn \end{bmatrix}$$

Mas como nenhum desses sistemas é possível, utilizaremos o método para transformar o sistema em um sistema quadrado e em vez de achar X, acharemos " $\sim X$ ", que é o valor aproximado de X.

O novo sistema a ser resolvido será A^tA ($\sim X$) = A^tB e o erro dessa aproximação terá a forma $||A(\sim X) - B||$ que, caso X fosse exato, deveria ser 0.

Coleta de dados:

Os dados disponibilizados sobre cada escola participante do ENEM forneceram os três valores utilizados para estabelecer a relação desejada: nível socioeconômico, taxa de aprovação e nota. Entretanto, o nível socioeconômico não estava em um formato que pudesse ser utilizado numericamente e sim com categorias do tipo: Muito Alto, Alto, Médio Alto, Médio, Médio Baixo, Baixo, Muito Baixo. A fim de poder utilizar o dado da melhor maneira possível foi utilizado a seguinte atribuição numéria representando porcentagens:

MUITO ALTO	100
ALTO	83.333333333
MÉDIO ALTO	66.666666666
MÉDIO	50
MÉDIO BAIXO	33.333333333
BAIXO	16.666666666
MUITO BAIXO	0

De forma que a transição de um nível para o próximo seja sempre igual para todo par de níveis.

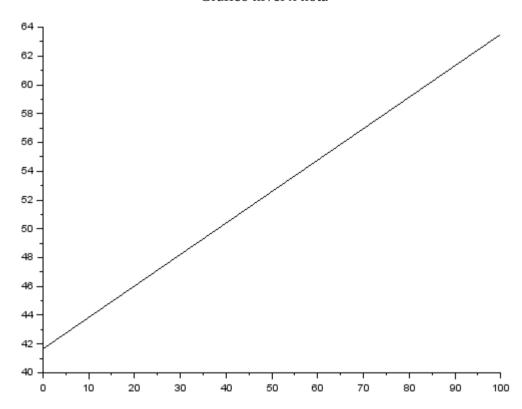
Além disso, as notas em vez de serem avaliadas de 0 a 1000 foram avaliadas de 0 a 100, pois tanto a taxa quanto o nível socioeconômico são apresentados de 0 a 100, evitando qualquer tipo de transformação e mantendo todos os dados no mesmo tipo de unidade (porcentagem).

Resultados

→ Relação da nota apenas com o nível socioeconômico:

a = 0.2185861 // b = 41.651921

Gráfico nível x nota



Ao utilizar os pontos (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.3333333333333, 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNota([50],x)
nota =

52.581225

--> nota = preveNota([66.666666666],x)
nota =

56.224326

--> nota = preveNota([16.666666666],x)
nota =

45.295022

--> nota = preveNota([83.333333333],x)
nota =

59.867427
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95

Os casos de nível máximo e mínimo geraram a seguinte saída:

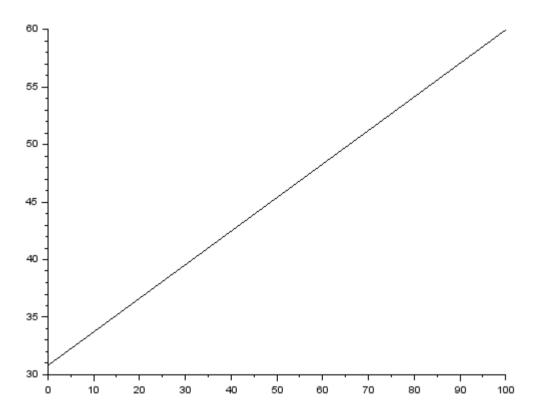
```
--> nota = preveNota([100],x)
nota =
63.510529

--> nota = preveNota([0],x)
nota =
41.651921
```

→ Relação da nota apenas com a taxa de aprovação:

$$nota = a*taxa + b$$

Gráfico taxa x nota



Ao utilizar os pontos (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.33333333333, 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNota([70.30],y)
nota =

51.290577

--> nota = preveNota([100],y)
nota =

59.947339

--> nota = preveNota([95.50],y)
nota =

58.635708

--> nota = preveNota([90.20],y)
nota =

57.090899
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95 Os casos de taxa máximo e mínimo geraram a seguinte saída:

```
--> nota = preveNota([100],y)
nota =

59.947339

--> nota = preveNota([0],y)
nota =

30.799993
```

→ 1^a proposta:

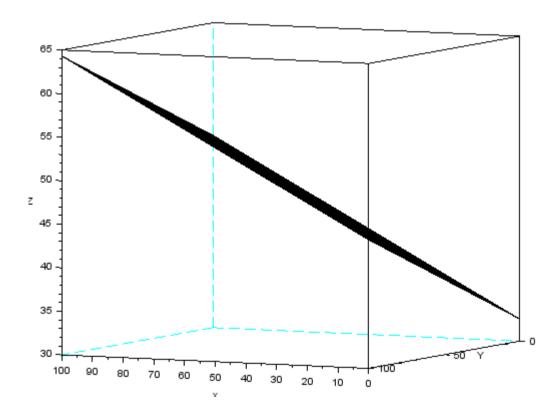
nota = a*nível + b*taxa + c

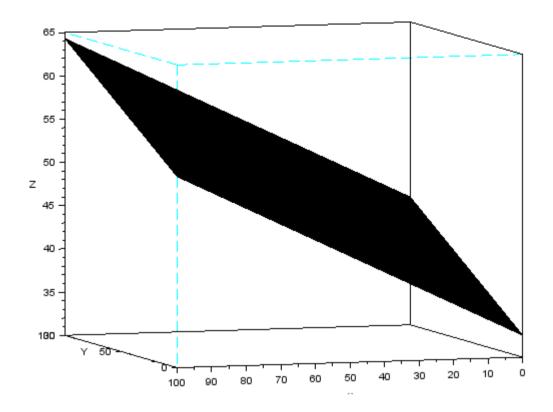
Tenta ver se a relação da nota com o nível e a taxa obedece a uma relação linear

O erro obtido é consideravelmente alto, levando em conta que uma determinada nota esta sendo avaliada de 0 a 100.

'x' é um vetor com os coeficientes que foram encontrados pelo método de Mínimos quadrados.

O gráfico gerado pela função encontrada pelo método tem a seguinte distribuição espacial.





Ao utilizar os pontos (50, 70.30, 448.70), (66.66666666, 100.00, 610.00), (16.666666666, 95.50, 513.20) e (83.3333333333333, 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNotaLinear([50;70.30],x)
nota =

50.901827

--> nota = preveNotaLinear([66.666666666;100],x)
nota =

57.793587

--> nota = preveNotaLinear([16.666666666;95.50],x)
nota =

47.465736

--> nota = preveNotaLinear([83.3333333333;90.20],x)
nota =

59.854101
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95

O caso em que tanto o nível quanto a taxa estão em seu valor máximo e mínimo gerou a seguinte saída:

```
--> nota = preveNotaLinear([100;100],x)
nota =
64.311891

--> nota = preveNotaLinear([0;0],x)
nota =
32.525976
```

→ 2^a proposta:

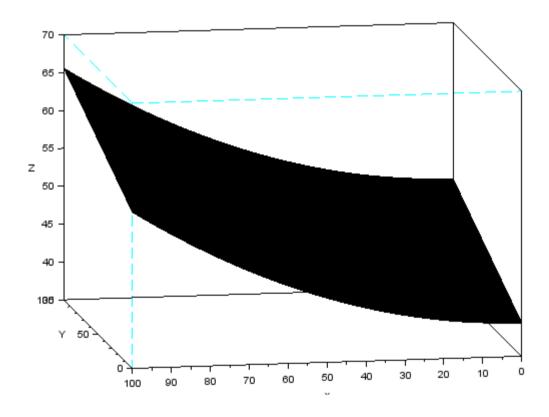
$$nota = a*nivel^2 + b*taxa + c$$

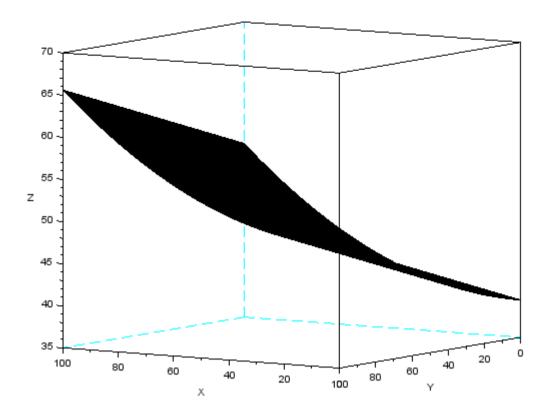
O erro obtido é consideravelmente alto, levando em conta que uma determinada nota esta sendo avaliada de 0 a 100.

 $\mbox{`x'}$ é um vetor com os coeficientes que foram encontrados pelo método de Mínimos quadrados.

$$a = 0.0016185 \: // \: b = 0.0999217 \: // \: c = 39.423113$$

O gráfico gerado pela função encontrada pelo método tem a seguinte distribuição espacial.





Ao utilizar os pontos (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.33333333333, 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNotaQuadNivel([50;70.3],ret)
nota =
50.493756

--> nota = preveNotaQuadNivel([66.6666666666;100],ret)
nota =
56.608435

--> nota = preveNotaQuadNivel([16.6666666666;95.50],ret)
nota =
49.415207

--> nota = preveNotaQuadNivel([83.3333333333;90.20],ret)
nota =
59.67535
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95

O caso em que tanto o nível quanto a taxa estão em seu valor máximo e mínimo gerou a seguinte saída:

```
--> nota = preveNotaQuadNivel([100;0],ret)
nota =

55.607705

--> nota = preveNotaQuadNivel([100;100],ret)
nota =

65.599874

--> nota = preveNotaQuadNivel([0;100],ret)
nota =

49.415283

--> nota = preveNotaQuadNivel([0;0],ret)
nota =

39.423113
```

→ 3^a proposta:

$nota = a*nivel + b*taxa^2 + c$

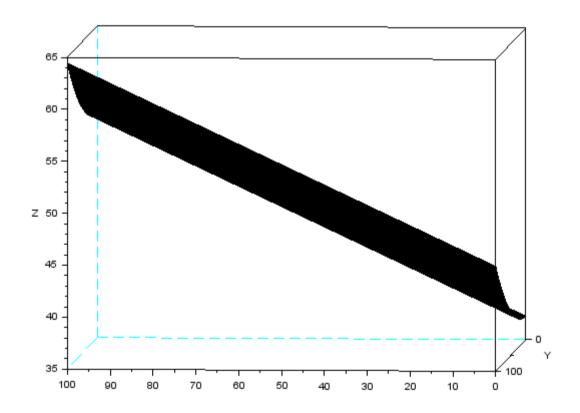
```
--> [ret,erro] = AproximacaoQuadTaxa()
erro =
6893.7969
ret =
0.1936356
0.0007829
37.271531
```

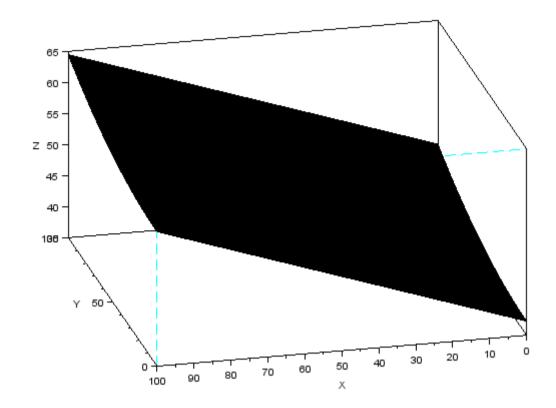
O erro obtido é maior do que todos os anteriores até o momento o que demonstra um distanciamento maior da aproximação ideal.

'ret' é um vetor com os coeficientes que foram encontrados pelo método de Mínimos quadrados:

$$a = 0.1936356 // b = 0.0007829 // c = 37.271531$$

O gráfico gerado pela função encontrada pelo método tem a seguinte distribuição espacial.





Ao utilizar os pontos (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.3333333333333333), 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNotaQuadTaxa([50;70.30],ret)
nota =

50.822581

--> nota = preveNotaQuadTaxa([66.666666666;100],ret)
nota =

58.009789

--> nota = preveNotaQuadTaxa([16.666666666;95.50],ret)
nota =

47.639234

--> nota = preveNotaQuadTaxa([83.333333333;90.20],ret)
nota =

59.777714
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95

O caso em que tanto o nível quanto a taxa estão em seu valor máximo e mínimo gerou a seguinte saída:

```
--> nota = preveNotaQuadTaxa([100;100],ret)
nota =
64.464309

--> nota = preveNotaQuadTaxa([0;0],ret)
nota =
37.271531
```

→ 4^a proposta:

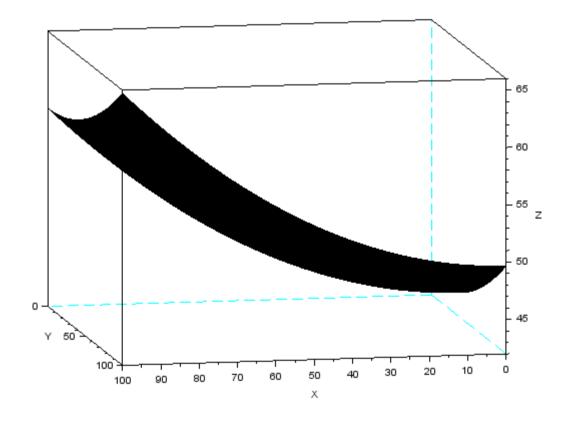
$$nota = a*nivel^2 + b*taxa^2 + c$$

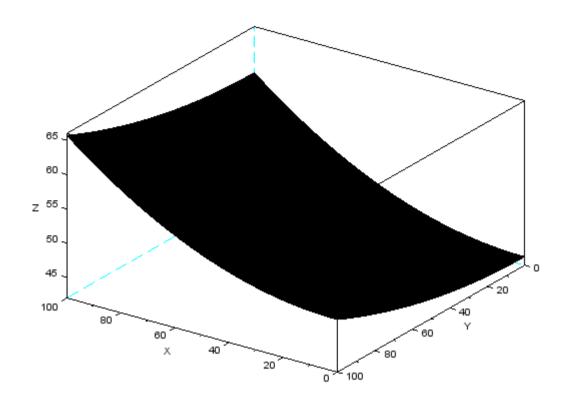
```
--> [ret,erro] = AproximacaoQuadratica()
erro =
6895.5383
ret =
0.0016048
0.0006428
43.239379
```

 $\mbox{`x'}$ é um vetor com os coeficientes que foram encontrados pelo método de Mínimos quadrados:

$$a = 0.0016048 // b = 0.0006428 // c = 43.239379$$

O gráfico gerado pela função encontrada pelo método tem a seguinte distribuição espacial.





Ao utilizar os pontos (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.33333333333333, 90.20, 819.55) — modelo tipo (nível, taxa, nota) — para testar como o a previsão de uma determinada nota se comporta os seguintes valores foram retornados.

```
--> nota = preveNotaQuad([50;70.30],ret)
nota =

50.428186

--> nota = preveNotaQuad([66.666666666;100],ret)
nota =

56.799868

--> nota = preveNotaQuad([16.666666666;95.50],ret)
nota =

49.54757

--> nota = preveNotaQuad([83.333333333;90.20],ret)
nota =

59.613816
```

Onde os valores esperados eram, respectivamente: 44.87, 61.00, 51.32, 81.95

O caso em que tanto o nível quanto a taxa estão em seu valor máximo e mínimo gerou a seguinte saída:

```
--> nota = preveNotaQuad([100;100],ret)
nota =
65.715608

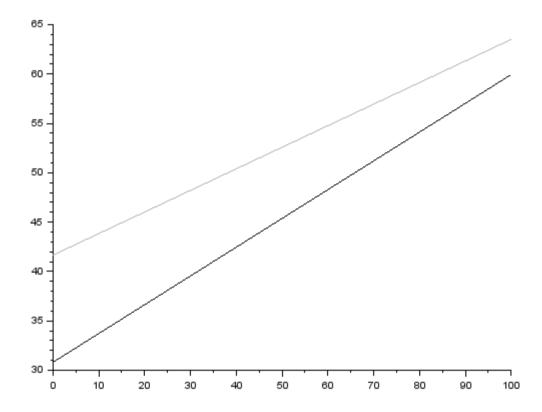
--> nota = preveNotaQuad([0;0],ret)
nota =
43.239379
```

Discussão e Conclusões

Ao comparar a relação linear encontrada para nível x nota com a relação encontrada para taxa x nota é possível perceber que a primeira se aproxima mais dos pontos dados do que a segunda.

ERRO DAS APROXIMAÇÕES			
NÍVEL X NOTA	1884.9978		
TAXA X NOTA	3126.6882		

Comparando a relação linear da nota com a taxa e a relação da nota com o nível (a linha mais clara representa a relação nível x nota e a mais escura representa a relação taxa x nota)



Percebe-se que a inclinação do gráfico que representa a relação da nota com a taxa é maior do que a inclinação da relação do nível com a nota, sendo assim, uma pequena variação na taxa provoca um crescimento mais rápido da nota do que essa mesma variação no nível socioeconômico. Em outras palavras, a nota é mais sensível à taxa de aprovação do que ao nível socioeconômico.

Se levarmos em conta que o erro da relação taxa x nota é consideravelmente superior ao da relação nível x nota, então a afirmativa de que a nota é mais sensível à taxa pode não ser tão verdadeira tendo em vista que a reta nível x nota conseguiu se aproximar mais de todos os pontos dados do que a reta taxa x nota.

Uma outra propriedade interessante de se avaliar neste gráfico é que durante todo o intervalo de 0 a 100 a nota determinada pelo nível sempre foi superior a nota determinada pela taxa e que o valor máximo deste intervalo foi 65. Notas acima de 65 não poderiam sequer ser representadas em nenhuma das duas relações lineares.

Avaliando o erro das relações nota x nível x taxa:

ERRO DAS APROXIMAÇÕES			
Identificador	Descrição	Valor do erro	
1ª proposta	Nota = a*nível + b*taxa + c	6878.8912	
2ª proposta	Nota = a*nível² + b*taxa + c	6883.3769	
3 ^a proposta	Nota = $a*nivel + b*taxa^2 + c$	6893.7969	
4ªproposta	Nota = $a*nivel^2 + b*taxa^2 + c$	6895.5383	

Das quatro propostas a que melhor se aproxima dos pontos é a primeira, onde a nota é linearmente dependente do nível socioeconômico e da taxa de aprovação, visto que conforme termos quadráticos foram sendo introduzidos, maior o erro foi ficando.

Mas mesmo a primeiro tendo sido, dentre as quatro, a que mais se aproximou, o erro desta é consideravelmente superior ao erro das relações nível x nota e taxa x nota, o que leva a conclusão que a melhor maneira de avaliar a discrepância entre as escolas de diferentes níveis e nichos é avaliando unicamente a relação do nível socioeconômico com a nota.

Quanto ao gráfico de cada uma das propostas, algumas coisas ficaram em comum em todos eles: quanto mais próximo do (0,0) menores ficavam os valores da nota e quanto mais próximos de (100,100) maiores ficavam esses valores. Isso pode ser observado, também, pelos valores de a, b e c encontrado para cada uma das propostas.

SOLUÇÃO APROXIMADA ENCONTRADA				
Identificador	a (referente ao nível)	b (referente à taxa)	c (constante)	
1ª proposta	0.1955491	0.12231	32.525976	
2ª proposta	0.0016185	0.0999217	39.423113	
3ª proposta	0.1936356	0.0007829	37.271531	
4ª proposta	0.0016048	0.0006428	43.239379	

SOLUÇÃO APROXIMADA ENCONTRADA			
Identificador	a (referente à variável)	b (constante independente)	
Nível x nota	0.2185861	41.651921	
Taxa x nota	0.2914735	30.799993	

Como todos os valores que multiplicam tanto o nível socioeconômico quanto a taxa de aprovação são positivos, a nota é diretamente proporcional a esses dois valores, ou seja, a nota obtém o seu valor mínimo quando tanto a taxa de aprovação quanto o nível socioeconômico estão em seus valores mínimos, e o mesmo para o valor máximo. Esta relação já era esperada socialmente falando, pois, de fato, quanto mais pobre uma escola, menor a sua nota, e quanto mais rica, maior a nota. O mesmo para a taxa que representa a aprovação dos alunos da escola.

Logo, nenhuma das propostas teve um comportamento ilógico em suas estimativas.

Como sabemos que os pontos de mínimo de todas as propostas serão no (0,0) e os de máximo no (100,100), é fácil compará-los.

VALORES DE NOTAS MÁXIMO E MÍNIMO				
Identificador	Máximo	Mínimo (constante independente)		
nível x nota	63.510529	41.651921		
taxa x nota	59.947339	30.799993		
1ª proposta	64.311891	32.525976		
2ª proposta	65.599874	39.423113		
3ª proposta	64.464309	37.271531		
4ªproposta	65.715608	43.239379		

Outra análise interessante é verificar que mesmo entre as quatro propostas, nenhuma nota chegou em 66, comportamento semelhante à relação da nota com o nível e com a taxa separadamente. Isto pode significar que há uma quantidade significativamente baixa de escolas cuja média é superior a 66.

Por último, seria interessante comparar os valores estimados por cada função. (50, 70.30, 448.70), (66.666666666, 100.00, 610.00), (16.6666666666, 95.50, 513.20) e (83.333333333, 90.20, 819.55)

ESTIMATIVAS APROXIMADAS DAS NOTAS							
Identificador	nível x nota	taxa x nota	1ª proposta	2ª proposta	3ª proposta	4ª proposta	Correto
(50, 70.30)	52.581225	51.290577	50.901827	50.493756	50.822581	50.428186	44.87
(66.66, 100.00)	56.224326	59.635708	57.793587	56.608435	58.009789	56.799868	61.00
(16.66, 95.50)	45.295022	58.635708	47.465736	49.415207	47.639234	49.54757	51.32
(83.33, 90.20)	59.867427	57.090899	59.854101	59.67535	59.777714	59.613816	81.955

O esperado, olhando apenas para o erro, seria que a estimativa do nível x nota chegasse mais perto das estimativas corretas, tendo em vista que seu erro é consideravelmente menor que o erro das demais funções, entretanto, a quarta proposta foi a que mais se aproximou da maioria dos pontos escolhidos. Claro que os pontos escolhidos são poucos para chegar em qualquer generalização, mas é interessante perceber como, apesar das funções serem bastante diferentes, os números estimados estão bem próximos entre si.

Com isso foi possível perceber que independente de qual função se escolha para aproximar os pontos, todas elas apresentam o mesmo tipo de comportamento linear, sendo o nível e a taxa sempre diretamente proporcional a nota e com notas máximas e mínimas consideravelmente próximas, o que demonstra uma concentração razoável das mesmas entre 30 e 66 (ou seja, entre 300 e 660, como o ENEM é calculado de 0 a 1000).

Referências

Os dados usados como entrada no programa foram retirados do site do inep referente ao ano de 2015:

http://inep.gov.br/microdados