

Qualidade do Sucesso Escolar: Turmas do Distrito de Évora em análise e discussão pelo Algoritmo de CHAID e de CRT ou CART

Luzia Barroso

luzia.barroso@sapo.pt

Universidade de Évora

Curso de Mestrado em Ciências da Educação/Administração, Regulação e Políticas Educativas

Merciana Rita

mercianarita@gmail.com

Universidade de Évora

Curso de Mestrado em Ciências da Educação/Administração, Regulação e Políticas Educativas

Resumo: O nosso estudo centra-se nas turmas dos agrupamentos escolares\escolas não agrupadas públicas do 3.º Ciclo do Ensino Básico do Alentejo Central, NUT 3, relativamente às quais se dispõe de informação sobre a avaliação final do 1.º Período Escolar do 7.º, 8.º e 9.º ano de escolaridade (variável resultado) do ano letivo de 2016/17. As características das turmas em estudo foram organizadas em torno do contexto extraescolar, centrado em variáveis socioculturais, económicas e territoriais relativas ao aluno e respetiva família e do contexto intraescola relacionado com aspetos de política educativa e organizacionais-pedagógicos. As árvores de decisão foi a técnica utilizada na construção e classificação de modelos de análise de dados. Com a aplicação do algoritmo CHAID, as turmas foram segmentadas pela percentagem de classificações positivas, pela percentagem de mães com 1.º ciclo ou menos, pelas percentagens de alunos sem ASE e pela percentagem de mães com o ensino superior. Assim, os dados fundamentam a importância do contexto socioeconómico e cultural das famílias dos alunos das turmas. Já com algoritmo CRT ou CART, as turmas foram segmentadas sucessivamente pela percentagem de classificações positivas. Os dados fundamentam a importância do número de classificações acima de 3 na qualidade do sucesso dos alunos.

Palavras-chave: árvore de decisão (algoritmo de CHAID e de CRT ou CART); perfis de turmas; resultados escolares; qualidade do sucesso educativo.

Quality of school success: Alentejo accent Classes in analysis and discussion by CHAID algorithm and CRT or CART

Abstract: The study focuses on classes of the third cycle public primary schools of Alentejo Central, NUT 3, for which offers info on the final evaluation of the first Term of the 7th, 8th and 9th year of schooling (result variable) of the school year from 2016/17. The characteristics of the groups under study were organized around the extracurricular context, centred on socio-cultural, economic and territorial variables relating to student and their family and the inter-school context related to aspects of educational policy and teaching organizational. Decision trees were the technique used in building and classifying of data analysis models. With the application of the CHAID algorithm the classes were segmented according to the percentage of positive evaluation, the percentage of mothers that have attended the fourth grade or less, the percentage of students without economic aid and the percentage of mothers who have a university degree. So the data fundamentals the importance of social, economic and cultural context of the students' families. With the help of the CRT or CART algorithm the classes were successively segmented according to the percentage of positive marks. The data gives fundament to the importance of the number of grades above 3 and of the quality and success of the students.

Keywords: decision tree (CHAID, CRT or CART algorithm); class profiles; school results; quality of school success.

Introdução

Um dos grandes desígnios da educação é inculcar em cada um de nós as capacidades de podermos comandar o nosso próprio desenvolvimento e de podermos ser proficiente que nos assegurem sermos cidadãos críticos e ativos. A educação deve, “de facto, fazer com que cada um tome o seu destino nas mãos e contribua para o progresso da sociedade em que vive, baseando o desenvolvimento na participação responsável dos indivíduos e das comunidades” (Delors, 1996, p. 82). Como afirma Paulo Freire (1997) “se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda” (p.67).

O mundo está cada vez mais imprevisível, os desafios futuros são muitos, o que coloca grande pressão nas escolas. Temos de preparar os nossos alunos para a imprevisibilidade e para a resolução de problemas.

Neste artigo abordaremos as questões do sucesso educativo e da sua qualidade. Serão os fatores de contexto socioeconómico determinantes da qualidade do sucesso educativo? Ou a nível local existem escolas e turmas que fazem a diferença?

“A escola é o local onde o conhecimento se torna público e onde se democratiza a produção e a distribuição do saber” (Sebarroja, 2001, p. 71), sendo que

o carácter público do ensino comporta o cumprimento de uns requisitos mínimos para todos os estabelecimentos sem exceção: igualdade de oportunidades para toda a população escolar com medidas de discriminação positivas para os sectores social e culturalmente mais desfavorecidos; gratuidade do ensino e dos serviços complementares básicos; pluralismo democrático e respeito pelos direitos humanos; não aceitação de nenhum tipo de discriminação por razões sociais, éticas ou de sexo no acesso e durante a escolaridade; e qualidade do ensino. (*idem*, pp. 91-92)

A escola é por tradição o local de ensino, mas cada vez mais deve ser encarada como um local onde se aprende, e aprendemos todos, alunos, professores e órgãos de gestão e lideranças. Como tal, é fundamental avaliar a sua prestação enquanto organização educativa. A escola faz a diferença?

A hipótese do efeito de escola pressupõe que o sistema educativo é constituído pela “agregação de unidades múltiplas, cada uma das quais produzindo efeitos sensivelmente diferentes sobre os resultados escolares e/ou sobre a seleção escolar” (Cousin, 1998, citado por Lima, 2008, p. 419).

A investigação sobre a eficácia das escolas tem por base a necessidade de contestar a tese de que as escolas não produzem efeitos nos alunos. O pressuposto de que a escola produz efeito no desempenho dos alunos, efeitos sobre os seus resultados escolares, reforça a ideia de que existe uma “diversificação da produção escolar,” (Lima, 2008, p. 419) havendo por isso, escolas boas e escolas menos boas.

Moreira (2014) considera que:

é hoje (re)conhecido o falhanço generalizado na garantia do sucesso escolar para todos os alunos e isto por várias ordens de razões: a entrada em massa dos alunos oriundos de todos os grupos sociais encontrou um número crescente de casos de abandono e insucesso escolar; a equidade social e o esbatimento das condições sociais pré-existent não estão garantidas, antes pelo contrário, tal como já afirmaram Bourdieu e Passeron, a escola reproduz as desigualdades (Perrenoud, 2005) e reforça a segregação social (Dubet, 2004); uma formação académica cada vez mais prolongada e especializada nem sempre encontra correspondência com as ambições de mobilidade social e com as necessidades do mercado de trabalho, quando, paradoxalmente, este é cada vez mais exigente (Crahay, 2002; Pires, 2000; Gather Thurler e Perrenoud, 1994). (p. 16)

Como defende Brunet (1992) “cada escola tem a sua personalidade própria que a caracteriza e que formaliza o comportamento dos seus membros” (p.138). Assim,

as escolas constituem uma *territorialidade* espacial e cultural, onde se exprime o jogo dos atores educativos internos e externos; por isso, a sua análise só tem verdadeiro sentido se conseguir mobilizar todas as dimensões pessoais, simbólicas e políticas da vida escolar, não reduzindo o pensamento e a ação educativa a perspectivas técnicas, de gestão ou de eficácia. (Nóvoa, 1992, p. 16)

Segundo Brunet (citado por Nóvoa,1992), “as organizações escolares, ainda que estejam integradas num contexto cultural mais amplo, produzem uma cultura interna que lhes é própria e que exprime os valores (ou os ideais sociais) e as crenças que os membros da organização partilham” (p. 29).

Schein (1985) definiu “cultura organizacional” como:

um padrão de assunções básicas – inventadas, descobertas, ou desenvolvidas por um dado grupo de acordo com a forma como ele aprendeu a lidar com problemas de adaptação ao exterior ou de integração no interior da organização – que foram trabalhadas até serem consideradas válidas e, portanto, comunicadas aos novos membros como o método correto de perceber, pensar e sentir relativamente aqueles problemas. (p. 9)

Como tal, se numa turma, escola, agrupamento de escolas e município existir um “conjunto de valores, crenças, ideologias, normas, regras, representações, rituais, símbolos, hábitos, rotinas, mitos, cerimoniais, formas de interação, formas de comunicação e, até mesmo, as práticas dos atores localizados em determinada organização” (Torres, 1997, p. 14) que tenha uma cultura que valorize o sucesso escolar, todos os atores educativos serão impulsionados para fazerem mais e melhor.

Defendemos que as escolas fazem a diferença, os professores fazem a diferença, as turmas fazem a diferença, o território faz a diferença! As culturas existentes influenciam também o desempenho dos seus alunos. Se a cultura existente for meritocrática cada um tenderá a melhorar o seu desempenhar, o que incentivará a melhoria do sucesso.

Método

A amostra é composta por turmas dos agrupamentos escolares\escolas não agrupadas públicas do 3.º Ciclo do Ensino Básico do Alentejo Central, NUT 3, relativamente às quais se dispõe de informação sobre a avaliação final do 1.º Período Escolar do 7.º, 8.º e

9.º ano de escolaridade (variável resultado) do ano letivo de 2016/17 e sobre as seguintes variáveis independentes do contexto extraescolar e intraescolar: número total de alunos das turmas (Nº de Alunos), média do número de anos de escolaridade das mães (N.º Médio_Anos_Habilit_Mães), percentagem de mães com ensino superior (% Mães com 1.º ciclo ou menos), percentagem de mães com 1.º ciclo ou menos (% Mães com 1.º ciclo ou menos), percentagem de alunos no escalão A da Ação Social Escolar (% ASE A), percentagem de alunos sem escalão da Ação Social Escolar (% Sem ASE), percentagem de alunos com Necessidades Educativas Especiais (% NEE), média do desvio etário (Média desvio etário) e área territorial municipal onde estão localizadas as escolas (Município).

Estas nove variáveis independentes são as variáveis explicativas escolhidas para análise das inter-relações com o nível de desempenho escolar e que permitem identificar e analisar perfis de desempenho das escolas relacionados com o tamanho da escola (número total de alunos que as frequentam), com o capital cultural escolar das mães das famílias de proveniência dos alunos (habilitações académicas das mães), com as carências económicas dos alunos e respetivas famílias (apoio social escolar), com o histórico de progressão escolar e não desfasamento etário (densidades de jovens no ensino básico regular) e com o município de localização da escola.

Na amostra não estão consideradas as turmas das escolas de Territórios Educativos de Intervenção Prioritária (TEIP).

As árvores de decisão são uma técnica na construção e classificação de modelos de análise de dados. A sua representação é caracterizada por uma estrutura hierarquizada que traduz uma árvore invertida, apresenta-se da raiz (nó zero) até às folhas (nós terminais). A estrutura da pesquisa desenvolve-se do geral para o particular, isto é, a cada novo nível de nós descendentes limita-se o valor de mais um atributo explicativo.

É um método de classificação supervisionado, já que uma variável dependente é explicada por n variáveis independentes.

Exploramos este método através da aplicação de dois algoritmos: o CHAID e o CRT ou CART.

O algoritmo CHAID (Chi-Squared Automatic Interaction Detection), o seu nome vem do teste estatístico utilizado para calcular os pontos de corte ideais (splits) durante o seu processo de divisão. O algoritmo CHAID é um algoritmo de classificação para implementação de árvores de decisão. Este algoritmo foi desenvolvido por Kass em 1980. Em 1991 surge o Exhaustive CHAID, que representa uma evolução em relação ao CHAID, nomeadamente no que toca a qualidade da análise do processo de corte (Biggs, 1991). O algoritmo CHAID começa o seu processamento por criar e analisar a tabela de contingência entre cada uma das variáveis de input e a variável dependente, de forma a analisar a relação entre as diferentes variáveis independentes e a variável dependente, de seguida é realizado um teste estatístico – teste chi-square – de forma a determinar a independência entre as variáveis. Este algoritmo define automaticamente um máximo de três níveis de profundidade.

O algoritmo CRT ou CART (Classification and Regression Trees) foi desenvolvido por Breiman, em 1984. A construção da árvore é feita de forma recursiva, neste processo os nós são repetidamente divididos, até se chegar às folhas (nós terminais). A divisão recursiva inicia-se na raiz da árvore, e começa pela seleção da melhor variável para usar na divisão, posteriormente são definidos os pontos de divisão ideais da variável. A cada nó terminal (folha), correspondente apenas uma classe, e cada nó terminal é definido por um conjunto de regras único (Lewis, 2000). Este algoritmo trabalha com uma profundidade máxima de cinco níveis e as árvores geradas por este algoritmo são sempre binárias.

As árvores de decisão e os seus algoritmos apresentam como principais vantagens a possibilidade de utilizar variáveis de diferentes escalas de medida, poder usar a mesma variável independente em diferentes níveis de profundidade evidenciando sucessões de efeitos sobre outras variáveis, não requerer processos de logaritmização ou de normalização nem quaisquer outras transformações e dispensar pressupostos de aplicabilidade; como principal desvantagem, a grande quantidade de dados que requerem. (Verdasca, 2013, p. 180)

Caraterização da amostra

A amostra é constituída por 166 turmas das escolas públicas da Região do Alentejo Central, como se pode constatar através do Tabela 1:

Tabela 1: Distribuição das turmas por município

<i>Municípios</i>	7.º ano	8.º ano	9.º ano	Totais
<i>Alandroal</i>	2	2	2	6
<i>Arraiolos</i>	3	3	3	9
<i>Borba</i>	3	3	3	9
<i>Estremoz</i>	3	3	3	9
<i>Évora</i>	21	21	20	62
<i>Montemor-o-Novo</i>	6	9	5	20
<i>Mora</i>	2	2	2	6
<i>Portel</i>	3	2	3	8
<i>Redondo</i>	3	3	3	9
<i>Reguengos de Monsaraz</i>	7	5	4	16
<i>Viana do Alentejo</i>	4	4	4	12
<i>Totais</i>	57	57	52	166

As características das turmas em estudo foram organizadas em torno do contexto extraescolar, centrado em variáveis socioculturais, económicas e territoriais relativas ao aluno e respetiva família (N.º Médio_Anos_Habilit_Mães, % Mães com 1.º ciclo ou menos, % Mães com 1.º ciclo ou menos, % ASE A, % Sem ASE) e do contexto intraescolar relacionado com aspetos de política educativa e organizacionais-pedagógicos (Nº de Alunos, % NEE, Média desvio etário e Município).

Tabela 2: Caraterísticas da amostra

	Mínimo	Máximo	Média	Desvio padrão	Variância
<i>Alunos por Turma</i>	9	30	20,4	4,5	20,1
<i>%NEE</i>	0,0	23,5	3,7	5,4	28,8
<i>%ASE A</i>	0,0	55,6	22,3	11,8	139,9
<i>%Sem ASE</i>	22,2	92,3	62,6	14,8	219,2

<i>Média desvio etário</i>	0,0	1,4	0,6	0,3	0,1
<i>N.º médio anos Habilitações Mães</i>	6,1	15,3	11,2	1,8	3,4
<i>%Mães com Ensino Superior</i>	0,0	77,3	21,5	14,7	217,4
<i>%Mães com 1º ciclo ou menos</i>	0,0	38,9	7,0	7,5	56,1

Da análise dos dados da Tabela 2 destacam-se: a) a elevada heterogeneidade das turmas no que respeita à percentagem de alunos sem ASE (219,2) com amplitudes que oscilam entre os 22,2 e os 92,3 e percentagem de alunos com ASE de escalão A (139,9) com amplitudes que variam entre 0 e 55,6; b) a elevada heterogeneidade ao nível da percentagem das habilitações académicas das mães sendo que existem turmas onde nenhuma das mães tem habilitação de ensino superior e outras há em que 77,3 das mães possuem certificação superior, assim a variância elevada de 217,4; c) a existência de heterogeneidade significativa de 51,6% em relação à percentagem de mães com a habilitação de 1.º Ciclo ou menos, a variância oscila entre 0 e 38,9 das turmas; d) quanto às restantes variáveis, o grau de disparidade entre as turmas é relativamente menor e ainda que subsistam diferenças elas não são tão extremas (valores < 30%), onde não existe praticamente variância 0,1% é na média ao desvio etário.

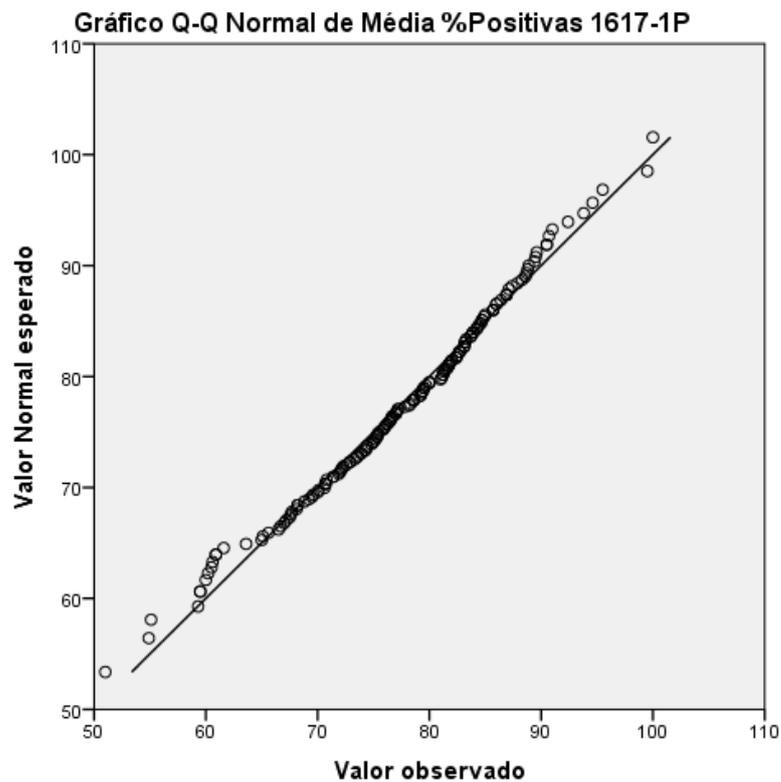
No que respeita à variável média de percentagens positivas no 1.º Período Escolar do ano letivo 2016/2017, esta apresenta uma distribuição tendencialmente simétrica já que os valores da média, mediana e moda apresentam valores muito próximos (77,5; 78,3; 78,6 respetivamente). No entanto, a distribuição é ligeiramente assimétrica negativa já que a mediana tem valores superiores à média, mas inferiores à moda (-,325) conforme se verifica na Tabela 3.

Tabela 3: Estatísticas descritivas da variável média de % de Positivas no ano letivo 2016/2017 – 1.º Período Escolar

	Mínimo	Máximo	Média	Mediana	Moda	Desvio padrão	Variância	Assimetria
<i>Média de %Positivas 16/17- 1.ºPeríodo Escolar</i>	51,0	100,0	77,5	78,3	78,6	9,0	81,3	-,325

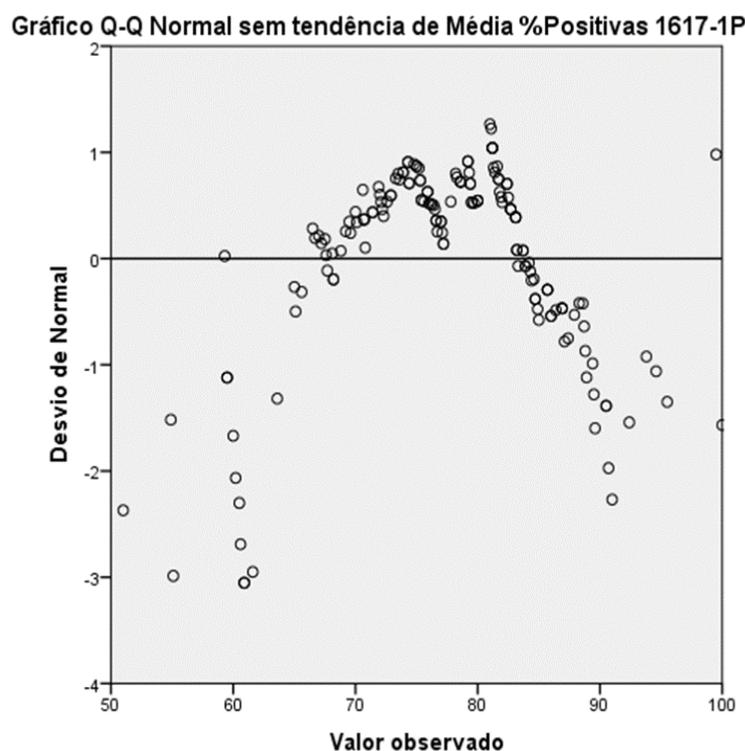
A distribuição da variável média de % de Positivas (Gráfico 1) confirma a tendência da normalidade da amostra considerada, já que a distribuição se aproxima tendencialmente da normalidade. Assim, podemos afirmar que a normalidade é sustentada.

Gráfico 1: Distribuição normalizada da variável média de % de Positivas no ano letivo 2016/2017 – 1.º Período Escolar



Ainda no âmbito da variável acima indicada, a distribuição é tendencialmente simétrica com se pode verificar no Gráfico 2.

Gráfico 2: Distribuição normalizada sem tendência da variável média de % de Positivas no ano letivo 2016/2017 – 1.º Período Escolar



Relativamente à variável média da Qualidade do Sucesso no 1.º Período Escolar do ano letivo 2016/2017 verificamos que uma distribuição tendencialmente simétrica já que os valores da média, mediana e moda apresentam valores muito próximos (3,143; 3,1; 3,2 respetivamente), conforme o exposto na Tabela 4.

Tabela 4: Estatísticas descritivas da variável média da Qualidade do Sucesso no ano letivo 2016/2017 – 1.º Período Escolar

	Mínimo	Máximo	Média	Mediana	Moda	Desvio padrão	Variância	Assimetria
(n) Média da Qualidade do Sucesso 16/17 – 1.º Período Escolar	2,6	4,0	3,143	3,1	3,2	0,2523	0,064	0,559

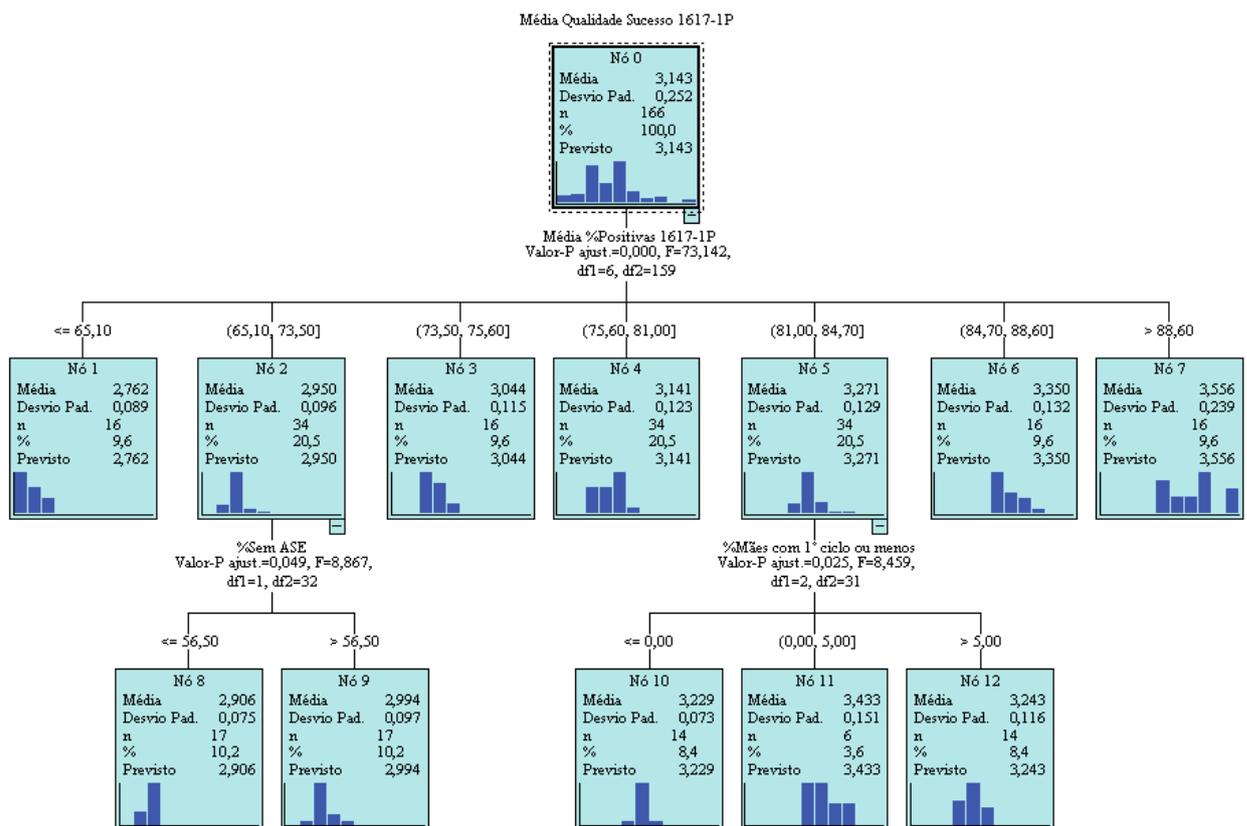
Apresentam-se de seguida as especificações e resultados da aplicação dos algoritmos CHAID e CRT ou CART e as respetivas soluções arbóreas geradas.

Árvore de classificação pelo método de crescimento CHAID

A metodologia CHAID utilizada segundo Pestana e Gageiro (2009) é um “método exploratório para estudar as relações entre uma variável de resposta e um conjunto de variáveis explicativas que podem interagir entre si” (p.366). Com a aplicação deste algoritmo chegamos a uma árvore com catorze nós, dos quais onze nós são terminais distribuídos por três níveis de profundidade.

A Figura 1 mostra a árvore gerada para a variável Média Qualidade do Sucesso, utilizando o algoritmo desenvolvido por (Kass, 1980), que de seguida iremos explorar.

Figura 1 – Solução arbórea gerada pelo método CHAID



Observando a Figura 1, constatamos que:

- a) A amostra geral está representada no nó zero (0);
- b) As estatísticas do nó zero (0) confirmam os valores da Tabela 4;

- c) A solução arbórea obtida tem uma estrutura não binária já que se segmenta em 11 nós;
- d) A árvore obtida estrutura-se em 3 níveis;
- e) Das variáveis independentes que o modelo fez interagir com a qualidade do sucesso escolar são mais significativas as variáveis por ordem de importância de classificação: a média de % positivas, a % de mães com 1.º ciclo ou menos, % de alunos Sem ASE e por último, a % de mães com ensino superior;
- f) A variável independente que mais diferencia as turmas quanto à qualidade do sucesso escolar é a percentagem de média de positivas e que segmentando as escolas em sete grupos, sendo quatro nós terminais (1,3,6 e 7);
- g) No último nível de profundidade apuraram os restantes nós finais (7).
- h) A segunda variável independente mais significativa, como já referido, é a % de mães com o 1.º ciclo ou menos emergindo três nós finais (12,13 e 14);
- i) A terceira variável mais significativa é a % de alunos sem ASE do qual emergem dois nós finais (8 e 9);
- j) A última variável independente mais significativa que devemos considerar é a % de mães com ensino superior da qual se ramifica em dois nós finais (10 e 11).

A solução gerada pelo algoritmo CHAID conduziu-nos à classificação das turmas em 11 perfis.

De seguida iremos apresentar a caracterização de cada um dos perfis gerados por ordem crescente da qualidade do sucesso educativo.

Perfil Nó 1: As turmas deste nó apresentam a média mais baixa (2,762) da média do total da amostra geral, com uma diferença de 0,381. Neste perfil estão representadas 16 turmas que correspondem a 9,6% da amostra geral, sendo que a média das percentagens positivas é igual ou inferior a 65,1.

Perfil Nó 8: Nesta subamostra encontram-se turmas onde a percentagem de alunos sem ASE é inferior ou igual a 56,5 com uma média de 2,906, o que corresponde 17 turmas da amostra geral;

Perfil Nó 9: Deste nó constam 14 turmas onde a percentagem de mães com o 1.º Ciclo ou menos de habilitação académica é superior a 56,5 com uma média de 2,994, o que corresponde 17 turmas da amostra geral. As turmas aqui incluídas, arredondado o valor, já se encontram com uma média de qualidade do sucesso escolar positivo;

Perfil Nó 3: As turmas incluídas neste nó possuem uma média positiva de 3,044, o que corresponde a um universo de 16 turmas, sendo a média das percentagens positivas entre os valores de 73,5 e 75,6.

Perfil Nó 10: Esta subamostra detém uma média de 3,112 para um universo de 25 turmas que se caracterizam por a percentagem das mães com ensino superior ser igual ou inferior a 27,3. As turmas têm uma média de percentagem de classificações positivas entre os valores de 75,6 a 81,0.

Perfil Nó 11: Esta subamostra detém uma média de 3,222 para um universo de 9 turmas que se caracterizam por a percentagem das mães com ensino universitário superior a 27,3. As turmas têm uma média de percentagem de classificações positivas entre os valores de 75,6 a 81,0.

Perfil Nó 12: Deste nó fazem parte 14 turmas onde não existem mães com o 1.º ciclo ou menos de habilitação académica. Esta subamostra tem uma média de 3,229, sendo a média de percentagem de classificações positivas entre os valores de 81 e 84,7.

Perfil Nó 14: As turmas deste nó têm uma média de 3,243. Neste subgrupo estão representadas 14 turmas onde a percentagem de mães com o 1.º ciclo ou menos é superior a 5,00. A percentagem de classificações positivas oscila entre os valores de 81 e 84,7.

Perfil Nó 13: Deste subgrupo estão representadas 6 turmas onde a percentagem de mães com o 1.º ciclo ou menos é inferior ou igual a 5,00. A média deste nó é de 3,433 de valor de classificação, sendo a média de percentagem de classificações positivas entre os valores de 81 e 84,7.

Perfil do Nó 6: Deste nó com uma média 3,350 de qualidade de sucesso, fazem parte 16 turmas a média de percentagem de classificações positivas entre os valores de 84,7 e 88,6.

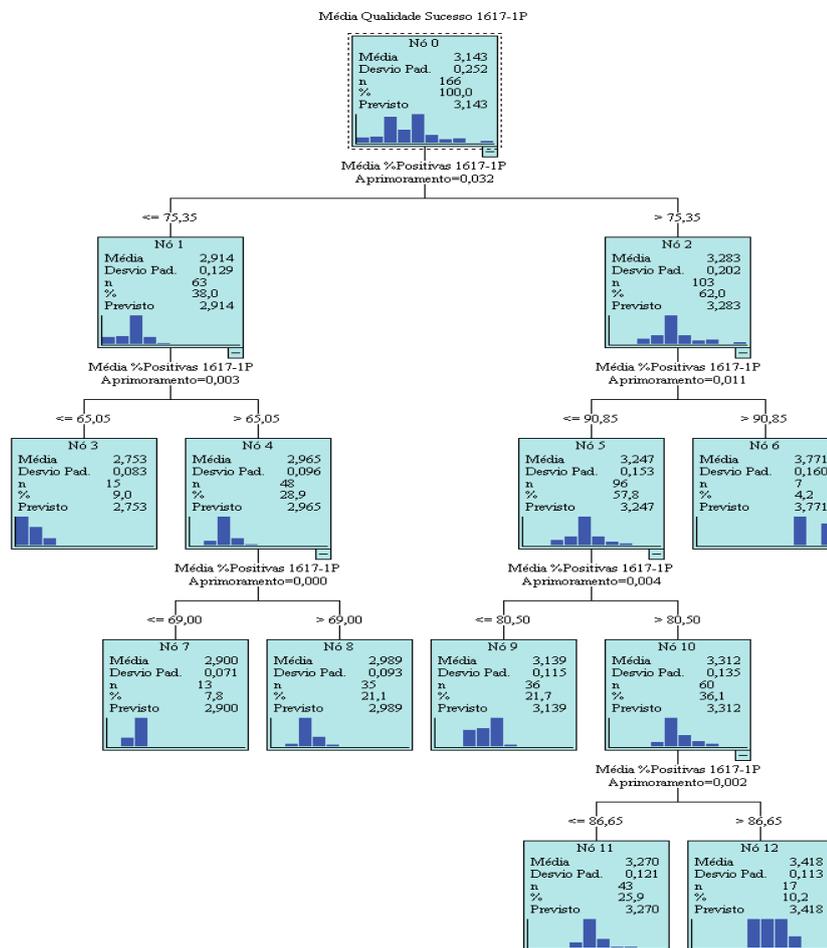
Perfil do Nó 7: De todos os nós terminais é o que regista a maior qualidade de sucesso educativo, o desempenho médio dos alunos das turmas é de 3,556, que arredondado

corresponde ao nível 4, situação única para um universo de 16 turmas. A média de percentagens de classificações positivas é superior a 88,6.

Em síntese, o algoritmo CHAID segmenta as turmas pela percentagem de classificações positivas, pela percentagem de mães com 1.º ciclo ou menos, pelas percentagens de alunos sem ASE e pela percentagem de mães com o ensino superior. Os dados fundamentam a importância do contexto socioeconómico e cultural das famílias dos alunos das turmas.

Árvore de classificação pelo método CRT ou CART

Figura 2 – Solução arbórea gerada pelo método CRT ou CART



Analisando a Figura 2, inferimos que:

- A amostra geral está representada no nó zero (0);
- As estatísticas do nó zero (0) confirmam os valores da Tabela 4;
- A solução arbórea obtida tem uma estrutura não binária já que se segmenta em 12 nós;

- d) A árvore obtida estrutura-se em 5 níveis;
- e) Das variáveis independentes que o modelo fez interagir com a qualidade do sucesso escolar a única que surge com este método é a média de % positivas, que se repete recursivamente;
- f) A variável independente que diferencia as turmas quanto à qualidade do sucesso escolar é a percentagem de média de positivas e que segmentando as escolas em dois grupos no primeiro nível de profundidade; no segundo nível de profundidade a ramificação faz-se em dois nós cada, sendo que dois nós são terminais (3 e 6); no terceiro nível de profundidade a ramificação faz-se em dois nós cada, sendo que três dos nós são terminais (7, 8 e 9); o último nível de profundidade ramifica-se em dois nós terminais (11 e 12).

A solução gerada pelo algoritmo CRT ou CART conduziu-nos à classificação das turmas em 7 perfis.

De seguida iremos apresentar a caracterização de cada um dos perfis gerados por ordem crescente da qualidade do sucesso educativo.

Perfil Nó 3: As turmas deste nó apresentam a média mais baixa (2,753) da média do total da amostra geral, com uma diferença de 0,39. Neste perfil estão representadas 15 turmas que correspondem a 9,0% da amostra geral, sendo que a média das percentagens positivas é igual ou inferior a 65,05.

Perfil Nó 7: Nesta subamostra encontram-se turmas onde o valor continua a ser negativo, mas mais elevado que no perfil anterior, o valor médio é de 2,9. Corresponde a 13 turmas da amostra geral;

Perfil Nó 8: Deste nó constam 35 turmas, o que correspondem a 21,1% da amostra geral. As turmas aqui incluídas arredondado o valor, já se encontram com uma média de qualidade do sucesso escolar positivo (2,989);

Perfil Nó 9: As turmas incluídas neste nó possuem uma média positiva de 3,139, o que corresponde a um universo de 36 turmas, correspondendo a 21,7 % do total de turmas da amostra.

Perfil Nó 11: Esta subamostra detém uma média de 3,270 para um universo de 43 turmas, correspondendo a 25,9 % do total de turmas da amostra.

Perfil Nó 12: Esta subamostra detém uma média de 3,418 para um universo de 17 turmas, correspondendo a 10,2 % do total de turmas da amostra.

Perfil Nó 6: Deste nó fazem parte apenas 7 turmas com uma média de 3,771, sendo a média de percentagem de classificações positivas superior a 90,85, correspondendo a 4,2 % do total de turmas da amostra.

Em síntese, o algoritmo CRT ou CART segmenta sucessivamente as turmas pela percentagem de classificações positivas. Os dados fundamentam a importância do número de classificações acima de 3 na qualidade do sucesso dos alunos, sendo que é uma variável tão robusta que não chegam a emergir as variáveis independentes de contexto socioeconómico e cultural das famílias dos alunos das turmas.

Discussão dos Resultados

A análise dos dados permitiu-nos definir vários perfis de turmas de 3.º Ciclo do Ensino Básico, com diferentes padrões de qualidade do sucesso educativo, ao nível da Região do Alentejo Central.

Através da aplicação do algoritmo CHAID, a probabilidade de erro na média de % de positivas é residual ($p=0,000$), em relação à % de mães com 1.º ciclo ou menos a probabilidade de erro é baixa ($p=0,001$), quanto à % de alunos sem ASE a probabilidade de erro também é pequena ($p=0,005$) o que revela que estas duas variáveis independentes são muito robustas na justificação da qualidade do sucesso dos alunos das turmas analisadas. No que diz respeito à % de mães com ensino superior a probabilidade de erro já é superior ($p=0,019$). Assim, estas variáveis são estaticamente significativas, pela ordem referidas, na explicação da qualidade do sucesso das turmas e que vão corroborar com o valor real observado.

Com a aplicação do algoritmo CRT ou CART, a margem de erro também é baixa, no entanto os resultados são menos aliciantes, já que a única variável independente que emerge é a média de percentagem de positivas.

Assim, as nossas reflexões ir-se-ão centrar nos resultados do algoritmo CHAID, que nos fornece mais dados pertinentes de análise.

Os resultados dizem-nos que as variáveis independentes: % de positivas, % de mães com 1.º ciclo ou menos, % de alunos sem ASE e % de mães com ensino superior explicam apenas uma parte da variância explicativa.

Assim, o meio sociocultural e económico de origem das famílias dos alunos explica em parte os desempenhos destes, no entanto, há variáveis locais que também explicam o desempenho dos alunos das turmas. É por isso que turmas de alunos com características socioculturais iguais revelam desempenhos de qualidade do sucesso diferentes.

Estas aferições são corroboradas pelas conclusões das publicações da DGEEC, sobre as desigualdades socioeconómicas e resultados escolares do 2.º e 3.º ciclo do ensino público geral (2016), as publicações apresentam os principais resultados sobre a relação entre o desempenho escolar dos alunos do 2.º e 3.º ciclo e o meio socioeconómico dos seus agregados familiares. Os estudos sugerem que no nosso país, no ensino público, existe uma muito forte relação entre o desempenho escolar dos alunos e o contexto socioeconómico e cultural das suas famílias.

Por exemplo, entre os alunos cujas mães têm licenciatura ou bacharelato, a percentagem de “percursos de sucesso”¹ no 3.º ciclo é de 71%, enquanto entre os alunos cujas mães têm habilitação escolar mais baixa, equivalente ao 4.º ano, a mesma percentagem de percursos de sucesso é de apenas 19%. (DGEEC, 2016, p. 2)

O meio sociocultural e económico de origem das famílias dos alunos explica em parte os desempenhos destes, no entanto, não existem apenas esses fatores de contexto. Apesar do nível socioeconómico do agregado familiar ter uma forte influência sobre os desempenhos escolares, existem outros fatores igualmente importantes que, ao exercer a sua influência de forma assimétrica entre as regiões, podem compensar e até superar os efeitos do nível socioeconómico no distrito. A influência de fatores locais como o dinamismo das escolas e dos seus professores, como o grau de importância atribuído ao ensino das crianças e ao trabalho escolar na cultura da região, poderá, porventura, sobrepor-se localmente ao efeito do nível socioeconómico, justificando assim o facto de alunos de regiões com nível socioeconómico baixo poderem, não obstante, ter níveis de desempenho escolares no 2.º ciclo francamente superiores à média nacional. (DGEEC, 2016, p. 13)

Contudo, as estatísticas apresentadas no estudo sugerem também que o nível socioeconómico não equivale a destino, ou seja, não determina de forma inapelável o desempenho escolar dos alunos. Prova disso é o facto dos alunos de certas regiões do país

¹ Por definição, um aluno com "percurso de sucesso" no 3.º ciclo é um aluno que obtém positiva nas duas provas finais de 9.º ano (Português e Matemática) após um percurso sem retenções nos 7.º e 8.º anos.

com indicadores socioeconómicos desfavoráveis, como Braga ou Viseu, terem, não obstante, indicadores de desempenho no 3.º ciclo francamente superiores à média nacional. Por exemplo, observa-se que, em média, os alunos do distrito de Braga cujas mães têm habilitação baixa, equivalente ao 6.º ano, têm um desempenho escolar no 3.º ciclo superior aos alunos do distrito de Beja cujas mães têm como habilitação o 12.º ano completo. Existem portanto outros fatores importantes em jogo, além do nível socioeconómico, fatores que importa investigar localmente e de forma mais aprofundada. (DGEE, 2016, p. 3)

Conclusões

Muitos estudos, nacionais e internacionais, demonstraram que os alunos provenientes de meios socioeconómicos mais desfavoráveis tendem a obter, em média, piores resultados escolares do que os oriundos de meios mais favorecidos. No entanto, não é um destino traçado, já que há alunos, em certas turmas e escolas, que rompem o determinismo do meio socioeconómico e que apesar da adversidade conseguem melhor sucesso educativo do que seria expectável.

Assim, há outros aspetos a considerar na qualidade do sucesso, que deverão ser investigados. Fatores que localmente fazem a diferença, tais como: o dinamismo das escolas e dos seus professores, o grau de importância atribuído à escola e ao ensino, a diversificação e gestão curricular, a articulação curricular, o trabalho colaborativo, a inovação pedagógica, a organização e funcionamento interno, o relacionamento com a comunidade e a cultura escolar e educacional do local.

Urge, então, estudar os bons exemplos e replicá-los, adequando-os ao contexto local. Porque ainda vale a pena lutar pelo sucesso e pela qualidade do sucesso educativo dos nossos alunos.

Referências bibliográficas

- Biggs, D. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, vol. 18, pp. 49-62.
- Brunet, L. (1992). Clima de trabalho e eficácia da escola. Em A. Nóvoa, *As organizações escolares em análise* (pp. 123-140). Lisboa: Dom Quixote.

- Delors, J. (1996). *Educação um Tesouro a Descobrir*. São Paulo: Cortez Editora.
- DGEEC. (2016). *Desigualdades socioeconómicas e resultados escolares II. 2.º ciclo do ensino público geral*. Lisboa: DGEEC.
- DGEEC. (2016). *Desigualdades socioeconómicas e resultados escolares. 3.º ciclo do ensino público geral*. Lisboa: DGEEC.
- Freire, P. (1997). *Pedagogia da Autonomia: saberes necessários à prática educativa*. São Paulo: Paz e Terra.
- Kass, G. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29 (2), pp. 119-127.
- Lewis, R. (2000). An introduction to classification and regression tree (CART) analysis. *Annual Meeting of the Society for Academy Emergency Medicine*.
- Lima, J. Á. (2008). *Em Busca da Boa Escola*. Vila Nova de Gaia: Fundação Manuel Leão.
- Moreira, L. (2014). *Projeto Fénix: Sentidos de um projeto emergente na construção do sucesso escolar*. Porto: Universidade Católica Portuguesa.
- Nóvoa, A. (1992). *As organizações escolares em análise*. Lisboa: Publicações Dom Quixote.
- Pestana, M. H., & Gageiro, J. N. (2009). *Análise Categórica, Árvores de Decisão e Análise de Conteúdo*. Lisboa: LIDEL.
- Pestana, M. H., & Gageiro, J. N. (2009). *Análise Categórica, Árvores de Decisão e Análise de Conteúdo*. Lisboa: LIDEL.
- Schein, E. (1985). *Organizational Culture and Leadership*. San Francisco: Jossey-Bas.
- Sebarroja, J. C. (2001). *A Aventura de Inovar. A mudança na escola*. Porto: Porto Editora.
- Torres, L. (1997). *Cultura Organizacional Escolar: Representação dos Professores numa Escola Portuguesa*. Oeiras: Celta Editora.
- Verdasca, J. (2013). Rankings escolares: "A César o que é de César, e a Deus o que é de Deus". Em J. Verdasca, *Educação: Temas e Problemas 12,13* (pp. 175-199). Évora: Universidade de Évora.

Apêndice I

Solução pelo método de crescimento CHAID

Tabela 5

Resumo do modelo		
Especificações	Método de crescimento	CHAID
	Variável dependente	Média Qualidade Sucesso 1617-1P
	Variáveis independentes	Alunos PTurma, %NEE, %ASE A, %Sem ASE, Média desvio etário, N° médio anos Habilitações Mães, %Mães com Ensino Superior, %Mães com 1° ciclo ou menos, Média %Positivas 1617-1P
	Validação	Nenhum
	Profundidade de árvore máxima	3
	Casos mínimos em nó pai	15
	Casos mínimos em nó filho	5
	Variáveis independentes incluídas	Média %Positivas 1617-1P, %Sem ASE, %Mães com 1° ciclo ou menos
Resultados	Número de nós	15
	Número de nós de terminal	11
	Profundidade	2

Tabela 6

Resumo de ganhos para nós

Nó	N	Porcentagem	Média
7	16	9,6%	3,556
11	6	3,6%	3,433
13	9	5,4%	3,422
14	7	4,2%	3,257
12	14	8,4%	3,243
10	14	8,4%	3,229
4	34	20,5%	3,141
3	16	9,6%	3,044
9	17	10,2%	2,994
8	17	10,2%	2,906
1	16	9,6%	2,763

Método de crescimento: CHAID
 Variável dependente: Média Qualidade Sucesso 1617-1P

Tabela 7

Risco	
Estimativas	Modelo padrão
,015	,002

Método de crescimento: CHAID
 Variável dependente: Média Qualidade Sucesso 16/17-1P

Apêndice II

Solução pelo método de crescimento CRT ou CART

Tabela 8

Resumo do modelo

Especificações	Método de crescimento	CRT	
	Variável dependente	Média Qualidade Sucesso 1617-1P	
	Variáveis independentes	Alunos PTurma, %NEE, %ASE A, %Sem ASE, Média desvio etário, N° médio anos Habilitações Mães, %Mães com Ensino Superior, %Mães com 1º ciclo ou menos, Média %Positivas 1617-1P	
	Validação	Nenhum	
	Profundidade de árvore máxima		5
Resultados	Casos mínimos em nó pai		30
	Casos mínimos em nó filho		5
	Variáveis independentes incluídas	Média %Positivas 1617-1P, N° médio anos Habilitações Mães, %ASE A, %Sem ASE, %Mães com Ensino Superior, Média desvio etário, %Mães com 1º ciclo ou menos, Alunos PTurma, %NEE	
	Número de nós		13
	Número de nós de terminal		7
	Profundidade		4

Tabela 9

Resumo de ganhos para nós

Nó	N	Porcentagem	Média
6	7	4,2%	3,771
12	17	10,2%	3,418
11	43	25,9%	3,270
9	36	21,7%	3,139
8	35	21,1%	2,989
7	13	7,8%	2,900
3	15	9,0%	2,753

Método de crescimento: CRT

Variável dependente: Média Qualidade Sucesso 16/17-1P

Tabela 10

Risco

Estimativas	Modelo padrão
,011	,001

Método de crescimento: CRT

Variável dependente: Média

Qualidade Sucesso 16/17-1P