

PODE UMA NOVA METOLOGIA DE ENSINO MELHORAR RESULTADOS? AVALIAÇÃO MICROECONOMÉTRICA DO SUCESSO ESCOLAR

Alcina Maria de Almeida Rodrigues Nunes, Escola Superior de Tecnologia e de Gestão-Instituto Politécnico de Bragança

Paula Odete Fernandes, Escola Superior de Tecnologia e de Gestão-Instituto Politécnico de Bragança

RESUMEN

La introducción de una metodología de enseñanza basada en la filosofía preconizada por el Proceso de Bolonia, en una de las carreras del área de Administración ministradas por una institución de enseñanza superior en Portugal, permitió la realización de un proceso de evaluación microeconómica del impacto de la filosofía de enseñanza en el suceso escolar de los alumnos a ella sujetos. Aplicando un método econométrico científicamente testado e internacionalmente aceptado, el *Propensity Score Matching*, concluyese que la introducción de una nueva metodología de enseñanza mejoró las tasas de aprobación de los alumnos sujetos a ella, cuando comparadas con las de otras carreras en el entorno del área de Administración; pero ya no cuando comparadas con las tasas de aprobación en todas las carreras ministradas por la institución. Todavía, cuando se compara la calidad de esas tasas de aprobación, los resultados no son positivos, independientemente del área científica de las carreras.

PALABRAS CLAVE: Metodologías de Enseñanza, Evaluación Microeconómica, *Propensity Score Matching*.

RESUMO: A introdução, num dos cursos da área de Gestão leccionados por uma instituição de ensino superior em Portugal, de uma metodologia de ensino baseada na filosofia preconizada pelo Processo de Bolonha, permitiu a realização de um processo de avaliação microeconómica do impacto da filosofia de ensino no sucesso escolar dos alunos a ela sujeitos. Aplicando um método econométrico científicamente testado e internacionalmente aceite, o *Propensity Score Matching*, conclui-se que a introdução de uma nova metodologia de ensino melhorou as taxas de aprovação dos alunos sujeitos à medida quando comparadas com as dos cursos na vizinhança da área da Gestão mas não quando comparadas com as taxas de aprovação em todos os cursos leccionados pela instituição. No entanto, quando se compara a qualidade dessas taxas de aprovação os resultados não são positivos, independentemente da área científica dos cursos.

PALAVRAS-CHAVE: Metodologias de Ensino, Avaliação Microeconómica, *Propensity Score Matching*.

1. INTRODUÇÃO

O ano lectivo de 2004/2005 caracterizou-se na Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança (ESTiG-IPB) pela reformulação de um dos nove cursos leccionados pela instituição, o

Curso de Informática de Gestão (IG)¹. Esta reformulação, visando colocar em prática o Tratado de Bolonha, adoptou não só uma estrutura curricular diferente, do que era comum nesta instituição, como também uma nova filosofia de ensino. A introdução da nova filosofia de ensino verificou-se essencialmente ao nível das metodologias de ensino com o objectivo de melhorar o sucesso escolar. Destaca-se a obrigatoriedade de participação nas aulas, a existência de uma avaliação contínua ao longo do período escolar assim como um contínuo, e mais personalizado, acompanhamento do aluno pelos docentes.

Estando as autoras interessadas em contribuir para a discussão relativa aos métodos pedagógicos de ensino na instituição onde leccionam, em particular, e no ensino superior, em geral, elaboraram uma avaliação microeconómica do impacto, das alterações metodológicas, no sucesso escolar daqueles a que a elas estiveram sujeitos. A avaliação e os resultados que aqui se apresentam correspondem a uma avaliação objectiva, que adopta uma abordagem metodológica distinta da que se realiza normalmente no interior das instituições, pretendendo-se que os resultados possam servir de base a uma discussão mais fundamentada e cientificamente alicerçada. Retirar inferências relativamente à adopção de uma nova filosofia de ensino, em oposição ao *status quo*, pode contribuir para aferir da sua eficiência e eficácia, relativamente aos objectivos que se pretendem atingir, e fornecer aos decisores um poderoso instrumento de análise.

Sujeitar um curso da ESTiG-IPB à adopção de uma nova filosofia de ensino - a que no futuro nos iremos referir como tratamento para facilitar a explicação do método econométrico utilizado - quando os restantes não lhe foram sujeitos, cria as condições para que se efectue um processo de avaliação quase-experimental cujo objectivo é quantificar o efeito causal da exposição ao tratamento. Ou seja, quantificar o efeito causado pela introdução da metodologia de ensino nos resultados dessa mesma metodologia.

O que se verifica é que os métodos estatísticos normalmente utilizados para tentar quantificar os efeitos, nos resultados, da exposição a um determinado tratamento correspondem a medidas estatísticas de associação, como a correlação, entre a variável de tratamento e a variável de resultado, e não necessariamente a medidas de estimação de um efeito causal entre o tratamento e o resultado. A associação entre duas variáveis nem sempre significa causalidade, isto é, nem sempre permite estabelecer uma relação causa-efeito entre essas variáveis.

Neste caso concreto, considera-se que a simples análise comparativa dos resultados do sucesso escolar entre diferentes cursos é insuficiente para concluir sobre o efeito da metodologia de ensino nos resultados dos alunos do curso de IG. Assim, para quantificar inferências causais acerca da exposição a um determinado tratamento, analisa-se a adopção de uma filosofia de ensino através de um método internacionalmente conhecido, nos processos de avaliação empírica de políticas sociais, como *Propensity Score Matching*². Esta metodologia de avaliação empírica é bastante intuitiva podendo ser explicada como uma técnica onde, para os elementos do grupo de tratamento serão encontrados elementos idênticos no grupo não sujeito a tratamento, o grupo de comparação, em termos das suas características observáveis, de forma a que a diferença nos seus resultados seja possível de atribuir ao que os distingue - a participação no tratamento.

¹ Por motivos institucionais apenas o curso de Informática de Gestão foi reformulado mantendo-se os restantes cursos, leccionados na ESTiG-IPB, inalterados até à entrada em funcionamento, no presente ano lectivo, do resultado da sua adequação de acordo com a actual legislação para o ensino superior. Simultaneamente, enquanto a reformulação dos restantes cursos foi sendo discutida e preparada, foi também discutida a oportunidade da extensão da metodologia de ensino introduzida no curso de IG aos restantes cursos da instituição.

² A designação de alguns métodos e técnicas adoptará a designação inglesa original pois a sua actualidade, e a sua não adopção em Portugal, não permitem encontrar tradução adequada em português.

Muitas das características observáveis para os alunos dos diferentes cursos da ESTiG-IPB correspondem a dados administrativos, que são registados pelo sistema de informação da instituição, ao qual se acedeu para realizar a presente avaliação. Outras das características dos alunos têm que ser obtidas questionando-os sobre as mesmas porque não correspondem a características pessoais ou educacionais objectivas. O questionário directo aos alunos constitui uma boa solução para obter alguns dos dados que se pretendem mas, sendo este um projecto de avaliação *ex-post*, não foi possível recolher tal informação³.

Se os resultados são positivos em termos de taxas de aprovação - a medida quantitativa mais significativa para o sucesso escolar - especialmente quando se analisam cursos na vizinhança da mesma área científica, neste caso a área de Gestão, tal não ocorre quando se tenta avaliar a qualidade dessas taxas de aprovação. Verifica-se que os alunos sujeitos à metodologia apresentam, em média, classificações positivas inferiores às dos seus colegas que não foram sujeitos à metodologia.

O presente trabalho divide-se da seguinte forma. Na secção seguinte apresenta-se e explica-se o método econométrico utilizado. Segue-se-lhe uma secção onde se explica como se adequou o método ao problema em questão e, após uma secção onde se apresentam resultados, conclui-se.

2. O PROBLEMA DA AVALIAÇÃO E A SUA RESOLUÇÃO

2.1. O “problema” da Avaliação

O problema de avaliação de programas sociais tem sido amplamente publicitado e é geralmente apresentado como um problema de “ausência de informação” podendo ser formalizado de uma forma simples. Recorrendo ao nosso estudo concreto, num dado momento de tempo, um aluno da ESTiG encontra-se numa de duas potenciais situações (D), cada uma delas originando um determinado resultado (Y): na situação 1 o aluno participa na filosofia em avaliação, a que nos referiremos como situação de tratamento; na situação 0 ele não participa no programa. O resultado desta formulação apresenta-se através do chamado Modelo de Rubin:

$$Y = DY_1 + (1 - D)Y_0 \quad (1)$$

Face ao anterior modelo faz sentido associar ambos os resultados e pensar na diferença entre eles como o impacto da participação, no aluno, isto é, o efeito causal, no indivíduo, da participação na política em causa pode ser expresso através da seguinte expressão:

$$\Delta = Y_1 - Y_0 \quad (2)$$

O problema de avaliação surge porque, para um aluno num momento particular de tempo, é impossível observar a sua participação (D=1) e, em simultâneo, a sua não-participação (D=0) - o aluno ou participa na metodologia sob avaliação ou não participa!

Assim, apenas um dos resultados é observado dando origem a um problema de avaliação de políticas sociais - não é possível conhecer o efeito causal de uma política para um determinado indivíduo porque não existe a evidência contrária, ou seja, um “contra facto” correspondente ao que aconteceria ao aluno na ausência do tratamento. A estimação do efeito de um tratamento baseia-se, então, na construção artificial de um “contra

³ Semestralmente, os alunos são inquiridos relativamente a questões como a motivação e a assiduidade e alguma informação subjectiva existe na ESTiG-IPB embora protegida por medidas de confidencialidade o que limita uma adequada avaliação.

facto”, isto é, na inferência acerca de um potencial resultado que teria sido observado se o aluno não tivesse sido sujeito a uma política (Rubin, 1974; Rosenbaum & Rubin, 1983).

Nesta análise em concreto, para percebermos o impacto da introdução de uma nova filosofia de ensino na probabilidade de os alunos obterem sucesso escolar é necessário realizar uma dedução acerca da mesma probabilidade no caso desses mesmo alunos não terem sido sujeitos à aplicação da filosofia.

Na literatura econométrica de avaliação todos os esforços para resolver o problema de avaliação têm sido desenvolvidos para responder à questão da construção do “contra facto”. Esses esforços distinguem-se, entre si, pela hipótese que assumem acerca da relação entre os dados que estão disponíveis e aqueles que estão em falta. No entanto, para além da discussão acerca das diferentes formas de construir um “contra facto”⁴, a literatura de avaliação concorda, em geral, com a impossibilidade de cálculo do efeito causal, Δ , para um único indivíduo. Desta forma, a solução estatística foi transferida do nível individual para o nível global da população como forma de estimar parâmetros de interesse médios. Neste trabalho o parâmetro de interesse é o efeito médio do tratamento naqueles que participaram:

$$\Delta P \equiv E(\Delta|D=1) = E(Y_1 - Y_0|D=1) = E(Y_1|D=1) - E(Y_0|D=1) \quad (3)$$

2.2. O método de *Matching*

Os métodos de *matching* têm sido bastante refinados com os mais recentes progressos na literatura de avaliação tornando-se actualmente num valioso instrumento de avaliação em termos de metodologia empírica (Blundell & Dias, 2002). Este tipo de métodos é intuitivamente apelativo. No nosso caso concreto corresponderá em fazer corresponder um aluno que participou na introdução de uma nova metodologia de ensino a um aluno que não participou mas que apresente as mesmas características observáveis de forma que a diferença entre os seus resultados apenas possa ser imputada à introdução da metodologia (Dehejia & Wahba, 2002).

Apesar de intuitivo, o processo de *matching* pode tornar-se complexo numa avaliação do tipo quase-experimental devido a problemas de selecção que podem enviesar os resultados. Assim, o processo de *matching* assume hipóteses bastante fortes. Uma dessas fortes hipóteses é a chamada Hipótese da Independência Condicional, que não podendo ser provada ou testada, pode ser discutida caso a caso. Esta hipótese assume que a selecção para o processo de tratamento (D), condicionada por um vector de características observáveis (X), é independente dos resultados potenciais (Y). Em termos de notação formal, esta hipótese corresponde a:

$$(Y_1, Y_0) \perp D | X \quad (4)$$

Em termos práticos, a hipótese significa que os alunos são seleccionados para o curso, que foi sujeito à introdução de uma nova filosofia de ensino, de acordo com um conjunto de características observáveis mas essa selecção não dependeu dos potenciais resultados em termos de sucesso escolar.

Se esta hipótese parece admissível, para o presente projecto de avaliação, ela pode criar problemas práticos de implementação se o vector de características X possuir uma elevada dimensão e variáveis contínuas. Atentos a este problema, Rosenbaum and Rubin (1983) mostraram que proceder a um ajustamento entre

⁴ Blundell and Dias (2002), por exemplo, dividem os métodos de avaliação em cinco categorias, de acordo com a modalidade de construção do “contra facto”: experiências sociais aleatórias, experiências naturais, métodos de *matching*, modelos de selecção e modelos estruturais de simulação.

indivíduos com base numa função escalar de X , como a que eles baptizaram de *Propensity score* - $P(X)$, é suficiente para equilibrar o vector de características X , entre os indivíduos tratados e aqueles que lhe servirão como termo de comparação. Os autores definiram *propensity score* como uma variável uni-dimensional que representa a probabilidade condicional de participação no tratamento, em questão, dado o vector de características observáveis.

Deste modo, se a hipótese acima mencionada é aceitável condicionada no vector de características X é também aceitável se for condicionada na variável *propensity score*, $P(X)$:

$$(Y_1, Y_0) \perp D | P(X), \text{ com } P(X) = \Pr(D = 1 | X) \quad (5)$$

Para além de exigir a *assumpção* de uma hipótese forte para a construção do grupo de comparação, o método de *matching* requer também uma hipótese de equilíbrio para garantir conteúdo empírico ao efeito médio de tratamento. Esta hipótese indica que o número de observações com a mesma *propensity score* estimada deve possuir a mesma distribuição de características, independentemente do estado de tratamento, isto é, $D \perp X | P(X)$. Por outro lado exige-se que, para satisfazer a condição anterior, deve existir para cada característica X tanto participantes como não-participantes, ou seja, $0 < P(X) = \Pr(D = 1 | X) < 1$.

2.3. A técnica de *matching* através do vizinho mais próximo

Decidida a adopção da *metodologia de propensity score* várias técnicas podem ser seleccionadas para realizar o ajustamento entre os alunos participantes e os que lhe vão servir de comparação, isto é, podem ser utilizadas diferentes técnicas para realização do processo de *matching*.

Cada técnica compreende a definição de um critério de proximidade, uma vizinhança, e a selecção de um adequada função de ponderação que sirva para associar os indivíduos no grupo de comparação a cada um dos indivíduos tratados. Ora a escolha de uma vizinhança baseia-se num *trade-off* entre a variância e o enviesamento associados com cada tipo de técnica de *matching* assim como no fardo computacional que pode acarretar. Em geral, por exemplo, diminuir a exigência de proximidade entre observações para construir o “*contra facto*” (para seleccionar o indivíduo que será comparável) irá reduzir a variância mas aumentar o enviesamento resultante da utilização, em média, de mais mas também mais pobres ajustamentos. Esta solução aumenta também as dificuldades de estimação computacional.

Neste estudo concreto será adoptada a técnica do “vizinho mais próximo” que significa que cada aluno de IG, isto é cada aluno sujeito a tratamento, será comparado com o aluno mais próximo, retirado dos cursos de comparação, que apresentar o mesmo valor em termos de *propensity score*. Adicionalmente, permitiremos que exista a reposição dessa observação de comparação, caso se prove que o aluno seleccionado, nos cursos que servirão de comparação, é a melhor observação para servir de comparação aos alunos de IG.

3. APLICAÇÃO EMPÍRICA DO MÉTODO ECONOMÉTRICO À AVALIAÇÃO DO CURSO DE IG

3.1. Selecção das disciplinas a avaliar

Da apresentação do método de *Propensity Score Matching* ficou claro que a sua adaptação a um processo concreto de avaliação requer a comparação entre aquilo que pode ser comparável. Será este o pressuposto de partida para o processo de avaliação efectuado e que a seguir se descreve.

Na ESTiG-IPB, no ano lectivo de 2004/2005, eram leccionados nove cursos cada um deles com as suas especificidades mas também com pontos em comum, objectivamente observáveis e potencialmente comparáveis. De facto, o curso de IG possuía em comum com os restantes cursos da ESTiG algumas disciplinas que são o objecto deste estudo de avaliação. Pela análise das fichas de disciplina das disciplinas do curso de IG verificou-se a existência de um programa idêntico ao de disciplinas de outros cursos assim como, frequentemente, a partilha da mesma carga horária, do mesmo docente ou do mesmo formato lectivo. Face a esta constatação, foram avaliados os resultados das disciplinas que apresentavam um programa, assim como outras características formais, comum aos vários cursos da ESTiG onde eram leccionadas⁵.

Da observação das fichas de disciplina foram seleccionadas as seguintes disciplinas do curso de IG: Álgebra Linear, Matemática I, Matemática II, Estatística, Investigação Operacional, Introdução à Informática, Contabilidade Financeira I, Contabilidade Financeira II e Marketing⁶.

Concluída a constatação de que as disciplinas são semelhantes em termos formais, chega-se à conclusão que apenas a metodologia de ensino é diferente entre elas. Mesmo assim, constatou-se que em algumas disciplinas se aplica também um dos pilares da metodologia de ensino que aqui se procura avaliar - a avaliação contínua. É o caso das disciplinas de Marketing ou de Introdução à Informática. No entanto, em nenhum dos cursos que serviram para comparação existe a obrigatoriedade de participação dos alunos nas aulas nem o acompanhamento mais contínuo e personalizado dos alunos é assumido como uma linha de orientação pedagógica obrigatória. Desta forma as diferenças nos resultados dos alunos apenas se deverão aquilo que diferencia as disciplinas nos diferentes cursos - a introdução de uma nova metodologia de ensino no curso de IG.

Quanto aos resultados, vamos utilizar duas medidas distintas de sucesso escolar. Ou seja, mediremos o sucesso escolar com base: (i) na percentagem de alunos com classificações iguais ou superiores a 10, nota que divide o sucesso e o insucesso escolar, e; (ii) na média da classificação dos alunos que obtiveram uma classificação positiva. O objectivo é avaliar não apenas sucesso escolar com base nos alunos que foram aprovados na avaliação mas também avaliar o que designaremos de qualidade desse sucesso escolar, de forma a melhor compreender o impacto da nova metodologia de ensino.

3.2. Comparação de médias antes de introduzir a metodologia de *Propensity Score matching*

Vejam os resultados de uma simples comparação de médias.

⁵ Para auxiliar a concluir da comparabilidade entre disciplinas foi analisado o grau de comparabilidade entre os programas das disciplinas com base na ponderação atribuída pelas autoras a cada um dos itens do programa. Sempre que se verificaram dúvidas relativas a esta mesma ponderação foram pedidos esclarecimentos aos responsáveis pelas disciplinas. Se, para algumas das disciplinas seleccionadas o grau de comparabilidade é máximo, significando a disciplina é absolutamente idêntica (disciplinas de Marketing, de Álgebra Linear, Estatística ou Investigação Operacional), para outras disciplinas foi possível encontrar graus de comparabilidade inferiores devido a diferenças não significativas. De facto, assumiu-se que pequenas diferenças no programa não devem ser factores de distinção entre resultados dos alunos.

⁶ As restantes disciplinas do curso apresentam demasiadas especificidades e a sua inclusão na análise não permitiria concluir se as diferenças nos resultados dos alunos se deveria à metodologia de ensino ou a outro factor qualquer relacionado com essas especificidades. As questões de confidencialidade não permitiram também utilizar variáveis que permitissem o tratamento estatístico de aspectos como a motivação do aluno relativamente a cada uma das disciplinas. A não utilização de variáveis que permitissem a aproximação estatística a aspectos subjectivos que podem influenciar o sucesso escolar, não sendo um factor restritivo da nossa análise, condiciona as conclusões que podem ser elaboradas após a obtenção dos resultados estatísticos.

Tabela 1: Percentagem de alunos aprovados antes de introduzirmos o *matching*.

DISCIPLINAS	CURSOS DE COMPARAÇÃO					
	TODOS			ÁREA DE GESTÃO		
	GT	GC	Diferença (Δ)	GT	GC	Diferença (Δ)
Álgebra Linear	26.5	41.1	- 14.6	26.5	37.3	- 10.8
Matemática I	18.6	28.8	- 10.2	18.6	20.7	- 2.1
Matemática II	-----	-----	-----	27.4	17.8	9.7
Estatística	53.2	43.6	9.6	53.2	32.3	20.9
Investigação Operacional	34.6	42.7	- 8.1	34.6	32.9	1.7
Introdução à Informática	-----	-----	-----	30.4	32.7	- 2.2
Marketing	90.7	71.8	18.9	90.7	69.0	21.7
Contabilidade Financeira I	73.7	29.5	44.2	73.7	25.4	48.3
Contabilidade Financeira II	-----	-----	-----	56.4	33.0	23.4

A Tabela 1 mostra a percentagem de alunos aprovados no curso de IG (grupo de tratamento - GT) e a mesma percentagem nos restantes cursos onde a disciplina é leccionada (grupo de comparação - GC), assim como a diferença entre os resultados dos alunos. Os resultados são apresentados assumindo como sub-grupos particulares de comparação, todos os cursos onde a disciplina de IG é comparável a outra e apenas os cursos na vizinhança daquela que designámos área de gestão - os cursos de Gestão de Empresas (GE) e Contabilidade e Administração (CA).

Da observação da Tabela 1, alguns factos devem ser salientados: (i) as disciplinas de Matemática II, Introdução à Informática e Contabilidade Financeira II só são comparáveis com disciplinas em cursos da área de Gestão; (ii) das seis disciplinas que é possível comparar com cursos não apenas na área da Gestão - Álgebra Linear, Matemática I, Estatística, Investigação Operacional, Marketing e Contabilidade Financeira I - três apresentam resultados piores para os alunos do curso de IG do que para os alunos que servem de comparação; (iii) restringindo a comparação das seis disciplinas mencionadas no ponto acima a cursos apenas na área da gestão verifica-se que os resultados melhoram, deixando mesmo a Investigação Operacional de apresentar resultados menos positivos para os alunos de IG do que para os restantes. Este facto deve-se a um notório decréscimo na aprovação dos alunos nos cursos da área da gestão; (iv) é interessante notar que a disciplina de Introdução à Informática, podendo ser considerada uma disciplina mais específica de um curso como o de IG apresenta piores resultados que disciplinas não tão relacionadas com o objecto específico do curso, como as Contabilidades ou o Marketing.

Tabela 2: Média das classificações positivas antes de introduzirmos o *matching*.

DISCIPLINAS	CURSOS DE COMPARAÇÃO					
	TODOS			ÁREA DE GESTÃO		
	GT	GC	Diferença (Δ)	GT	GC	Diferença (Δ)
Álgebra Linear	11.38	11.72	- 0.34	11.38	12.70	- 1.31
Matemática I	11.27	11.96	- 0.68	11.27	11.68	- 0.40
Matemática II	-----	-----	-----	11.75	12.32	- 0.57
Estatística	10.69	10.93	- 0.24	10.69	10.50	0.19

Investigação Operacional	11.00	11.39	- 0.39	11.00	11.62	- 0.62
Introdução à Informática	-----	-----	-----	12.71	11.16	1.55
Marketing	11.31	11.45	- 0.14	11.31	11.53	- 0.22
Contabilidade Financeira I	10.38	11.30	- 0.92	10.38	11.17	- 0.78
Contabilidade Financeira II	-----	-----	-----	11.21	11.44	- 0.23

A Tabela 2, que mostra a média das classificações positivas, ajuda a esclarecer o impacto da introdução de uma metodologia de ensino no sucesso escolar daqueles que a ela foram sujeitos. É notório que não existe qualquer impacto na qualidade das notas. Quando o grupo de comparação engloba todos os cursos verifica-se que, mesmo não apresentado diferenças significativas em termos de classificação positiva, os alunos de IG apresentam uma classificação positiva média inferior aos seus colegas. Quando o grupo de comparação se resume aos cursos na área de gestão, as diferenças apresentam-se mais significativas. No entanto, o pior resultado para os alunos de IG apresenta a excepção da disciplina de Introdução à Informática onde os mesmos possuem uma média de classificação positiva 1,55 pontos acima dos colegas de GE e CA.

Sem que se efectue o ajuste econométrico proposto para comparar os alunos de IG apenas com aqueles que, de facto, lhe são semelhantes, os resultados apenas se limitam a estabelecer possíveis relações entre a introdução da metodologia e os resultados obtidos sem que se defina uma verdadeira relação causa-efeito.

4. EFEITO CAUSAL DA INTRODUÇÃO DE UMA METODOLOGIA DE ENSINO NOS ALUNOS DE IG

Nesta secção serão apresentados os resultados da adopção de uma metodologia de *propensity score matching* para estimar o efeito causal da introdução da uma nova metodologia de ensino nos alunos do curso de IG. Em concreto, mostrar-se-á qual o efeito da metodologia nos alunos de IG comparando com o que lhes teria acontecido se eles tivessem frequentado a mesma disciplina mas num curso diferente, onde não foi utilizada a mesma filosofia de ensino. Para aferir dos resultados dos alunos do curso de IG, caso eles não tivessem sido sujeitos à nova metodologia de ensino, vão ser encontrados nos cursos de comparação - para cada uma das disciplinas seleccionadas para análise - alunos que apresentem as mesmas características observáveis que cada um dos alunos de IG. O *matching* entre alunos será realizado caso a caso e para cada aluno de IG vai ser encontrado o aluno, num dos restantes cursos, que lhe seja idêntico ou, no mínimo, muito semelhante⁷.

4.1. Efeito da introdução da metodologia na percentagem de aprovação

Da informação disponível e acessível publicamente foram seleccionadas as variáveis que constam na tabela seguinte.

Tabela 3: Distribuição das características observáveis, seleccionadas para a realização do processo de *matching*, entre os grupos de tratamento e comparação.

DISCIPLINAS	GRUPOS	CARACTERÍSTICAS
-------------	--------	-----------------

⁷ O ideal no processo de *matching* entre alunos seria poder dispor de um conjunto vasto de variáveis correspondentes a todas as características que podem influenciar os resultados, no entanto, nem toda a informação estatística desejável existe e, por outro lado, nem toda a informação existente está disponível publicamente pelos motivos já avançados.

		Alunos	Idade	% Sexo (Masculino)	% Região Origem (Norte)	% Nacionalidade (Portuguesa)	% Isenção Propinas (Sim)	% Associativo (Sim)	% Bolseiro (Sim)
Álgebra Linear	GT	49	25.3	71.4	20.4	77.6	4.1	18.4	6.1
	GC - Todos os Cursos	525	24.8	71.0	27.2	85.5	2.1	26.5	22.3
	GC - Área de Gestão	118	26.1	41.5	29.7	85.6	0.8	47.5	33.1
Matemática I	GT	59	24.7	67.8	18.6	78.0	5.1	20.3	8.5
	GC - Todos os Cursos	794	24.9	64.2	24.4	85.3	2.0	27.6	22.2
	GC - Área de Gestão	299	25.6	41.8	23.4	83.9	1.3	35.1	27.8
Matemática II	GT	62	24.9	69.4	21.0	77.4	4.8	17.7	9.7
	GC - Todos os Cursos	-----	-----	-----	-----	-----	-----	-----	-----
	GC - Área de Gestão	321	25.8	43.6	24.3	83.8	0.9	34.3	26.2
Estatística	GT	47	25.4	48.9	25.5	89.4	2.1	34.0	25.5
	GC - Todos os Cursos	621	25.7	57.5	27.2	85.7	1.3	41.1	28.5
	GC - Área de Gestão	217	26.2	37.8	27.2	86.6	1.4	46.5	34.1
Investigação Operacional	GT	26	25.4	50.0	34.6	92.3	3.8	19.2	26.9
	GC - Todos os Cursos	419	26.1	55.6	27.9	88.5	1.2	39.4	28.4
	GC - Área de Gestão	167	26.2	38.3	27.5	86.8	1.2	49.1	36.5
Introdução à Informática	GT	23	25.6	73.9	21.7	73.9	4.3	4.3	21.7
	GC - Todos os Cursos	-----	-----	-----	-----	-----	-----	-----	-----
	GC - Área de Gestão	150	25.4	36.7	20.0	80.0	0.7	13.3	24.7
Marketing	GT	43	25.0	39.5	27.9	79.1	0.0	20.9	37.2
	GC - Todos os Cursos	78	26.6	25.6	29.5	76.9	0.0	39.7	35.9
	GC - Área de Gestão	71	26.7	28.2	31.0	74.6	0.0	40.8	32.4
Contabilidade Financeira I	GT	40	25.1	65.8	13.2	71.1	0.0	18.4	7.9
	GC - Todos os Cursos	78	24.5	47.4	20.5	80.8	0.0	14.1	19.2
	GC - Área de Gestão	71	24.7	50.7	22.5	80.3	0.0	14.1	14.1
Contabilidade Financeira II	GT	57	25.8	63.6	25.5	76.4	0.0	18.2	7.3
	GC - Todos os Cursos	-----	-----	-----	-----	-----	-----	-----	-----
	GC - Área de Gestão	97	24.9	44.3	18.6	79.4	0.0	22.7	16.5

Na Tabela 3 apresentou-se a distribuição das características seleccionadas pelas disciplinas analisadas, tanto para o grupo de tratamento como para os grupos de comparação.

É possível verificar que existem diferenças na distribuição das características seleccionadas entre os grupos de análise. Esta diferente distribuição de características mostra a importância de se proceder a um processo de *matching* entre os alunos dos diferentes grupos para perceber qual o verdadeiro efeito causal da política nos alunos de IG. De facto, a diferença entre os alunos em termos de características observáveis pode estar na origem da diferença entre os seus resultados e assim, afirmar-se que a diferença entre os resultados dos alunos de diferentes cursos - relativamente ao curso de IG - é consequência da introdução da política em IG e não nos restantes é abusiva e pode originar erros de interpretação.

Em resultado da constatação que existem diferenças entre as características dos alunos nos diferentes cursos, foi estimado um modelo *logit* binomial - cujos parâmetros são as variáveis correspondentes às características seleccionadas na Tabela 3 - para a obtenção da probabilidade de um aluno estar inscrito num curso de IG e não num outro curso⁸, isto é, para estimar a *propensity score* que permitirá o ajustamento entre os alunos.

Da estimação resulta que, embora nem todas as variáveis sejam estatisticamente significativas ao nível individual, elas apresentam significância estatística conjunta mostrando que são relevantes para a determinação

⁸ Neste trabalho não serão apresentados os resultados do modelo *logit*, pela extensão dos mesmos e pelas particularidades econométricas cuja dificuldade de explicação pensamos dever ficar fora do texto do estudo. No entanto, os resultados serão fornecidos caso haja interesse e sempre que forem necessárias explicações adicionais.

da probabilidade de participar no curso. Aliás é possível mostrar os resultados das taxas de previsão correcta (TPC) da participação, isto é, é possível mostrar como é que os modelos *logit* paramétricos estimados prevêm, de forma correcta, os alunos que efectivamente são alunos de IG. Os resultados constam na Tabela 4.

Tabela 4: Taxa de previsão correcta da participação de alunos no curso de IG.

CURSOS		DISCIPLINAS								
		Álgebra Linear	Matemática I	Matemática II	Estatística	Investigação Operacional	Introdução à Informática	Marketing	Contabilidade Financeira I	Contabilidade Financeira II
TPC (%)	TODOS	71.43	67.8	66.13	57.45	76.92	69.57	74.42	76.32	61.82
	ÁREA GESTÃO	77.55	62.71	-----	63.83	73.08	-----	72.09	76.32	-----

De facto todos os modelos apresentam taxas correctas de previsão da participação em disciplinas do curso de IG acima dos 57%, chegando alguns modelos a apresentar taxas perto dos 80%. As taxas de previsão consideram-se muito boas o que indicia que as variáveis seleccionadas são um excelente ponto de partida para permitir encontrar nos grupos de comparação alunos idênticos, ou quase idênticos, aos alunos de IG de acordo com a sua probabilidade de participação no curso de IG condicionada num vector de características observáveis⁹.

Calculada a probabilidade de participação - a partir das características que auxiliam o processo de *matching* - no curso de IG para cada aluno, em cada disciplina, foi encontrado para cada um dos alunos do curso de IG aquele aluno no grupo de comparação que possui a mesma probabilidade. Assim, a média dos resultados para os alunos de IG será agora comparada com a média dos resultados daqueles que lhe são idênticos mas frequentam a mesma disciplina num outro curso. A média dos resultados dos alunos que serviram de comparação, representa o resultado que o aluno de IG teria obtido se não fosse obrigado a assistir às aulas, não estivesse sujeito a uma avaliação contínua e não fosse acompanhado pelo docente de uma forma mais sistemática.

Agora é possível obter uma aproximação mais rigorosa ao verdadeiro efeito causal da introdução da nova metodologia, naqueles que a ela foram sujeitos.

Os resultados serão apresentados de seguida, mas antes é importante referir a qualidade do procedimento. De facto, ao ajustarem-se os alunos, de acordo com a sua semelhança em termos de probabilidade de participação, pretende-se que eles se assemelhem o mais possível em termos de distribuição das características observáveis. Em termos estatísticos, significa que o teste de igualdade entre a média de cada variável para os grupos em análise deve possuir significância estatística.

Com reduzidas excepções, para todas as variáveis na análise concreta de cada uma das disciplinas, foi encontrada significância estatística para a igualdade entre a média de cada uma das variáveis entre os grupos. A quantidade de variáveis e de disciplinas analisadas não permite que seja possível a apresentação desses resultados. No entanto, pode ser apresentado o resultado do cálculo do valor absoluto médio do enviesamento

⁹ Os modelos não produzem taxas de previsão iguais a 100%, nem esperávamos que tal ocorresse pois, tal como já admitimos, verifica-se a inexistência de informação que poderia explicar os resultados (como a média de entrada no curso ou a prova específica de acesso, por exemplo) ou então existe informação que não pode ser acedida (como a motivação do aluno na disciplina).

padronizado¹⁰, relativamente ao conjunto de características observáveis, entre o grupo de tratamento e os grupos de comparação antes e depois da aplicação do nosso procedimento de *matching*.

Tabela 5: Média absoluta do enviesamento padronizado entre o grupo de tratamento e os grupos de comparação antes e depois do procedimento de *matching*.

GC	Enviesamento	Álgebra Linear	Matemática I	Matemática II	Estatística	Investigação Operacional	Introdução à Informática	Marketing	Contabilidade Financeira I	Contabilidade Financeira II
TODOS	Antes	18.58	16.87	29.66	9.86	17.58	23.68	21.35	23.6	21.03
	Depois	0.54	6.58	8.37	9.94	12.6	22.83	5.82	10.98	14.37
ÁREA GESTÃO	Antes	39.92	30.06	-----	14.83	25.45	-----	23.87	19.95	-----
	Depois	8.43	4.81	-----	13.14	10.07	-----	13.43	18.35	-----

A Tabela 5 permite observar que a média do enviesamento padronizado entre os grupos é, em geral, bastante elevado antes de se proceder ao *matching* entre os alunos, reduzindo-se bastante após este procedimento. Esta redução significativa do enviesamento médio entre os grupos é um bom indicador da qualidade do procedimento de *matching* traduzindo-se, na prática, pela indicação de que foi possível encontrar nos grupos de comparação alunos muito idênticos aos alunos de IG, no conjunto das características analisadas.

Assim, é possível olhar para os resultados do efeito causal da introdução da nova metodologia de ensino com confiança e a certeza que serão a melhor aproximação para o efeito real da metodologia naqueles que a ela estiveram sujeitos. Estes resultados encontram-se na Tabela 6.

Começando por observar o efeito da metodologia nos alunos de IG, admitindo a possibilidade de eles frequentarem a disciplina em qualquer curso da ESTiG-IPB que a possua na sua estrutura curricular, verifica-se que, com a exceção das disciplinas de Estatística, Marketing e Contabilidade Financeira I, os alunos de IG teriam obtido melhores resultados em termos de aprovação nas disciplinas de Álgebra Linear, Matemática I e Investigação Operacional se as tivessem frequentado num outro curso que não IG. Quanto aos alunos de IG, nas disciplinas de Álgebra Linear, Matemática I e Investigação Operacional, os resultados mostram que eles possuem menos 4 a 19%, respectivamente, de probabilidade de obter uma classificação positiva do que se tivessem frequentado as disciplinas noutros cursos leccionados na ESTiG, independentemente da área científica em questão.

Tabela 6: Percentagem de alunos aprovados após realizarmos o *matching* entre alunos.

DISCIPLINAS	CURSOS DE COMPARAÇÃO					
	TODOS			ÁREA DE GESTÃO		
	GT	GC	Diferença (ΔP)	GT	GC	Diferença (ΔP)
Álgebra Linear	26.5	30.6	- 4.1	26.5	16.3	10.2
Matemática I	18.6	30.5	- 11.9	19.0	13.8	5.2
Matemática II	-----	-----	-----	27.9	11.5	16.4
Estatística	53.2	46.8	6.4	53.2	19.1	34.0
Investigação Operacional	34.6	53.8	- 19.2	34.6	23.1	11.5

¹⁰ Enviesamento padronizado: $\frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{(V_1(X_1) + V_0(X_0))/2}}$, onde $\bar{X}_1(V_1)$ é a média (variância) no grupo de tratamento e $\bar{X}_0(V_0)$ é a simbologia análoga para o grupo de comparação.

Introdução à Informática	----	----	----	30.4	26.1	4.3
Marketing	92.9	73.8	19.0	95.1	65.9	29.3
Contabilidade Financeira I	73.0	27.0	45.9	73.0	29.7	43.2
Contabilidade Financeira II	----	----	----	56.4	16.4	40.0

No entanto, se compararmos os alunos apenas com os colegas de cursos da área da gestão, verifica-se que os alunos de IG beneficiaram bastante por frequentarem o curso de IG e não os cursos de CA ou GE. Se olharmos com cuidado, verifica-se que este resultado é uma consequência dos baixos níveis de aprovação nos cursos de CA e GE.

Curioso é a observação de que a disciplina de Informática de Gestão é a que parece apresentar o menor efeito positivo para os alunos de IG. Uma explicação poderá ser a obrigatoriedade de, qualquer que seja o curso, o aluno se sujeitar a avaliação contínua o que dilui o efeito da introdução da metodologia no curso de IG. Os alunos do curso de IG possuem apenas cerca de 4% mais probabilidade de obter aprovação na disciplina de Introdução à Informática do que se tivessem frequentado a disciplina num dos outros cursos da área da gestão. De facto, se analisarmos, por exemplo, a disciplina de Estatística verifica-se que é apenas obrigatória uma avaliação contínua para o curso de IG, não se prevendo esta alternativa de avaliação nos restantes cursos, e que o efeito causal da política indica que os alunos de IG possuem cerca de 34% mais probabilidade de serem aprovados na disciplina do que se tivessem frequentado a disciplina num outro curso da área da Gestão.

Os factos mencionados servem como pistas de análise para avaliar o efeito da introdução da metodologia. Para auxiliar essa análise é também interessante comparar os resultados obtidos antes e depois do procedimento de *matching*. Pode observar-se que não existem grandes diferenças nos resultados antes e depois da aplicação do procedimento de *matching* quando se analisa a instituição como um todo, no entanto, tal já não é verdadeiro para a comparação entre os cursos na área da Gestão. Neste último caso, verifica-se que o ajustamento entre alunos permite obter resultados bastante mais positivos para o efeito causal da introdução da nova metodologia de ensino naqueles que a ela foram sujeitos. Para a análise em que se faz a comparação com todos os cursos da ESTiG os resultados menos claros podem indiciar que existam factores relacionados com as áreas científicas, específicas de cada curso, que podem influenciar os resultados e não foram captadas pela presente avaliação.

4.2. Efeito da introdução da metodologia na classificação positiva

Para a análise da qualidade do sucesso escolar foram apenas seleccionados os alunos que obtiveram uma classificação positiva, ou seja, maior ou igual a 10. Para estes indivíduos é apresentada a distribuição de características observáveis, à semelhança do que se fez na secção anterior. Ver Tabela 7.

Como é óbvio temos uma diminuição do número de observações adequadas para análise o que deteriora um pouco, embora não restrinja, um procedimento de *matching*. De facto, a constatação de que existem diferenças na distribuição do conjunto de características mostra que é importante realizar esse procedimento, tal como já se tinha concluído para a avaliação do sucesso escolar. Como o procedimento é exactamente o mesmo não iremos repetir as explicações já fornecidas limitando-nos a apresentar os resultados mais significativos.

Um desses resultados a salientar refere-se ao pior ajustamento entre alunos de IG e os colegas. A diminuição do número de observações, em ambos os grupos, deixava adivinhar que tal pudesse acontecer. A

degradação da qualidade do procedimento de *matching*, como pode ser constatado na Tabela 8, é a consequência prática de, com menos alunos, ser mais difícil encontrar, para cada aluno de IG, um aluno semelhante no conjunto de características seleccionadas para a análise.

Tabela 7: Distribuição das características observáveis, seleccionadas para o processo de *matching*, entre os grupos de tratamento e comparação.

DISCIPLINAS	GRUPOS	CARACTERÍSTICAS							
		Alunos	Idade	% Sexo (Masculino)	% Região Origem (Norte)	% Nacionalidade (Portuguesa)	% Isenção Propinas (Sim)	% Associativo (Sim)	% Bolseiro (Sim)
Álgebra Linear	GT	13	23.8	30.8	23.1	92.3	0.0	30.8	7.7
	GC - Todos os Cursos	125	23.3	57.6	32.0	93.6	0.0	27.2	49.6
	GC - Área de Gestão	43	24.0	30.2	34.9	93.0	0.0	44.2	58.1
Matemática I	GT	11	23.6	36.4	27.3	63.6	0.0	27.3	18.2
	GC - Todos os Cursos	229	23.9	56.3	23.1	88.6	0.0	31.0	39.7
	GC - Área de Gestão	62	24.3	24.2	25.8	85.5	0.0	50.0	56.5
Matemática II	GT	8	23.9	37.5	37.5	62.5	0.0	25.0	25.0
	GC - Todos os Cursos	----	----	----	----	----	----	----	----
	GC - Área de Gestão	57	24.5	24.6	21.1	84.2	0.0	42.1	47.4
Estatística	GT	13	25.2	38.5	23.1	92.3	0.0	30.8	46.2
	GC - Todos os Cursos	71	24.8	39.4	26.8	87.3	0.0	38.0	43.7
	GC - Área de Gestão	16	24.5	0.0	31.3	93.8	0.0	31.3	68.8
Investigação Operacional	GT	9	24.4	33.3	33.3	88.9	0.0	11.1	55.6
	GC - Todos os Cursos	179	25.0	52.0	24.6	91.1	0.0	37.4	36.9
	GC - Área de Gestão	55	24.6	16.4	23.6	96.4	0.0	47.3	56.4
Introdução à Informática	GT	7	26.1	85.7	42.9	85.7	0.0	0.0	42.9
	GC - Todos os Cursos	----	----	----	----	----	----	----	----
	GC - Área de Gestão	49	23.7	40.8	10.2	81.6	0.0	18.4	40.8
Marketing	GT	39	24.9	33.3	25.6	76.9	0.0	20.5	35.9
	GC - Todos os Cursos	56	25.9	23.2	30.4	82.1	0.0	46.4	46.4
	GC - Área de Gestão	49	26.0	26.5	32.7	79.6	0.0	49.0	42.9
Contabilidade Financeira I	GT	13	24.5	53.8	23.1	84.6	0.0	30.8	7.7
	GC - Todos os Cursos	23	23.4	30.4	17.4	87.0	0.0	17.4	39.1
	GC - Área de Gestão	18	23.4	33.3	22.2	83.3	0.0	16.7	27.8
Contabilidade Financeira II	GT	19	25.2	36.8	26.3	84.2	0.0	26.3	10.5
	GC - Todos os Cursos	----	----	----	----	----	----	----	----
	GC - Área de Gestão	32	23.9	37.5	18.8	78.1	0.0	37.5	25.0

Por outro lado, ao considerarmos que os resultados devem ser obtidos numa região de suporte comum, ou seja, num intervalo em que todos os alunos possuam a mesma distribuição na probabilidade de participação, foi necessário abandonar algumas observações no grupo de tratamento (no máximo 2 observações correspondentes a alunos de IG). Estes alunos foram retirados da análise porque se concluiu que seria impossível encontrar nos restantes cursos alunos com uma probabilidade de participação idêntica, isto é, não seria possível encontrar para esses alunos de IG um *match* adequado.

Tabela 8: Média absoluta do enviesamento padronizado entre o grupo de tratamento e os grupos de comparação antes e depois do procedimento de *matching*.

GC	Enviesamento	Álgebra Linear	Matemática I	Matemática II	Estatística	Investigação Operacional	Introdução à Informática	Marketing	Contabilidade Financeira I	Contabilidade Financeira II
TODOS	Antes	35.96	29.21	36.28	9.85	30.2	54.47	25.87	38.2	24.18
	Depois	1.38	24.50	40.81	28.37	20.68	30.43	10.57	32.53	40.68

ÁREA	Antes	31.43	39.03	-----	25.53	29.73	-----	24.35	30.35	-----
GESTÃO	Depois	17.25	39.50	-----	51.25	34.35	-----	18.9	19.8	-----

Apesar da pior qualidade do *matching*, percebida em termos do valor absoluto médio do enviesamento padronizado, verificou-se que os testes estatísticos à igualdade das médias, de cada uma das características seleccionadas, entre os grupos são significativos na maioria das variáveis¹¹.

Feitas as devidas ressalvas, podem agora analisar-se os resultados do procedimento de *matching* na qualidade da classificação positiva dos alunos de IG. Esses resultados apresentam-se na Tabela 9.

Tabela 9: Média das classificações positivas após realizarmos o *matching* entre alunos.

DISCIPLINAS	CURSOS DE COMPARAÇÃO					
	TODOS			ÁREA DE GESTÃO		
	GT	GC	Diferença (ΔP)	GT	GC	Diferença (ΔP)
Álgebra Linear	11.50	11.50	0.00	11.38	11.38	0.00
Matemática I	11.27	11.91	-0.64	11.27	11.45	-0.18
Matemática II	-----	-----	-----	11.75	12.13	-0.38
Estatística	10.73	10.45	0.27	10.73	10.18	0.55
Investigação Operacional	11.00	11.33	-0.33	11.00	13.22	-2.22
Introdução à Informática	-----	-----	-----	12.33	10.67	1.67
Marketing	11.31	11.54	-0.23	11.31	11.51	-0.21
Contabilidade Financeira I	10.45	11.00	-0.55	10.45	10.91	-0.45
Contabilidade Financeira II	-----	-----	-----	11.21	11.84	-0.63

A primeira constatação é que se os alunos de IG possuem maior probabilidade de obter aprovação escolar por frequentarem disciplinas no curso de IG e não noutros cursos, especialmente nos da área da gestão, tal não acontece em termos do valor das suas classificações positivas. Com algumas excepções, os alunos de IG obteriam melhores classificações se frequentassem as disciplinas em causa, noutros cursos. As excepções são a disciplina de Informática de Gestão - verifica-se que os alunos de IG possuem uma média de classificações positivas superior em 1,67 valores do que se tivessem frequentado a disciplina num dos cursos da área da gestão. No outro extremo, encontra-se a disciplina de Investigação Operacional onde se verifica que os alunos que frequentam a disciplina no curso de IG possuem uma média de classificações positivas inferiores em quase 2,22 valores do que se tivessem sido aprovados na mesma disciplina mas a frequentar outro curso.

De facto, os resultados em termos de qualidade do sucesso escolar não são tão positivos para os alunos abrangidos pela medida.

5. CONCLUSÃO

Com base nos resultados obtidos conclui-se que a introdução de uma filosofia de ensino, tal como a preconizada pelo Tratado de Bolonha, é importante para a melhoria das taxas de aprovação dos alunos a que a ela ficam sujeitos. Tal verifica-se pela maior taxa de aprovação dos alunos de IG quando comparados com alunos que frequentam as mesmas disciplinas em cursos diferentes mas na mesma vizinhança da área científica.

¹¹ Como já foi explicado, a elevada dimensão dos resultados não permite a sua adequada apresentação no corpo do texto.

A não verificação desta conclusão quando o estudo se estende a todos os cursos leccionados pela ESTiG-IPB, independentemente da sua área científica, pode estar relacionada com factores específicos dos cursos que não foram analisados neste processo de avaliação.

Se as taxas de aprovação são a medida quantitativa preferida dos decisores políticos, considera-se importante que não seja a única a ser analisada.

De facto, em termos da consistência das taxas de aprovação, que no presente trabalho foi decidido designar por qualidade, não foi possível concluir pela importância da aplicação de novas metodologias de ensino.

REFERÊNCIAS BIBLIOGRÁFICAS

BLUNDELL, R. and DIAS, M. Costa; (2002); “Alternative Approaches to evaluation in Empirical Microeconomics”; *Portuguese Economic Journal*; 1(2), pp. 91-115.

DEHEJIA, R. H. and WAHBA, S.; (2002); “Propensity Score-Matching Methods for Nonexperimental Causal Studies”; *The Review of Economics and Statistics*; 84(1), pp.151-161.

LEUVEN, E. and SIANESI, B.; (2003); "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing"; <http://ideas.repec.org/c/boc/bocode/s432001.html.Version 1.2.3>

ROSENBAUM, P. R. and RUBIN, D. B.; (1983); “The Central Role of the Propensity Score in Observational Studies for Causal Effects”; *Biometrika*; 70(1), pp. 41-55.

RUBIN, D.B.; (1974); “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies”; *Journal of Educational Psychology*; 66, pp. 688-701.