



UNIVERSIDADE FEDERAL DO RIO GRANDE DO
SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Escores de Propensão: Aplicações à Epidemiologia

Autor: Maria Cláudia Schardosim C. de Souza
Orientador: Prof. Álvaro Vigo

Porto Alegre, 15 de dezembro de 2010.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Escores de Propensão: Aplicações à Epidemiologia

Autor: Maria Cláudia Schardosim C. de Souza

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Prof Dr Álvaro Vigo
Prof^a Dr^a Suzi Alves Camey

Porto Alegre, 15 de dezembro de 2010.

"Si tu savais Comme j'ai envie d'être avec toi..."

Leticia Thompson

Agradecimentos

À minha família, mãe, pai e Juh, por entender a minha ausência, mesmo muitas vezes sem entender o que eu estava fazendo. Sempre torceram pelo meu sucesso, apostaram na minha capacidade e me apoiaram de todas as maneiras que podiam.

Ao meu namorado Henrique que, mesmo do outro lado do oceano, me escuta, me entende, me apóia, e que deve ser o único “não estatístico” que não me acha doida. Se não fosse ele, hoje não estaria onde estou. Nem a distância temporária impedirá a continuação dessa relação profunda e cúmplice, que só o amor explica.

Às amigas que eu fiz na faculdade. Eu sei que sou a caçula, e que estamos em momentos diferentes da vida, mas nem por isso deixaram de ser as melhores amigas que eu poderia querer. Vou ter essa amizade como uma das coisas mais importantes que conquistei nesses quatro anos de UFRGS.

À equipe de estatística do Projeto ELSA pela amizade e compreensão há mais de dois anos. Sabem quando estou feliz ou triste e preocupada, sempre me apoiando em momentos difíceis e me ajudando a superar momentos que eu não conseguiria sozinha, principalmente com ensinamentos fundamentais para as disciplinas ao longo do curso. Às vezes eu passo mais tempo com eles do que com a minha família, mas tudo bem, pois são ótimos colegas para trabalhar. Espero ainda ter muitos anos pela frente com a equipe.

Aos excelentes professores que tive na graduação, pois todos influenciaram muito na pessoa que sou hoje, e na bacharel em estatística que pretendo ser. Professoras Suzi e Vanessa, quando eu crescer, eu quero ser uma estatística como vocês!

Mas principalmente ao Professor Álvaro Vigo, meu orientador, professor e chefe. Ele soube quando e quais emails responder, enquanto a minha ansiedade lotava a caixa de email dele. Ele sabe que eu sou “faladeira” e teve paciência e sabedoria para lidar comigo. Com a monografia, mostrou que é muito difícil ser criativa. No início do semestre ele me disse que me daria os trilhos para começar a escrever, e talvez ele não saiba, mas me mostrou onde estava o trem e me ensinou a mover a locomotiva!

Obrigada!

Resumo

Em estudos observacionais usualmente não é possível fazer alocação aleatória dos indivíduos em grupos (intervenção ou controle) e, como consequência, as distribuições de probabilidade de algumas variáveis preditoras podem ser diferentes entre os grupos. Nestas situações é importante utilizar métodos de análise que permitam controlar o potencial confundimento na associação entre os fatores e o desfecho. Os escores de propensão podem ser usados para ajustar o efeito de um tratamento ou intervenção através de pareamento, estratificação, ponderação ou como uma variável de ajuste. O pareamento de indivíduos entre os grupos faz com que as distribuições sejam similares, tornando os grupos comparáveis, de forma que as associações estimadas podem ser atribuídas unicamente à intervenção ou ao tratamento. Este trabalho explora o uso do método de escores de propensão no contexto epidemiológico, cobrindo aspectos da estimação e sua utilização para pareamento. Rotinas computacionais em R e SAS foram utilizadas para fazer o pareamento. Um conjunto de dados hipotéticos foi utilizado para ilustrar as etapas de estimação dos escores de propensão, o pareamento e as comparações entre grupos. Diferentes níveis de exigência foram usados para o pareamento. Os resultados demonstram empiricamente o grande potencial do método para estimar acuradamente (sem confundimento) o efeito ou associação de uma intervenção com o desfecho. O nível de exigência na similaridade dos escores de propensão dos indivíduos entre os grupos tem impacto sobre o número de pares encontrados e, portanto, deve ser escolhido com cautela. No exemplo, as análises sem e com escores de propensão e com ajuste multivariável ilustram o papel do confundimento, pois a direção e a magnitude da associação podem ser substancialmente diferentes.

Palavras-chave: Escores de propensão, pareamento, confundimento, ajuste multivariável

Abstract

In observational studies it is not always possible to randomly allocate individuals to groups (treated and control) and as consequence, the distributions of some predictors may be different among groups. In these situations, it's important use methods of analysis that allows to control for potential confounding the association of the outcome with factors. The propensity scores can be used to adjust the treatment effect by means of matching, stratification, weighting or regression adjustment. The matching with subjects between groups can be used to balance the covariates distributions, making groups more comparable, so that the differences can be assigned to intervention or treatment. This paper explores the use of propensity score methods in epidemiological context, focusing on the estimation and their use for matching. An hypothetical data set was used to show the steps of propensity scores estimation, matching and comparison between groups. A SAS macro called "**%Match**" was used to do the matching. The results show empirically the great potential of propensity score matching technique to estimate accurately (without confounding) the effect or association among the intervention and the outcome. In this example the analyses using and not using propensity scores, as well as with multivariable adjustment, illustrate the role of confounding, since both the direction and magnitude of association may be very different.

Sumário

1 Introdução.....	8
2 Métodos.....	10
2.1 Definições formais.....	14
2.2 Métodos de estimação dos escores de propensão.....	16
2.3 Utilização dos escores de propensão.....	17
2.4 Recursos computacionais.....	20
2.4.1 R.....	20
2.4.2 SAS.....	22
3 Aplicação.....	24
4 Considerações finais.....	30
5 Referências.....	33
6 Anexos	
6.1 Anexo 1 – Rotina computacional em R com a função ps do pacote twang.....	35
6.2 Anexo 2 – Anexo 2 – Rotina computacional em SAS com a macro MATCH.....	38

1 Introdução

Em estudos clínicos ou epidemiológicos, estudos experimentais são feitos para avaliar uma nova intervenção ou tratamento com respeito a um determinado desfecho. Estudos com alocação ao acaso dos indivíduos aos grupos são mais recomendados, pois a aleatorização tende a tornar os grupos comparáveis com o aumento do tamanho da amostra, de tal forma que os efeitos (ou associações) entre o desfecho e os fatores podem ser atribuídos à intervenção.

Entretanto, quando não é possível alocar ao acaso os indivíduos aos grupos, é necessário utilizar métodos que permitam ajustar os efeitos ou associações pelas possíveis diferenças nas distribuições dos preditores entre os grupos. Existem várias técnicas de ajuste, tais como estratificação, pareamento ou ajuste multivariável. No contexto epidemiológico, também é comum utilizar escores de propensão para redução de viés, que é uma forma eficiente para corrigir as diferenças das distribuições dos preditores entre os grupos.

Neste trabalho são explorados aspectos do método de escores de propensão, contemplando definições básicas, estimação e sua utilização em análises de estudos epidemiológicos. Existem vários métodos para estimar os escores de propensão, mas o mais usual é utilizar o modelo de regressão logística. Estes escores também podem ser utilizados de diferentes maneiras nas análises, sendo a forma mais usual o pareamento dos indivíduos entre grupos com valores similares nos escores de propensão.

O objetivo do trabalho é explorar e caracterizar a aplicação de escores de propensão para análises de dados epidemiológicos, com enfoque no pareamento dos indivíduos, utilizando esses escores. A técnica é exemplificada utilizando um conjunto de dados hipotéticos e rotinas computacionais dos programas R e SAS.

O Capítulo 2 apresenta uma breve revisão da literatura e definições formais do método de escores de propensão e aspectos computacionais. No Capítulo 3 é ilustrada a aplicação dos escores de propensão quando existem diferenças nas distribuições dos preditores entre os grupos. Foram realizadas análises com e sem ajuste multivariável, bem como com

pareamento, ilustrando a necessidade de métodos que permitam a redução de vieses dos resultados. Nos anexos foram incluídas as rotinas computacionais R e SAS utilizadas.

2 Métodos

Ensaio clínico randomizado são considerados padrão-ouro para avaliar a eficácia de medicamentos, tratamentos ou procedimentos clínicos. A alocação aleatória permite a geração de grupos verdadeiramente comparáveis, de modo que cada participante tem a mesma probabilidade de pertencer a um dos grupos (intervenção ou controle), desde que todos preencham os critérios de elegibilidade. Com isso, a distribuição dos fatores relacionados ao desfecho tende a ser igual nos dois grupos, e as associações encontradas podem ser atribuídas à intervenção.

Entretanto, existem situações nas quais não é possível fazer alocação aleatória dos indivíduos aos grupos. Em pesquisas para a prevenção de doenças crônicas, por exemplo, ensaios clínicos randomizados frequentemente não são factíveis, principalmente devido ao tamanho, tempo, custo, restrições éticas ou capacidade de generalização.

Estudos não-experimentais enfrentam críticas por não considerar a aleatorização dos sujeitos aos grupos, podendo gerar viés de seleção. Como o pesquisador não tem controle sobre a alocação dos indivíduos aos grupos, as distribuições de probabilidades das covariáveis (exposições ou confundidores) podem ser diferentes e, como consequência, os grupos não são totalmente comparáveis, podendo produzir estimativas viesadas do “efeito” (ou associação) do tratamento ou da exposição sobre o desfecho. Nesse tipo de situação é necessário utilizar métodos de análise que permitam reduzir viés de seleção e confundimento¹.

A primeira tentativa de usar um único escore derivado de várias covariáveis para o ajuste de confundimento foi proposto por O.S. Miettinen em 1976². Rosenbaum e Rubin (1983) desenvolveram o conceito de escores de propensão estimados utilizando as observações realizadas na linha de base para controlar viés em estudos de coorte². O método foi originalmente proposto como uma forma de corrigir o “desbalanceamento” das distribuições dos preditores (covariáveis) em estudos observacionais, quando a aleatorização não é possível³, mas também tem sido utilizado em estudos experimentais, principalmente em pesquisas na área de farmacoepidemiologia.⁴

O escore de propensão é definido com a probabilidade (ou a propensão) de um indivíduo da amostra ser alocado a um determinado grupo (uma intervenção, como um tratamento ou medicamento), condicional aos valores dos preditores (exposições ou confundidores). Sua utilização como método para a redução de viés pode ser particularmente eficiente por que indivíduos que possuem o mesmo valor do escore de propensão (ou parecidos) têm as mesmas distribuições de probabilidades nos preditores e, por isso, os grupos podem ser comparados³.

O número de publicações na área clínica e epidemiológica utilizando o método de escores de propensão tem aumentado nos últimos anos. Até 1998 existiam apenas oito artigos publicados, e aumentou para 71 artigos em 2003⁵. Uma pesquisa realizada no Pubmed com o termo “propensity scores” retornou 755 publicações (pesquisa realizada em 11/11/2010 limitando a busca para humanos e publicações a partir de 01/01/2009). A sintaxe completa da pesquisa realizada é mostrada abaixo:

```
("propensity score"[MeSH Terms] OR ("propensity"[All Fields] AND "score"[All Fields]) OR "propensity score"[All Fields] OR ("propensity"[All Fields] AND "scores"[All Fields]) OR "propensity scores"[All Fields]) AND ("humans"[MeSH Terms] AND ("2009/01/01"[PDAT] : "3000"[PDAT]))
```

Limits Activated: Humans, Publication Date from 2009/01/01)

Diferenças entre os grupos produzidas por viés de seleção fazem com que os grupos já sejam diferentes na linha de base. Se os grupos não são comparáveis, as conclusões sobre o efeito do tratamento ou da exposição podem ser viesadas.

O uso de escores de propensão para fazer pareamento em estudos na área de cardiologia está aumentando, mas ainda é pequeno e mal implementado⁶. Esse crescimento deve-se ao fato de que na cardiologia, na maioria das vezes a alocação aleatória dos sujeitos do estudo não é possível, fazendo com que as distribuições das covariáveis sejam diferentes

na linha de base. No entanto, essas diferenças podem ser equilibradas com o uso de escores de propensão.

Uma revisão das publicações da área de cardiologia entre 1996 e 2003 concluiu que dos 47 estudos examinados, apenas dois descreveram de maneira completa e correta os métodos estatísticos que foram utilizados para a análise usando pareamento por escores de propensão. Entre os anos de 2004 e 2006, foram examinados 60 estudos, e nenhum descreveu adequadamente os métodos utilizados para o pareamento dos dados⁶.

Antes do uso de escores de propensão, a técnica mais comum utilizada para fazer pareamento de observações era a métrica de Mahalanobis, usualmente com uma grande quantidade de preditores. A distância de Mahalanobis consiste em ordenar aleatoriamente os sujeitos e, então, calcular a distância entre o primeiro do grupo tratado e todos os sujeitos do grupo controle, para todos os sujeitos tratados e sujeitos controle.

Em um estudo com 4144 pacientes sobre o efeito de estatina como terapia em pacientes com infarto do miocárdio, os escores de propensão foram estimados usando o modelo de regressão logística de duas formas: a primeira incluiu um número restrito de variáveis (38 variáveis mais quatro termos quadráticos) e a segunda um número mais amplo (52 variáveis mais 6 termos quadráticos). Na comparação dos resultados das análises entre os grupos não pareados e os grupos pareados usando o modelo reduzido, foi possível observar a eficácia dos escores de propensão para equilibrar os grupos, que antes tinham médias significativamente diferentes na linha de base. Com o uso do pareamento pelos escores de propensão, passaram a ser comparáveis, pois as distribuições das variáveis preditoras se tornaram semelhantes. Porém, a comparação das análises das variáveis observadas na linha de base revelou diferenças entre os pareamentos realizados com o modelo reduzido e do modelo expandido. Utilizando o modelo reduzido, para algumas variáveis preditoras havia diferenças significativas nas distribuições entre os grupos, enquanto o uso do modelo expandido equilibrou todas as distribuições das variáveis preditoras⁷.

Foram realizadas comparações considerando três estratégias de análise: análise sem pareamento e pareamento com escores de propensão estimados pelos modelos reduzido e

expandido. Na análise sem pareamento, a estimativa de razão de azares (“*hazard ratio*”) para a ocorrência de infarto agudo do miocárdio em indivíduos que receberam terapia com estatina em relação aos que não receberam foi $HR=2,11$ (IC 95%: 1,46-3,04). Utilizando pareamento por escores de propensão estimados pelo modelo reduzido esta estimativa foi $HR=1,01$ (IC 95%: 0,62-1,29). Porém, quando os escores de propensão foram estimados pelo modelo expandido, a estimativa de razão de azares foi $HR=0,69$ (IC 95%: 0,52-0,93), indicando que o resultado da análise sem pareamento tem forte confundimento, o qual é corrigido pelo uso do pareamento utilizando escores de propensão estimados pelo modelo expandido⁷. Nesse trabalho, foi avaliado como melhor modelo, aquele que estimou a menor razão de azares, isto é, o modelo expandido.

Uma vez que os escores já foram calculados, eles podem ser utilizados de várias formas para reduzir viés de seleção ou o confundimento em estudos de coorte não-experimentais⁵. As estratégias mais comuns são o pareamento utilizando os escores de propensão, a ponderação e a estratificação, ou ainda uma combinação dessas três estratégias³.

Vários métodos podem ser usados para estimar os escores de propensão, tais como análise discriminante, árvores de decisão e regressão logística, os quais podem ser ajustados utilizando programas estatísticos usuais, tais como SAS, IBM - SPSS, R, STATA, etc. Porém, a utilização dos escores de propensão para fazer pareamento dos indivíduos entre os grupos exige rotinas computacionais especiais que serão discutidas no Capítulo 3. A próxima seção aborda alguns aspectos conceituais do método e as seções 2.2 e 2.3 resumem, respectivamente, aspectos da estimação e da utilização dos escores de propensão. A Seção 2.4 mostra duas rotinas computacionais para o pareamento de indivíduos usando escores de propensão.

2.1 Definições formais

Os aspectos explorados desta seção baseiam-se nos artigos de Rosenbaum & Rubin² e D'Agostino³, publicados em 1983 e 1998, respectivamente. Considere uma amostra aleatória de n indivíduos selecionada de determinada população e que cada indivíduo pode ser alocado a um de dois grupos. Sem perda de generalidade, seja a variável que Z representa os grupos, tal que

$$Z_i = \begin{cases} 1, & \text{Intervenção} \\ 0, & \text{Controle} \end{cases}$$

$\forall i = 1, 2, \dots, n$. Ainda, considere o vetor de covariáveis contendo exposições e confundidores $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ associado ao i -ésimo indivíduo. Assim, o escore de propensão é definido como a probabilidade de alocar um indivíduo ao grupo intervenção, condicional aos valores observados das covariáveis², ou seja,

$$e_i(\mathbf{x}_i) = P(Z_i = 1 | \mathbf{x}_i), \quad \forall i = 1, 2, \dots, n.$$

Em experimentos aleatorizados, os resultados dos dois grupos freqüentemente podem ser diretamente comparados porque as unidades experimentais provavelmente são similares, enquanto que nos não aleatorizados, comparações diretas podem gerar resultados incorretos, pois as unidades para as quais foi alocada a intervenção podem ser sistematicamente diferentes das do grupo controle. Escores de equilíbrio ou de balanço ("*balancing scores*") podem ser usados para agrupar as unidades do grupo intervenção e controle tal que comparações diretas podem ser realizadas de forma correta².

Um escore de equilíbrio $b(\mathbf{x})$ é uma função dos preditores observados \mathbf{x} tal que a distribuição condicional de \mathbf{x} dado $b(\mathbf{x})$ é mesma para as unidades do grupo intervenção e controle. A função de escore de equilíbrio mais trivial é $b(\mathbf{x}) = \mathbf{x}$, mas a mais robusta ("*coarsest function*") é o escore de propensão².

Experimentos aleatorizados e não aleatorizados diferem entre si, pois os primeiros têm a distribuição de Z_i determinada pelo mecanismo aleatório especificado. Em um ensaio

aleatorizado o escore de propensão é uma função conhecida, ou seja, existe uma especificação aceita para $e(\mathbf{x})$, ao passo que nos ensaios não aleatorizados, o escore de propensão é quase sempre uma função desconhecida, tal que não existe uma especificação para $e(\mathbf{x})$. Contudo, $e(\mathbf{x})$ pode ser estimado a partir dos dados observados, utilizando, por exemplo, um modelo logístico².

Assumindo que cada indivíduo i ($\forall i = 1, 2, \dots, n$) recebe apenas a intervenção ou o controle, a resposta observada pode ser representada por r_{1i} (no grupo intervenção) ou r_{0i} (para o grupo controle). Em um ensaio aleatorizado, a alocação do tratamento z e a resposta são condicionalmente independentes, dado o vetor de preditores \mathbf{x} , porém isto usualmente não está satisfeita para experimentos não aleatorizados.

Se a alocação do tratamento é fortemente ignorável, então para cada valor específico do escore de propensão, a diferença entre as médias dos grupos intervenção e controle é uma estimativa não viesada do efeito médio da intervenção para aquele valor do escore de propensão. Assim, nesta situação, o pareamento, a subclassificação e o ajuste multivariável baseados no escore de propensão produzem estimativas não viesadas do efeito da intervenção ou tratamento².

Intuitivamente, o escore de propensão representa uma medida da probabilidade de que um indivíduo seja alocado ao grupo de intervenção, baseado apenas nos valores observados para os preditores³.

Estudos nos quais a alocação aleatória não é possível, os escores de propensão são usados para ajustar o efeito do tratamento ou intervenção, fazendo com que o estudo possa ser classificado como “quase-experimental”, pois apresenta resultados muito similares aos de estudos experimentais aleatorizados³.

O pareamento usando escores de propensão é uma forma eficaz de eliminar as diferenças entre os grupos causadas pelas variáveis que forem incluídas no modelo utilizado para estimar esses escores. O uso dos escores de propensão é preferível em relação a um

modelo com todas essas variáveis, principalmente pela parcimônia, ou seja, para manter o modelo o mais simples possível.

Se os sujeitos do grupo de tratamento e controle têm o mesmo (ou parecido) escore de propensão, então tendem a ter as mesmas (ou parecidas) distribuições de probabilidade para os preditores, de tal forma que o ajustamento usando o escore de propensão em média removerá os vieses subjacentes as distribuições dessas variáveis.

Não é muito comum encontrar exatamente o mesmo escore para dois sujeitos, cada um em um grupo, de tal forma que é usual definir intervalos nos quais podemos assumir que os escores são aproximadamente equivalentes. Uma maneira usual é a utilização de quartis, decis ou quintis dependendo do tamanho da amostra e dos escores obtidos. Adicionalmente, muitas vezes a relação funcional entre o escore de propensão e o desfecho pode não ser linear, e nestes casos também pode ser mais apropriado utilizar quantis dos escores de propensão, considerados no modelo como variável categórica⁸.

O método de escores de propensão pode ser usado tanto em tamanhos de amostras grandes quanto em amostras de tamanhos pequenos, sendo suficiente para remover as diferenças nas distribuições das covariáveis observadas². Deve-se ter cuidado, porém, quando o tamanho da amostra é pequeno, pois alguns métodos de análise exigem que este tamanho seja grande. Em particular, o pareamento utilizando escores de propensão usualmente necessita um tamanho de amostra grande.

2.2 Métodos de estimação dos escores de propensão

Escore de propensão podem ser estimados de formas diferentes, mas a maneira mais utilizada é através do modelo de regressão logística, onde os escores são a probabilidade condicional de alocação a um determinado grupo dados os valores observados nas variáveis predictoras. Neste trabalho será considerado apenas o modelo de regressão logística como método para estimação do escore de propensão. Neste modelo, o desfecho é a variável dicotômica que identifica o grupo ao qual cada indivíduo foi alocado (ou a categoria de

exposição) e os preditores são as variáveis confundidoras (incluindo termos quadráticos ou de ordem maior e interações). A seleção das variáveis no modelo deve respeitar aspectos conceituais relacionando os confundidores com o desfecho, bem como critérios de significância estatística. Alguns aspectos importantes a serem considerados são as possíveis diferenças nas distribuições de probabilidade dos confundidores entre os grupos.

Existem outros métodos para o cálculo do escore de propensão, como por exemplo, análise por árvore de regressão e classificação (*Classification and Regression Tree Analysis – CART*). Esse é um método de árvore de decisão não-paramétrico que pode dividir eficientemente uma população em grupos homogêneos.

2.3 Utilização dos escores de propensão

Em estudos observacionais, os escores de propensão são usados, principalmente, para a redução de vieses decorrentes de diferenças nas distribuições dos preditores e para aumentar a precisão das estimativas de efeitos. A principal forma é o pareamento, mas também podem ser usados para ponderação, estratificação ou como variável de ajuste. Cada uma dessas técnicas é uma forma de ajustar a estimativa do efeito da intervenção pelas diferenças nas distribuições dos preditores entre os grupos³.

No pareamento, a idéia é criar pares de indivíduos dos grupos intervenção e controle com valores iguais (ou parecidos) nos escores de propensão. Existem várias maneiras para fazer isso, como por exemplo, a utilização de decis ou quartis, depende do número de intervalos que se julga melhor.

Existem vários métodos de pareamento utilizando dos escores de propensão estimados, tais como:

- Pareamento por estratificação (*Stratified Matching*)
- Pareamento pelo vizinho mais próximo (*Nearest Neighbor Matching*)
- Pareamento N : N (*N : N Matching*)
- Pareamento pelo Raio (*Radius Matching*)

- Pareamento pelo método de Kernel (*Kernel Matching*)
- Pareamento pela Métrica de Mahalanobis (*Mahalanobis Matric Matching*)
- Pareamento por calibração (*Caliper Matching*)

Rosenbaum e Rubin (1998) exemplificaram 3 técnicas para pareamento com escores de propensão, os quais serão descritos a seguir:

Pareamento pelo valor mais próximo com escores de propensão estimados (*Nearest available matching on the estimated propensity score.*)

Esse método consiste em ordenar os sujeitos de cada grupo (intervenção e controle) pelo valor do escore de propensão, identificar o primeiro indivíduo do grupo intervenção e então localizar o sujeito do grupo controle com o escore de propensão mais próximo. Ambos os sujeitos são removidos e o processo é repetido consecutivamente até formar todos os pares. Os autores sugerem usar o logito da estimativa do escore de propensão para o pareamento, ou seja,

$$\hat{q}(\mathbf{x}) = \ln \frac{1 - \hat{e}(\mathbf{x})}{\hat{e}(\mathbf{x})}$$

por que a distribuição de $\hat{q}(\mathbf{x})$ é aproximadamente normal.

Pareamento usando métrica de Mahalanobis incluindo escores de propensão (*Mahalanobis metric matching including the propensity score.*)

Esse procedimento utiliza a definição de distância de Mahalanobis, definida por $d_{ij} = (\mathbf{u} - \mathbf{v})' \mathbf{C}^{-1} (\mathbf{u} - \mathbf{v})$, onde \mathbf{u} e \mathbf{v} são os vetores contendo os valores das variáveis usadas para o pareamento e \mathbf{C} é a matriz de variâncias e covariâncias amostral³, incorporando adicionalmente a estimativa do logito do escore de propensão $\hat{q}(\mathbf{x})$ no cálculo da distância. Quando as covariáveis têm distribuição normal multinormal, e os grupos intervenção e controle têm a mesma matriz de covariâncias, então o pareamento pela métrica de Mahalanobis é uma técnica de redução de viés com percentual igual em todas as

covariáveis (*Equal Per cent Bias Reducing - EPBR*), onde o viés é a média do grupo intervenção menos a média do grupo controle.

Pareamento usando a métrica de Mahalanobis mais próxima disponível com calibração interna usando escores de propensão (*Nearest available Mahalanobis metric matching within calipers defined by propensity score*).

Esse método é a combinação dos dois métodos descritos anteriormente. Os sujeitos do grupo intervenção são aleatoriamente ordenados e o primeiro sujeito deste grupo é selecionado. Todos os sujeitos do grupo controle com valor de escore de propensão já estimado são selecionados e a distância de Mahalanobis é calculada entre estes indivíduos e os do grupo intervenção. O sujeito controle mais próximo e o seu par (sujeito tratado) são excluídos e o processo é repetido. O valor de calibração é determinado pelo investigador e será usado para definir quantos pares do grupo controle cada indivíduo do grupo intervenção terá³.

Entre os três métodos de pareamento descritos, o terceiro método é o melhor, mas os autores sugerem utilizar o primeiro, pois é uma técnica mais fácil para implementar computacionalmente.³

A estratificação, também chamada de subclassificação, é muito usada em estudos observacionais para controlar diferenças sistemáticas entre os grupos tratados e controles. O método consiste em agrupar em determinados estratos (grupos) baseado em características anteriores observadas. Uma vez que os estratos forem definidos, os indivíduos tratados e os controles que estiverem no mesmo grupo, podem ser comparados diretamente.

Muitos dos problemas da estratificação acontecem quando o número de covariáveis aumenta. O aumento do número de preditores leva ao crescimento exponencial do número de estratos. Como o escore de propensão já é um resumo de dos preditores, a estratificação pode balancear suas distribuições nos grupos sem o crescimento exponencial do número de estratos.

Ao contrário da estratificação e do pareamento, a ponderação não é tão utilizada como método de ajuste. Na ponderação por escores de propensão, os grupos intervenção e controle são ponderados para que sejam mais representativos da população. A ponderação para os sujeitos do grupo intervenção são definidos como o inverso do escore de propensão $\left(w_i = \frac{1}{\hat{e}(\mathbf{x})}\right)$, e para os sujeitos do grupo controle são definidos como o inverso do complementar do escore de propensão $\left(w_i = \frac{1}{1 - \hat{e}(\mathbf{x})}\right)$.

Uma limitação importante da ponderação é que se as estimativas dos escores de propensão estiverem próximas de zero ou de um, então produz pesos não realistas

2.4 Recursos computacionais

A estimação de escores de propensão usando regressão logística pode ser realizada com diversos programas de análise de dados, tais como SAS, Stata, R, IBM-SPSS. No entanto, para a utilização dos métodos de pareamento por escores de propensão são necessárias rotinas especiais as quais muitas vezes não estão disponíveis, mas podem ser implementadas.

Nas seções a seguir serão apresentadas duas formas de estimar os escores de propensão e para fazer o pareamento, a primeira usando o programa R, e depois utilizando o programa SAS.

2.4.1 R

Escores de propensão podem ser usados para fazer o pareamento no software livre R, utilizando a função `ps` disponível no pacote `twang`. O quadro abaixo mostra uma instrução genérica em linguagem R para a estimação dos escores de propensão e o pareamento.

Detalhes podem ser obtidos na página <http://rss.acs.unt.edu/Rdoc/library/twang/html/ps.html> (acessado em 19/11/2010).

```
Install.packages("twang",dependencies = T  
Lybrary (twang)  
ps(formula = formula(data),  
data,  
title=NULL,  
stop.method = stop.methods[1:2],  
plots="all",  
pdf.plots=FALSE,  
n.trees = 10000,  
interaction.depth = 3,  
shrinkage = 0.01,  
perm.test.iters=0,  
print.level = 2,  
verbose = TRUE)
```

As primeiras duas linhas são para instalar (quando será necessário escolher um repositório para fazer o download do pacote) e carregar o pacote *twang*, respectivamente. Os argumentos da instrução do quadro acima são descrito a seguir:

formula: equação do modelo que estima os escores de propensão, sendo que à esquerda deve constar a variável que identifica os grupos (tratamento ou intervenção, por exemplo) e à direita os preditores do modelo;

data: nome do banco de dados;

title: um texto curto como título que será usado nos gráficos e arquivos que serão salvos;

stop.method: um método, ou uma lista de métodos, com regras e métricas de avaliação da qualidade do escore de propensão;

plots: um vetor indicando quais gráficos devem ser criados. As opções são “todos”, apenas alguns dos gráficos como, por exemplo, boxplot, otimização dos escores de propensão, histograma de peso, valores p dos testes t e ks, ou ainda nenhum gráfico;

pdf.plots: Se “TRUE” então todos os gráficos são salvos em um arquivo .PDF com o título especificado no argumento “title”;

n.trees: número de interações gbm (*Generalized Boosted Regression Modeling*);

interaction.deph: Especificação do modelo gbm, onde “1” indica um modelo aditivo, “2” indica um modelo com interações até ordem dois, e assim por diante;

shrinkage: parâmetro para gbm;

perm.test.iters: número inteiro não negativo indicando o número de interações do teste de permutação para a estatística ks;

verbose: Se "TRUE" então toda informação será listada no monitor durante o processo de ajustamento.

O Anexo 1 apresenta um exemplo de sintaxe R para o cálculo do escore e para o pareamento utilizando os escores de propensão. Os resultados são organizados em diferentes arquivos de dados R (data frames), sendo que os mais importantes são brevemente descritos abaixo:

- **ps**: contém as estimativas dos escores de propensão, onde cada coluna representa as estimativas correspondentes às opções especificadas no argumento *stop.methods*;
- **w**: com ponderações dos escores de propensão (e se foram definidas ponderações, então já estão incorporadas);
- **plot.info**: contém as informações para gerar os gráficos; e,
- **desc**: tabela contendo estatísticas descritivas.

A visualização dos resultados pode ser facilitada usando o pacote PSAgraphics⁹, cujas funções mais importantes são *cat.psa*, *box.psa*, *circ.psa* e *loess.psa*, entre outras. Por exemplo, o quadro abaixo mostra a sintaxe para produzir um diagrama de caixas (boxplot) dos escores de propensão por grupo.

```
box.psa(continuous, treatment = NULL, strata = NULL,
boxwex = 0.17, offset = 0.17, col = c("yellow",
"orange", "black", "red", "darkorange3"),
xlab = "Stratum", legend.xy = NULL,
legend.labels = NULL, pts = TRUE)
```

2.4.2 SAS

No programa SAS, o pareamento utilizando escores de propensão é feito em duas etapas, primeiro estimando o escore de propensão, e em seguida fazendo o pareamento. A

estimação dos escores pelo modelo logístico pode ser realizada utilizando os procedimentos PROC LOGISTIC ou PROC GENMOD.

Para fazer o pareamento pode-se utilizar a macro chamada “%Match” disponibilizada por L. Parsons¹⁰, constante no Anexo 2. A sintaxe para chamar a macro é mostrada no quadro abaixo:

```
%MACRO MATCH(  
  Lib,          /*nome da libname*/  
  dataset,     /*arquivo de dados com todas observações*/  
  Matched,     /*arquivo de dados com observações pareadas*/  
  SCase,       /*arquivo de dados só com casos*/  
  Scontrol,    /*arquivo de dados só com controles*/  
  depend,     /*variavel dependente */  
  digits       /*numero de digitos para parear com os escores*/  
);
```

O pareamento pode ser realizado com diferentes níveis de exigência na similaridade dos valores dos escores entre os grupos, conforme especificado no argumento “*digits*”. Por exemplo, se o valor especificado é igual a 0,1, então os pares de participantes são criados utilizando indivíduos com mesmo valor na primeira casa decimal do escore de propensão. Para tornar o pareamento mais exato, pode-se escolher participantes com mesmo valor na segunda (*digits=0,01*) ou terceira (*digits=0,001*) casa decimal e assim por diante. Naturalmente, quanto maior o nível de exigência na formação dos pares, menos indivíduos podem ser incluídos na análise (ou seja, menor vai ser o número de pares encontrados). Os indivíduos não pareados são descartados.

Esta macro faz pareamento 1:1, ou seja, um controle para cada caso. É possível (e sensato) utilizar mais indivíduos do grupo controle para cada caso (pareamento 1:N), mas isso depende do número de indivíduos disponível do grupo controle. A macro “%OneToManyMATCH” foi implementada por L. Parsons¹¹ permite fazer este tipo de pareamento.

O próximo capítulo descreve aspectos básicos da aplicação de escores de propensão utilizando um conjunto de dados hipotético.

3 Aplicação

Para ilustrar aspectos da utilização do método de escores de propensão, um conjunto de dados foi gerado artificialmente. A função RANDNORMAL do programa SAS foi utilizada para gerar uma amostra com 5000 observações e oito variáveis de uma distribuição normal multivariada, com matriz de variâncias e covariâncias pré-especificada. A documentação da referida função pode ser encontrada na página http://support.sas.com/documentation/cdl/en/imlug/59656/HTML/default/modlib_sect21.htm (acessada em 23/11/2010). A seguir, duas dessas variáveis foram dicotomizadas, uma para representar o desfecho e outra para representar os grupos. Transformações de locação e de escala foram realizadas para criar diferenças nas distribuições dos preditores em relação aos grupos e ao desfecho. Adicionalmente, outra variável foi dicotomizada para representar o sexo dos indivíduos.

A Tabela 1 descreve as variáveis da amostra hipotética e resume suas distribuições por grupo. As comparações entre grupos foram realizadas utilizando o teste de Wilcoxon para amostras independentes para os preditores quantitativos. Para variáveis dicotômicas foi utilizado o teste igualdade de proporções baseado na estatística qui-quadrado de Pearson com correção de Yates. A normalidade dos dados foi avaliada mediante o teste de Anderson-Darling, tendo sido rejeitada a hipótese nula que especifica que a distribuição é normal ($P < 0,005$) para todos os preditores.

Os modelos de regressão logística e de regressão de Poisson robusta foram utilizados para ilustrar a aplicação do método de escores de propensão. É importante salientar que a escolha do modelo depende da medida de associação que se deve estimar. Se o desfecho representa dados de incidência ou de prevalência, o modelo de regressão de Poisson robusta (ou logbinomial, se este convergir) deveria ser utilizado para estimar risco relativo ou razão de prevalências, respectivamente. Esta recomendação é mais importante quando o desfecho é comum ($>10\%$), como ocorre neste exemplo. O Anexo 2 contém a rotina computacional do programa SAS utilizada para o ajustes dos modelos explorados neste capítulo.

Tabela 1 – Médias (DP) ou % das variáveis da amostra hipotética, por grupo.

Variável	Controle (N=2498) Média (DP)	Intervenção (N=2502) Média (DP)	P
Desfecho**	30,0	52,8	<0,0001*
Sexo			
Masculino**	59,6	40,4	<0,0001*
Altura (cm)	170,7 (7,35)	172,3 (8,85)	<0,0001†
Cintura (cm)	58,2 (6,29)	71,7 (27,17)	<0,0001†
Idade (anos)	63,0 (7,12)	71,9 (7,33)	<0,0001†
IMC (kg/m ²)	22,8 (3,48)	28,5 (4,03)	<0,0001†
Peso (kg)	66,0 (6,89)	83,9 (7,22)	<0,0001†

* Teste igualdade de proporções baseado na estatística qui-quadrado de Pearson com correção de Yates

† Teste de Wilcoxon para amostras independentes

DP – Desvio padrão

** N (%)

A Tabela 2 apresenta os resultados da análise utilizando o modelo de regressão logística para estimação da razão de chances da ocorrência do desfecho sem a utilização de escores de propensão. Como foi destacado no parágrafo acima, o primeiro aspecto que deve ser observado são as enormes diferenças, em magnitude, entre as correspondentes estimativas de razão de chances e risco relativo. Comparando a razão de chances para o desfecho estimada do modelo univariável não ajustado (RC=2,6; IC 95%: 2,32-2,92) com a razão de chance ajustada para cintura, idade, IMC e sexo, é perceptível a presença de forte confundimento na associação entre grupo e o desfecho (RC=0,11; IC 95%:0,09-0,14), provavelmente devido às diferenças nas distribuições das variáveis nos grupos controle e intervenção. O mesmo comportamento foi observado para as estimativas de risco relativo (RR=0,62; IC 95%:0,57-0,68).

Tabela 2 – Razões de chances e riscos relativos da ocorrência do desfecho estimados pelos modelos de regressão logística e regressão de Poisson robusta univariável e multivariável.

Preditor	RC* (IC 95%)		RR** (IC 95%)	
	Univariável	Multivariável [†]	Univariável	Multivariável [†]
Grupo	2,60 (2,32-2,92)	0,11 (0,09-0,14)	1,76 (1,64-1,89)	0,62 (0,57-0,68)
Altura (cm)	0,95 (0,94-0,96)	-	0,97 (0,97-0,97)	-
Cintura (cm)	1,25 (1,23-1,26)	1,23 (1,21-1,24)	1,01 (1,01-1,01)	1,00 (1,00-1,00)
Idade (anos)	1,13 (1,12-1,14)	1,10 (1,09-1,12)	1,06 (1,05-1,06)	1,05 (1,04-1,06)
IMC (kg/m ²)	1,30 (1,27-1,32)	1,27 (1,24-1,31)	1,11 (1,10-1,12)	1,11 (1,10-1,11)
Peso (kg)	1,10 (1,09-1,10)	-	1,05 (1,04-1,05)	-
Sexo				
Masculino	1	1	1	1
Feminino	0,67 (0,60-0,75)	1,24 (1,05-1,46)	0,79 (0,74-0,84)	1,08 (1,02-1,15)

* RC – Razão de chances estimada pelo modelo de regressão logística

** RR – Risco relativo estimado usando regressão de Poisson robusta

[†] No modelo multivariável foi considerada a variável IMC ao invés de peso e altura

Os escores de propensão podem ser usados para estimar a associação ajustada entre desfecho e grupo. Assim, no modelo de regressão logística ajustado pelo escore de propensão, a estimativa desta associação foi RC=0,05 (IC 95%:0,039-0,068). Nesse modelo, são incluídas apenas as variáveis que identifica o grupo e o escore de propensão. Similarmente, utilizando o modelo de regressão de Poisson robusta a estimativa de risco relativo ajustada pelo escore de propensão foi RR=0,42 (IC 95%: 0,40-0,45). Novamente foi observada uma grande diferença em magnitude entre as estimativas de razão de chances e risco relativo.

A aplicação do uso de pareamento pelo escore de propensão será ilustrada a seguir, utilizando a macro “%Match” descrita no Anexo 2. Três níveis de exigência (exatidão) foram usados no pareamento, especificando os valores digits=0,1 (nível baixo), digits=0,01 (nível médio) e digits=0,001 (nível alto). Cabe ressaltar que esta denominação serve meramente para diferenciar as três situações.

A Figura 1 resume as diferenças entre as distribuições dos escores de propensão estimados, separados por grupo (0=Controles, 1=Intervenção). Pode-se observar que a grande maioria das estimativas dos escores de propensão dos indivíduos do grupo intervenção está concentrada acima de 0,6, enquanto que para os indivíduos do grupo controle estão concentradas abaixo de 0,4. Este aspecto pode explicar por que o número de pares criados para o pareamento foi pequeno.

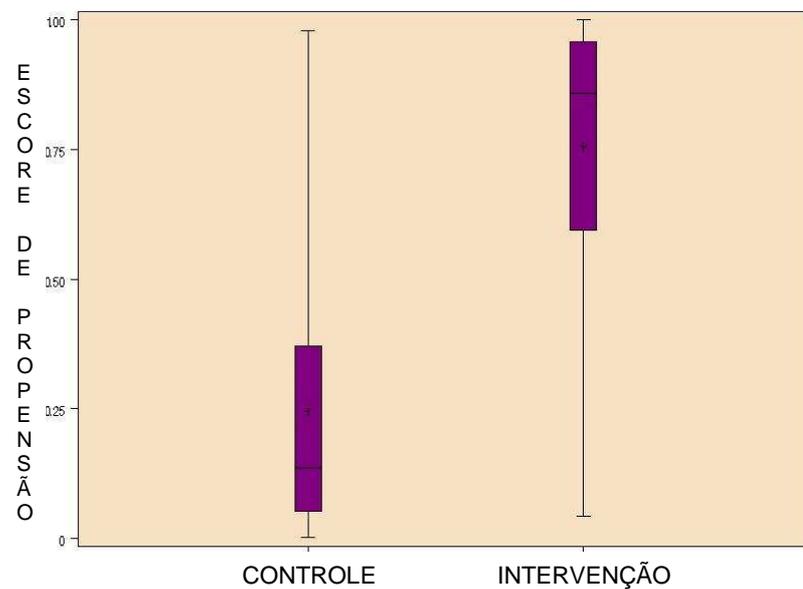


Figura 1 – Diagrama de caixas das estimativas do escore de propensão, por grupo.

A Tabela 3 descreve as distribuições dos preditores por grupo para estas três possibilidades de pareamento. É importante observar que aumentando o nível de exigência na similaridade dos escores de propensão entre os grupos para o pareamento, o número de pares diminuiu.

Tabela 3 – Médias (DP) ou % dos preditores por grupo com diferentes níveis de exatidão dos escores de propensão para o pareamento.

Variáveis	Nível Baixo (<i>digits=0,1</i>)			Nível Médio (<i>digits=0,01</i>)			Nível Alto (<i>digits=0,001</i>)		
	Controle (N = 884)	Intervenção (N = 884)	P	Controle (N = 854)	Intervenção (N = 854)	P	Controle (N = 676)	Intervenção (N = 676)	P
Sexo									
Masculino	50	50	0,9618	50,5	49,5	0,7699	49,6	50,4	0,8266
Feminino	50,1	49,9		49,6	50,4		50,3	49,7	
Cintura (cm)	62,2 (6,25)	65,1 (20,32)	0,2167	62,0 (6,30)	62,0 (5,77)	0,1460	62,1 (6,39)	61,9 (5,74)	0,0741
Idade (anos)	67,5 (6,47)	67,7 (6,60)	0,7999	67,3 (6,52)	67,3 (6,33)	0,3736	67,4 (6,66)	67,6 (6,66)	0,9133
IMC (kg/m ²)	25,8 (3,11)	26,2 (3,61)	0,1819	25,8 (3,21)	25,8 (2,79)	0,5886	25,8 (3,37)	25,8 (2,79)	0,8089

As razões de chances brutas estimadas por regressão logística condicional não ajustadas utilizando estes pareamentos foram $RC=0,07$ (IC 95%:0,04-0,10), $RC=0,08$ (IC 95%: 0,05-0,11) e $RC=0,06$ (IC 95%:0,04-0,10), respectivamente para os níveis de pareamento baixo, médio e alto. No entanto, para dados de incidência ou prevalência não é recomendável utilizar regressão logística condicional, pois também superestima o risco relativo ou razão de prevalências. Apesar de não parecer muito vantajoso, devido à grande redução no tamanho de amostra, uma alternativa seria utilizar o modelo de regressão de Poisson robusta para dados pareados.

4 Considerações finais

Em estudos observacionais podem existir diferenças importantes nas distribuições das variáveis preditoras entre os grupos, podendo gerar uma estimativa viesada do efeito (ou associação) da intervenção com o desfecho. O uso de escores de propensão pode ser extremamente útil para equilibrar as distribuições das variáveis preditoras entre os grupos (isto é, tornar os grupos comparáveis) e, conseqüentemente, obter uma estimativa não viesada do impacto da intervenção sobre o desfecho.

O pareamento dos indivíduos utilizando os valores estimados do escore de propensão é a técnica de ajuste mais utilizada e foi descrita neste trabalho. A rotina computacional “*%Match*” escrita na linguagem SAS foi aplicada aos dados hipotéticos para descrever aspectos práticos desta técnica, ilustrando empiricamente a grande impacto na redução de viés para estimar a associação do preditor principal (grupo) com o desfecho dicotômico.

O exemplo ilustrou com muita clareza o papel do confundimento nos resultados. Na análise univariável a associação entre o desfecho e o fator (grupo) foi $RC=2,60$ (IC 95%: 2,32-2,92), contrastando com o resultado do modelo logístico multivariável $RC=0,11$ (IC 95%:0,09-0,14). O mesmo comportamento foi observado com o modelo de regressão de Poisson robusta, mediante o qual foram obtidas as estimativas de risco relativo $RR=1,76$ (IC 95%:1,64-1,89) e $RR=0,62$ (IC 95%:0,57-0,68), respectivamente para o modelo ajustado e não ajustado. Fica evidente que as diferenças nas distribuições nos preditores conduzem a uma estimativa viesada da magnitude e da direção da associação. No entanto, é possível que ainda exista confundimento residual, pois as estimativas de razão de chances e risco relativo ajustadas pelo escore de propensão foram, respectivamente, $RC=0,05$ (IC 95%:0,04-0,07) e $RR=0,42$ (IC 95%: 0,40-0,45).

O investigador deve considerar a natureza do delineamento epidemiológico que gera os dados para escolher o modelo adequado para estimar a medida de associação. O exemplo mostra empiricamente enormes diferenças entre as estimativas de razão de chances e risco relativo, para todos os modelos ajustados.

Ao invés de utilizar o modelo ajustado pelo escore de propensão, uma alternativa é fazer o pareamento. Utilizando o pareamento 1:1 e com nível baixo de similaridade dos escores, a estimativa da associação foi $RC=0,07$ (IC 95%:0,04-0,10), que não é muito diferente da estimativa anterior. Isto pode sugerir que para este caso, não parece haver um ganho substancial, talvez por que o número de pares ($N=884$) foi relativamente pequeno, considerando o tamanho inicial da amostra. Adicionalmente, aumentando a exigência de similaridade dos escores de propensão para a criação dos pares de indivíduos, as estimativas da associação foram $RC=0,08$ (IC 95%:0,05-0,11) e $RC=0,06$ (IC 95%:0,04-0,10), respectivamente para os níveis médio e alto. É importante notar que o número de pares diminuiu, e aparentemente não há um ganho com este aumento da exigência no pareamento.

Como visto empiricamente, uma desvantagem do método é o grau de exigência para o pareamento com os escores de propensão. O ideal é que os escores de propensão de cada par de indivíduos (intervenção e controle) seja o mais parecido possível, pois representa igual probabilidade de receber a intervenção. Se o pesquisador considera suficiente que escores de propensão sejam parecidos na primeira casa decimal, então o número de pares criados será maior do que exigir pares com escores de propensão com valor igual até a segunda ou terceira casa decimal. Assim, quanto maior o nível de similaridade exigido, menor o número de pares considerados na análise, sendo que os demais indivíduos são descartados. O uso do pareamento 1:N pode ser uma alternativa para aumentar o tamanho de amostra, mas exige que um grande número de indivíduos no grupo que não recebe a intervenção. Isto pode ser factível em estudos observacionais nos quais não há alocação ao acaso aos grupos.

Os escores de propensão apresentam um grande potencial para redução do viés de confundimento, onde parecem ser capazes de equilibrar as distribuições dos preditores e tornar os grupos intervenção e controle comparáveis, fazendo com que as diferenças encontradas sejam provavelmente devido à intervenção. A escolha de como utilizar os escores de propensão é do pesquisador, mas neste trabalho foi mostrado que os resultados são semelhantes entre o uso para fazer pareamento ou como variável de ajuste. Contudo, é visível a vantagem do ajuste multivariável pelo escore de propensão, particularmente quando se

deseja estimar risco relativo ou razão de prevalência, pois o modelo de regressão de Poisson robusta para dados não pareados pode ser utilizado. Ao contrário, utilizando o pareamento dos indivíduos pelos valores do escore de propensão, além de reduzir substancialmente o tamanho de amostra, será necessário utilizar o modelo de regressão de Poisson para dados pareados, haja vista que a regressão logística condicional superestima o risco relativo ou a razão de prevalências.

Em muitas situações práticas, pode ser importante fazer pareamento 1:N, ou seja, mais de uma observação do grupo controle para uma observação do grupo intervenção. No exemplo deste trabalho, o pareamento 1:N não é adequado, haja vista que existem poucos indivíduos do grupo controle em relação ao grupo intervenção, pois os dois grupos têm aproximadamente o mesmo número de indivíduos. Mesmo assim, o uso da macro ***“%OneToManyMATCH”*** apresentada por L.S. Parsons¹¹, não funcionou adequadamente, sendo necessária uma investigação mais aprofundada.

Com o software R, não foi encontrada uma função que fizesse o pareamento 1:N, nem uma possibilidade de adaptar a função ps apresentada para que essa fizesse o pareamento 1:N.

5 Referências

1. Rubin D. B. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Available at: <http://www.ingentaconnect.com/content/klu/hsor/2001/00000002/F0020003/05088585> [Acessado Novembro 9, 2010].
2. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983;70(1):41-55.
3. D'Agostino RB. Tutorial in Biostatistics - Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265-2281.
4. Wang J, Donnan PT. Propensity score methods in drug safety studies: practice, strengths and limitations. *Pharmacoepidem. Drug Safe*. 2001;10(4):341-344.
5. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology*. 2006;59(5):437.e1-437.e24.
6. Austin PC. Primer on Statistical Interpretation or Methods Report Card on Propensity-Score Matching in the Cardiology Literature From 2004 to 2006. *Circulation: Cardiovascular Quality and Outcomes*. 2008;1(1):62 -67.
7. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med Care*. 2007;45(10 Supl 2):S143-148.
8. Pasta DJ. Using Propensity Score to Adjust for Group Differences, SUGI 25.
9. James E. Helmreich, Robert M. Pruzek. PSAgraphics: An R Package to Support Propensity Score Analysis. *Journal of Statistical Software*.
10. Using SAS Software to Perform a Case-Control Match on Propensity Score. Available at: <http://support.sas.com/dsearch?Find=Search&ct=&qt=Using+SAS%AE+Software+to+Perform+a+Case-Control+Match+on+Propensity+Score&col=suppprd&nh=10&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0> [Acessado Novembro 4, 2010].
11. Parsons LS. Performing a 1:N Case-Control Match on Propensity Score, SUGI 29.

6 Anexos

Anexo 1 – Rotina computacional em R com a função ps do pacote twang

Anexo 2 – Rotina computacional em SAS com a macro MATCH

Anexo 1 – Rotina computacional em R com a função ps do pacote twang

```
library(twang)
attach(EP) # "fixa" qual o banco
ps.EP <- ps(GRUPO ~ IMC + IDADE + CINTURA + SEXO,
            data = EP,
            title="EP - banco simulado",
            stop.method=stop.methods[c("ks.stat.mean","ks.stat.max"]),
            # generate plots?
            plots="all",
            pdf.plots=TRUE,
            # gbm options
            n.trees=2000,
            interaction.depth=3,
            shrinkage=0.005,
            perm.test.iters=0,
            verbose=FALSE)

names(ps.EP) # resultados da função ps
# 2: treat
ps.EP[[2]] # banco inteiro: sujeito tratado = 1; não tratado = 0
# 3: treat.var
ps.EP[[3]] # variável "tratamento", no meu caso: GRUPO
# 4: desc
ps.EP[[4]] # medidas para cada método: ess, n.treat, n.ctrl, max.es, mean.es, max.ks, max.ks.p, mean.ks, bal.table
# ess: tamanho efetivo da amostra do grupo controle (talvez um missing...)
# n.treat: número de sujeitos do grupo controle
# n.ctrl: número de sujeito do grupo intervenção
# max.es: maior magnitude de efeito entre covariáveis
# mean.es: a média absoluta do tamanho do efeito
# max.ks: a maior estatística KS entre as covariáveis
# mean.ks: a média da estatística KS entre covariáveis
# bal.table: uma tabela resumida da qualidade das ponderações para "equilibrar" as distribuições entre os grupos

# 5: ps
ps.EP[[5]] # ver escores calculados por cada método de avaliação dos escores de propensão calculados
ps.EP[[5]][1] # ver escores calculados pelo método 1: ks.stat.mean
ps.EP[[5]][2] # ver escores calculados pelo método 2: ks.stat.max

# 6: w
ps.EP[[6]] # ver ponderações
ps.EP[[6]][1] # ver ponderação do método 1
```

```

ps.EP[[6]][1] # ver ponderação do método 2
mean(ps.EP[[6]]) #calcula a média dos pesos para cada método

# 7: plot.info
ps.EP[[7]] # lista contendo dados em linha usados para gerar gráficos
ps.EP[[7]][1] # dados do primeiro método. Com não tem gráfico, não tem dados. APARECE NULL
ps.EP[[7]][2] # lista contendo dados para o gráfico pelo método ks.stat.mean
ps.EP[[7]][3] # lista contendo dados para o gráfico pelo método ks.stat.max

# 8: datestamps
ps.EP[[8]] #grava data da análise

# 9: parameters
ps.EP[[9]] #salva o comando da função ps

# 10: alerts
ps.EP[[10]] #texto contendo qualquer warning acumulado durante a estimação

# get the balance tables
bal.table(ps.EP)
#o que tem dentro da bal.table? (PARA CADA VARIÁVEL)
#tx.mn: a média do grupo intervenção
#tx.sd: o desvio padrão do grupo intervenção
#ct.mn: a média do grupo controle
#ct.sd: o desvio padrão do grupo controle
#std.eff.sz: o tamanho do efeito padronizado Se dp do grupo intervenção for pequeno ou 0, então std.eff.sz pode ser grande ou
INF
#stat: teste t para variáveis numéricas e teste quiquadrado para variáveis contínuas
#p: p valor associado ao 'stat'
#ks: estatística ks
#ks.pval: p valor da estatística ks usando aproximação analítica

a <- dx.wts(ps.EP,data=EP,treat.var="GRUPO")
print(a) # tabela resumo que compara métodos
bal.table(a) # para cada variável, medidas da bal.table dividido por métodos

w <- with(ps.EP, ps/(1-ps))
w[EP$GRUPO==1,] <- 1 #define em 1 pra quem tem ponderação maior que 1
dx.wts(w,data=EP,treat.var="GRUPO", perm.test.iters=0)

p <- ps.EP$ps #escores de propensão como ponderação
dx.wts(p,data=EP,treat.var="GRUPO",x.as.weights=FALSE)

```

```

# olhando para os escores de propensão
names(ps.EP$ps)
hist(ps.EP$ps$ks.stat.max)
hist(ps.EP$ps$ks.stat.mean)
boxplot(split(ps.EP$ps$ks.stat.max, EP$GRUPO),
        main="Método ks.stat.max",
        ylab="Escores de propensão estimados",
        names=c("controle", "intervenção"))

#box plot dos escores de propensão separados por grupo
boxplot(split(ps.EP$ps$ks.stat.mean, EP$GRUPO),
        main="Método ks.stat.mean",
        ylab="Escores de propensão estimados",
        names=c("controle", "intervenção"))

# checando 'equilíbrio'
names(ps.EP$desc)
# sem ponderação
ps.EP$desc$unw
# otimização pelo método ks.stat.max
ps.EP$desc$ks.stat.max
# otimização depo método ks.stat.mean
ps.EP$desc$ks.stat.mean

# estimating the propensity score
summary(ps.EP$gbm.obj, n.trees=ps.EP$desc$ks.stat.max$n.trees)

bal.stat(data=EP, # compara os sujeitos dos grupos intervenção e controle pelo média, dp, tamanho do efeito, estatística KS
        w.all=w[,1],
        vars=names(EP),
        treat.var="GRUPO",
        get.means=TRUE, #computa médias e variâncias
        get.ks=TRUE, #computa estatística ks
        na.action="level") # como manipular missings. opções: "level", "exclude", "lowest"

```

Anexo 2 – Rotina computacional em SAS com a macro MATCH

```
title; title1;title2;

options ps=58 ls=80 nocenter nodate nonumber formchar='|----|+|----+=|-/\<>*' ;
title;
title1;
title2;

libname L1 V9 "C:\Vigo\Maria Claudia";
options ls=130;

data MC;
    set L1.dadosmc_121110;
run;
proc format;
    value grupof 0='Controle' 1='Caso';
run;

* Tabela 2 - RC (Univariaveis);
proc logistic data=MC descending;
    model DESFECHO = GRUPO / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = ALTURA / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = CINTURA / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = IDADE / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = IMC / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = PESO / rl;
run;
proc logistic data=MC descending;
    model DESFECHO = SEXO / rl;
run;

* Tabela 2 - RC (Multivariavel);
proc logistic data=MC descending;
    model DESFECHO = GRUPO CINTURA IDADE IMC SEXO / rl;
run;

* Tabela 2 - RP Robusta;
proc genmod data=MC ;
    titlel "RR Univariavel";
    class ID;
    model DESFECHO = GRUPO / dist=poisson link=log lrci;
    repeated subject=ID / type=ind;
    estimate 'Alto vs. Baixo' GRUPO 1 / exp;
run;

proc genmod data=MC ;
    titlel "RR Univariavel";
    class ID;
    model DESFECHO = ALTURA / dist=poisson link=log lrci;
    repeated subject=ID / type=ind;
    estimate 'Aumento lcm' ALTURA 1 / exp;
run;
```

```

proc genmod data=MC ;
  title1 "RR Univariavel";
  class ID;
  model DESFECHO = CINTURA / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Aumento lcm' CINTURA 1 / exp;
run;
proc genmod data=MC ;
  title1 "RR Univariavel";
  class ID;
  model DESFECHO = IDADE / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Aumento lano' IDADE 1 / exp;
run;
proc genmod data=MC ;
  title1 "RR Univariavel";
  class ID;
  model DESFECHO = IMC / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Aumento lkg/m2' IMC 1 / exp;
run;
proc genmod data=MC ;
  title1 "RR Univariavel";
  class ID;
  model DESFECHO = PESO / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Aumento lkg' PESO 1 / exp;
run;
proc genmod data=MC ;
  title1 "RR Univariavel";
  class ID;
  model DESFECHO = SEXO / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Feminino vs. Masculino' SEXO 1 / exp;
run;

* RP robusta multivariavel;
proc genmod data=MC ;
  title1 "RR Multivariavel";
  class ID;
  model DESFECHO = GRUPO CINTURA IDADE IMC SEXO / dist=poisson link=log
lrci;
  repeated subject=ID / type=ind;
  estimate 'Alto vs. Baixo' GRUPO 1 / exp;
  estimate 'Aumento lcm' CINTURA 1 / exp;
  estimate 'Aumento lano' IDADE 1 / exp;
  estimate 'Aumento lkg/m2' IMC 1 / exp;
  estimate 'Feminino vs. Masculino' SEXO 1 / exp;
run;

*****
      Criando ESCORES DE PROPENSAO
*****;
proc logistic data=MC descending;
  model GRUPO = IDADE SEXO IMC CINTURA / rl ;
  output out=ESCORE prob=prob;
run;

```

```

* Modelo RL ajustado pelo PS;
proc logistic data=ESCORE descending;
  titlel "RC Ajustada pelo PS";
  model DESFECHO = GRUPO PROB / rl;
run;

* Modelo RP Robusta ajustado pelo PS;
proc genmod data=ESCORE ;
  titlel "RR Ajustado pelo PS";
  class ID;
  model DESFECHO = GRUPO PROB / dist=poisson link=log lrci;
  repeated subject=ID / type=ind;
  estimate 'Alto vs. Baixo' GRUPO 1 / exp;
run;

*****
                          PAREAMENTO
*****;
%MACRO MATCH (
  Lib,          /*nome da libname*/
  dataset,     /*data set com todas observações*/
  Matched,     /*ordenado observações pareadas*/
  SCase,       /*ordenado pelos casos*/
  Scontrol,    /*ordenado pelos controles*/
  depend,      /*variavel dependente, no caso: GRUPO*/
  digits       /*numero de digitos para parear*/
);
data tcases (drop=cprob)
  tctrl (drop=aprob);
set &LIB.&dataset.;
if &depend. = 1 and prob ne .
then do;
  cprob = round(prob,&digits.);
  Cmatch = 0;
  Length RandNum 8;
  RandNum=ranuni(1234567);
  Label RandNum = 'Uniform Randomization Score';
  output tctrl;
end;
else if &depend. = 0 and prob ne .
then do;
  aprob = round(prob,&digits.);
  output tcases;
end;

run;
proc sort data=tcases
  out=&LIB.&SCase.;
  by prob;
run;
proc sort data=tctrl
  out=&LIB.&Scontrol.;
  by prob randnum;
run;
data &lib.&matched.
  (drop=Cmatch randnum aprob cprob start oldi curctrl matched);
  set &lib.&SCase. ;
  curob + 1;
  matchto = curob;
  if curob = 1 then do;
    start = 1;
    oldi = 1;
  end;

```

```

do i = start to n;
    set &LIB.&Scontrol. point = i nobs = n;

    if i gt n then goto startovr;
    if _Error_ = 1 then abort;

    curctrl = i;
    if aprob = cprob
then do;
        Cmatch = 1;
        output &LIB.&matched.;
        matched = curctrl;
        goto found;
    end;
    else if cprob gt aprob then goto nextcase;

    startovr: if i gt n then goto nextcase;
end; /*end of DO LOOP*/
nextcase:
if Cmatch=0 then start=oldi;

found:
if Cmatch = 1
then do;
    oldi = matched + 1;
    start = matched + 1;
    set &LIB.&SCase. point = curob;
    output &LIB.&matched.;
    end;

    retain oldi start;
    if _Error_=1 then _Error_=0;
run;
%MEND MATCH;

***** EXECUTANDO A MACRO *****;
title 'Pareamento com PS - Caso 0.1';
%MATCH (WORK,SCORE,Match,CASOS,CONTROLES,GRUPO1,.1);

* Medias com pareamento;
proc means data=MATCH n mean std maxdec=2;
class GRUPO;
var IDADE SEXO IMC CINTURA;
format GRUPO grupof.;
run;
proc logistic data=SCORE;
title2 'Regressao Logistica (nao pareada)- Sem Ajuste';
model DESFECHO(event='1')= GRUPO;
format GRUPO grupof.;
run;
proc logistic data=SCORE;
title2 'Regressao Logistica (nao pareada)- Ajustado por PS';
model DESFECHO(event='1')= GRUPO PROB;
format GRUPO grupof.;
run;
proc logistic data=MATCH;
title2 'Regressao Logistica Condicional - Sem Ajuste';
strata MATCHTO;
model DESFECHO(event='1')= GRUPO;
format GRUPO grupof.;
run;

```

```

*****;
title1 'Pareamento com PS - Caso 0.01';
%MATCH (WORK,ESCORE,Match,CASOS,CONTROLES,GRUPO1,.01);

* Medias com pareamento;
proc means data=MATCH n mean std maxdec=2;
  class GRUPO;
  var IDADE SEXO IMC CINTURA;
  format GRUPO grupof.;
run;
proc logistic data=MATCH;
  title2 'Regressao Logistica Condicional - Sem Ajuste';
  strata MATCHTO;
  model DESFECHO(event='1')= GRUPO;
  format GRUPO grupof.;
run;

*****;
title1 'Pareamento com PS - Caso 0.001';
%MATCH (WORK,ESCORE,Match,CASOS,CONTROLES,GRUPO1,.001);

* Medias com pareamento;
proc means data=MATCH n mean std maxdec=2;
  class GRUPO;
  var IDADE SEXO IMC CINTURA;
  format GRUPO grupof.;
run;
proc logistic data=MATCH;
  title2 'Regressao Logistica Condicional - Sem Ajuste';
  strata MATCHTO;
  model DESFECHO(event='1')= GRUPO;
  format GRUPO grupof.;
run;

```