**Primeira Lista de Exercícios – Avaliação de Políticas Sociais**

**Modelo de resultados potenciais e Matching**

**I – Questões teóricas**

**Questão 1 (Q4 2015):** Com base no arcabouço do modelo dos resultados potenciais, a expressão abaixo descreve os dois tipos de vieses que surgem quando se utiliza o estimador *naive* para análise causal de um “tratamento qualquer” em uma situação em que o processo de atribuição do tratamento não foi aleatório.

*E* (Y1|D=1) - *E*(Y0|D=0) = *E*(δ) + { *E*(Y0|D=1) - *E*(Y0|D=0) } +

(1- π)[{*E* (Y1|D=1) - *E*(Y0|D=1)} – {*E*(Y1|D=0) - *E*(Y0|D=0)}]

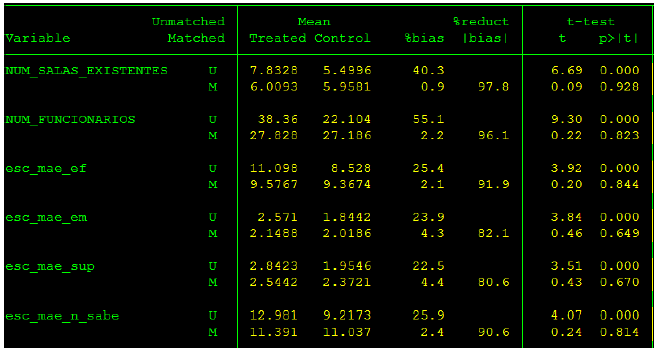
*Onde E*(δ) = ATE = *E* (Y1 – Y0) = *E*(Y1) – *E*(Y0).

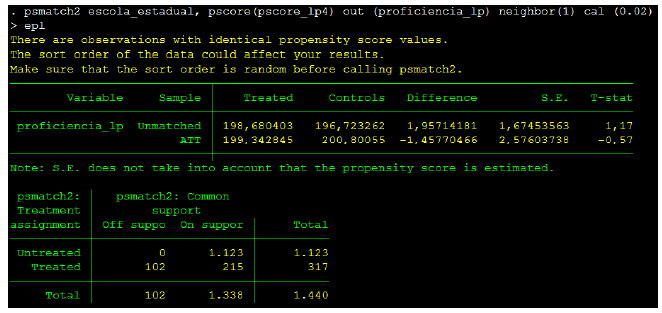
Os diferentes tracejados identificam os dois vieses. Comente cada um deles. Pense em um exemplo (pode ser fictício) diferente dos abordados em aula, para ilustrar tais vieses.

**Questão 2 (Q5 2015):** “O pareamento pode ser uma alternativa quando a seleção ao tratamento não foi aleatória.”

a) Explicite as hipóteses necessárias para estimar o efeito do tratamento sobre os tratados com base em um pareamento, admitindo que **S** seja o conjunto disponível de variáveis observáveis.

b) A figura abaixo mostra a saída de um teste de balanceamento após a realização do pareamento com uso do *propensity-score* bem como os resultados do próprio *propensity score matching*. Comente os resultados.





**Questão 3 (Q1 2016):** Diversos pesquisadores têm apontado os benefícios para as crianças de começar a estudar mais cedo, ou seja, de frequentar a educação infantil. O desempenho escolar futuro é um dos indicadores utilizados para captar tais benefícios. Suponha que você seja contratado para avaliar o impacto da frequência a educação infantil sobre o desempenho escolar. Os dados disponíveis são os resultados de um teste padronizado (tipo SAEB), de um determinado ano (por exemplo, de 2015) para alunos do 5º ano do ensino fundamental. Você também tem disponíveis as características das escolas, professores, alunos e seus familiares. Uma das informações presente é a respeito de quando a criança começou a estudar. No questionário do aluno há a seguinte pergunta:

*Quando você entrou na escola?*

1. *Na creche (0 a 3 anos).*
2. *Na pré-escola (4 a 5 anos).*
3. *Na primeira série ou primeiro ano (6 a 7 anos).*
4. *Depois da primeira série.*

a) formalize o problema de avaliação de impacto descrito na questão tendo por base o modelo de resultados potenciais.

b) discuta os possíveis vieses que surgiriam caso o estimador *naive* fosse utilizado para realizar a avaliação descrita.

c) Qual método de avaliação seria mais apropriado, dadas as informações disponíveis?

d) Quais as hipóteses de identificação para que essa estratégia seja válida?

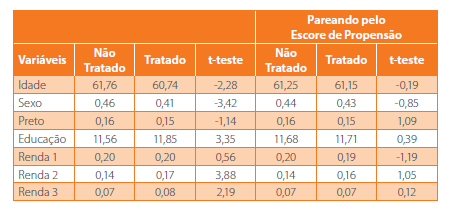
**Questão 4 (Q2 2016):** Compare o matching pelo vizinho mais próximo com reposição e kernel matching.

Na sua resposta você deve descrever brevemente como realizar cada um deles, discutir o trade-off entre viés e variância associado a cada um deles e, por fim, argumentar quando cada um deles é mais apropriado em função da disponibilidade dos dados.

**Questão 5:** Hirano e Imbens (2001) usam os métodos baseados no escore de propensão para estimar o efeito médio de um procedimento cardíaco, o cateterismo, sobre a probabilidade de sobrevivência dos indivíduos. Eles usam observações para 5.735 indivíduos, sendo 2.184 tratados e 3.551 não tratados. Para cada individuo, eles observam se um procedimento de cateterismo foi aplicado no período de 24 horas após a entrada no hospital com problemas cardíacos. Os indivíduos no grupo de tratamento são aqueles que sofreram a intervenção de um cateterismo, enquanto os indivíduos no grupo de controle são aqueles que não sofreram a intervenção. Além disso, eles observam um vetor com 72 características para cada um dos indivíduos, por exemplo, sexo, raça, educação, renda, se possui plano de saúde, tipo de doença que possui, peso, etc. Eles estimam o escore de propensão baseado em um modelo logístico.

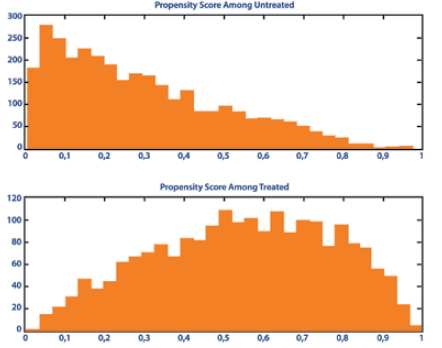
**a)** Antes de estimar o efeito médio do tratamento sobre os tratados, eles apresentam uma tabela que contem um teste de diferença de medias para as características observáveis. Eles

obtém resultados para esse teste antes e depois da ponderação pelo escore de propensão. Esta tabela encontra-se abaixo. Interprete estes resultados:



Nesta tabela, renda 1, renda 2 e renda 3 representam 3 baixas de renda-faixa, media e alta, respectivamente.

**b)** Antes de estimar os efeitos, eles também mostram os histogramas do escore de propensão para o grupo de tratados e para o grupo de não tratados. Interprete estes gráficos.



**c)** Eles usam diferentes métodos para estimar o efeito médio do tratamento sobre a taxa de sobrevivência. Usando um estimador de pareamento simples baseado nos vizinhos mais próximos, eles encontram um efeito de -0,081 (erro-padrão de 0,017). Ao combinar o método de pareamento com regressão simples, o efeito é de -0,063 (erro-padrão de 0,016). Estimando o efeito médio do tratamento por regressão linear, eles encontram um efeito médio de -0,065 (erro-padrão de 0,014), enquanto usando o método de reponderação, eles encontram efeito médio de -0,060 (erro-padrão de 0,018). Combinando os métodos de regressão e reponderação, eles encontram efeito médio de -0,062 (erro-padrão de 0,015). Interprete essas evidencias, dando ênfase as hipóteses em que se baseia cada estimador. Qual seria a conclusão sobre o efeito desse procedimento na taxa de sobrevivência?

**II – Questões Seminário**

\* Caliendo (2014) – “Unobservable, but unimportant?....”

1) Qual o objetivo principal do artigo e qual a estratégia adotada pelo autor?

**III – Questões Prática**

Você quer avaliar o impacto do número de filhos sobre a participação da mulher no mercado de trabalho (V4704). Você tem a sua disposição um instrumento que é a ocorrência de gêmeos.

A variável GEMEOS = 1 indica as mulheres que tiveram gêmeos.

Você também tem a sua disposição um conjunto X de variáveis independentes que caracterizam essas mulheres.

X = idade (V8005), raça (BRANCA), escolaridade (educa), local de residência (metropolitana ou não-metropolitana e macro-região) e o salário de reserva da mulher (renda da mulher que não é derivada do trabalho; por exemplo, o salário do marido). Além disso, porque você trabalha com vários anos da PNAD, um conjunto de variáveis dummies indicam a qual ano tais dados se referem.

Com base em procedimentos de matching, avalie o efeito do ‘numero de filhos’ sobre o Y de interesse.

1. Nearest neighboor – 1 vizinho sem reposição
2. Nearest neighboor –10 vizinhos com caliper = 0.0001 e com reposição
3. Kernel
4. Compare os resultados com uma estimativa por mqo simples.

Apenas para a letra ‘a’ verifique o balanceamento do matching.