

6 Estudo da Simulação

6.1. Introdução

Neste capítulo, serão apresentados os resultados dos estudos de simulação realizados sobre uma base de dados reais retirada da PNAD 2005: o arquivo dos domicílios permanentes particulares alugados da região metropolitana do Rio de Janeiro. A metodologia desenvolvida e apresentada no capítulo 5 é utilizada.

As simulações foram conduzidas para reproduzir a base original de dados, cujas matrizes de covariância e correlação são conhecidas, e podem servir para validar cada procedimento de emparelhamento estatístico proposto neste estudo, na habilidade de preservar a covariância e correlação originais nos arquivos sínteses gerados como se estivessem usando amostragem matricial.

O estudo investiga os resultados dos arquivos sintéticos comparando-os com os resultados do arquivo original.

Em um novo procedimento de emparelhamento estatístico, os métodos propostos usam informação auxiliar, num objetivo micro, irrestrito e com uma abordagem não-paramétrica, usando um procedimento *distance hot deck* realizado em classes.

O primeiro procedimento usa informação auxiliar sobre a relação dos percentis das variáveis Y e Z , sem somar resíduos, e os outros dois procedimentos usam além da informação auxiliar, a adição de resíduos.

Um estudo computacional intensivo realiza tanto as simulações dos métodos propostos como a do método sob pressuposto da CIA. A estrutura de informação auxiliar, se baseia na hipótese do relacionamento teórico de afiliação estocástica entre as variáveis de renda (Y) e de aluguel (Z).

Para cada procedimento de emparelhamento dos arquivos A e B , e uma vez criado o seu respectivo arquivo Síntese, verifica-se se a estrutura de covariância e de correlação das variáveis é preservada após a concatenação. O enfoque principal é investigar a preservação de \sum_{YZ} do arquivo original nos processos, que é uma das indicações da representatividade da base de dados emparelhada dos mesmos. Verifica-se se existem grandes diferenças entre os quatro procedimentos, através do erro quadrático médio e compararam-se os resultados de viés dos dois tipos de emparelhamento, a partir das suas estimativas da covariância e correlação das variáveis Y e Z .

Primeiramente, apresentam-se os resultados de uma análise exploratória realizada nos dados do arquivo original (X, Y, Z) (ver seção 6.1.1) de forma a evitar os zeros estruturais.

A seguir, serão discutidos os procedimentos propostos de emparelhamento estatístico. Primeiramente, investiga-se se o pressuposto da CIA conduz a resultados inconsistentes com a teoria econômica. A seguir, o uso da informação auxiliar é investigado para verificar se alguma consistência é recuperada.

Na simulação são construídos 500 arquivos síntese, para cada um dos métodos TNRIr e TNRIo, para a regressão e sob CIA. Tanto o arquivo A como o arquivo B serão usados, alternadamente como doadores. Estimam-se as covariâncias e as correlações dessas amostras concatenadas, para calcular as

suas médias, de forma a compará-las com as correlações e covariâncias originais.

Aqui os aspectos prévios de harmonização das bases, seção 2.3, e os cuidados descritos na seção 2.1, que antecedem o emparelhamento estatístico, não são discutidos.

Caso necessário, um processo de escolha das variáveis a serem usadas no emparelhamento estatístico deve ser investigado. Da mesma forma, não foi necessária a escolha das variáveis para compor o núcleo de integração.

Observe-se que o objetivo micro é particularmente atrativo porque fornece uma base de dados ampliada, importante para as variáveis sócio-econômicas dos domicílios, que ainda não foram coletados conjuntamente por alguma pesquisa. E uma aplicação micro é adequada para avaliar a representatividade do arquivo sintético a ser concatenado. Como estabelecido nos capítulos anteriores, quando se considera um objetivo micro, o emparelhamento estatístico primeiro estima um modelo plausível paramétrico ou não-paramétrico das variáveis de interesse, e então imputa a variável faltante no arquivo receptor.

Aqui, o objetivo micro é avaliar os dois diferentes procedimentos propostos e os procedimentos quando pressupõe a CIA e se usa regressão, na habilidade de preservar a estrutura de covariância e de correlação real \sum_{YZ} , comparando-as com as matrizes de covariância e de correlação estimadas, usando os arquivos sínteses obtidos nos quatro procedimentos, via simulação. Verifica-se se existem grandes diferenças entre os procedimentos, no que concerne o erro quadrático médio e o viés, ao comparar os resultados dos três tipos de emparelhamentos, a partir da suas estimativas obtidas para a covariância ou correlação de (Y,Z) .

A tese apresenta funções através dos algoritmos escritos na linguagem R, para obter o $\hat{\Sigma}_{YZ}$ estimado médio do arquivo emparelhado, ou melhor, a média das estimativas de $\hat{\Sigma}_{YZ}$ dos arquivos sínteses gerados nas simulações, tanto usando interpolação como verificando o valor observado mais próximo a ser imputado, a partir da obtenção de um intervalo; além dos métodos que usam regressão e a CIA, acompanhado de sua medida de precisão EQM e viés, cujos códigos estão disponíveis no apêndice anexo. Para mais informação sobre o projeto R visite o sítio <http://www.r-project.org/>.

Nas próximas seções, consideram-se os quatro diferentes métodos. Para cada um, investiga-se a preservação de Σ_{YZ} , que é uma das indicações da representatividade da base de dados emparelhada. Utilizam-se B e A como doadores, com a imputação de Z no arquivo A e Y no arquivo B . Finalmente, concatenam-se os respectivos arquivos sintéticos resultantes da doação dos arquivos A e B , em cada classe X , que resulta nos arquivos sintéticos relativos a cada um dos procedimentos. As doações são realizadas em cada classe.

Recorda-se que com o percentil das variáveis Y (ou Z), e os arquivos A e B se tornam:

- (i) (X_j, Y_j, p_{Y_j}) do arquivo A , de cada domicílio.
- (ii) (X_i, Z_i, p_{Z_i}) do arquivo B , de cada domicílio.

Nesse trabalho se propõe simular e comparar a aplicação de quatro métodos de emparelhamento estatístico na presença de variável contínua. Os procedimentos aqui propostos também se enquadram no caso trivariado. Um objetivo micro com abordagem não-paramétrica de um

emparelhamento estatístico irrestrito, com reposição, obtém o arquivo síntese. Não existe especificação sobre o modelo da distribuição conjunta de (X, Y, Z) . Opta-se pelo uso de uma informação auxiliar, baseada na afiliação estocástica da teoria econômica, evitando a CIA.

6.1.1. Qualidade da informação

Este estudo de caso se baseia em uma das pesquisas domiciliares mais importantes no Brasil, a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada anualmente, em todo território nacional, e que reúne dados sobre emprego, moradia, escolaridade, renda, etc.

O arquivo original (X, Y, Z) é oriundo de uma amostra complexa, a PNAD 2005. Consideram-se, apenas, os domicílios permanentes particulares alugados, que representam 11% da pesquisa na região metropolitana do Rio de Janeiro. Esta pesquisa é produzida pelo IBGE, e tem um desenho amostral complexo com estratificação dos domicílios, para maiores informações vide Albieri et al., 1998.

Neste estudo admite-se que o plano amostral da PNAD (base de dados utilizada na simulação) é equivalente ao de uma amostra simples, e o fator de correção de população finita é aproximadamente 1.

Após uma amostragem sistemática simples, o arquivo original (X, Y, Z) é dividido nos arquivos A e B . O arquivo A conterà as variáveis (X_j, Y_j) e o arquivo B as variáveis (X_j, Z_j) . Como consequência da divisão do arquivo

original, os arquivos A e B são considerados como amostras aleatórias independentes da mesma população, não existindo sobreposição de registros. Por conveniência supõe-se, que X , Y e Z são valores de vetores univariados.

No arquivo original trivariado (X,Y,Z) permaneceram 1206 registros. Dos 1210 registros iniciais, 4 registros foram excluídos devido às restrições lógicas tal como aluguel (Z) maior do que a renda (Y), devido ao domicílio ter tido sua renda diminuída, ou à sub-declaração de renda ou ao desemprego temporário de seu informante.

O vetor X representa a variável de número de cômodos, e os vetores Y e Z , logaritmos da variável renda e da variável aluguel respectivamente, de cada domicílio. As variáveis Y e Z serão chamadas renda e aluguel daqui em diante, a não ser que haja necessidade de explicitar o logaritmo.

Os registros são classificados em L classes de emparelhamento, preliminarmente realizado pela variável X , número de cômodos, no caso uma variável categórica ordinal.

Como antecipado na seção 5.2.1, uma transformação monotônica $g(\cdot)$ de Y e Z gera percentis p_Y e p_Z , respectivamente, e isso transforma o arquivo original (X,Y,Z) na quintupla ordenada (X,Y,Z,p_Y,p_Z) .

A amostra completa com (X,Y,Z) contém 1206 registros que estão representados em forma de histograma para Y (renda) e Z (aluguel), e o *box-plot* para os percentis p_Y e p_Z , na figura 5.

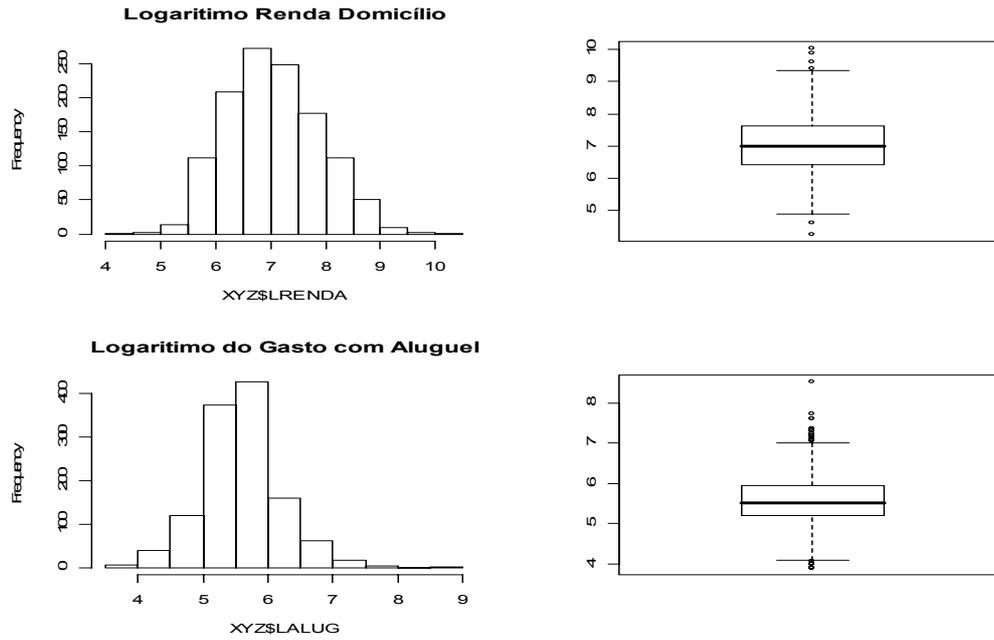


Figura 5 - Histograma e *Box-plot* da distribuição (Y, Z) e seus percentis

Quando se visualizam os dados sem considerar as classes, surge uma função não-linear monotônica crescente, entre variáveis p_Y e p_Z , ver figura 6.

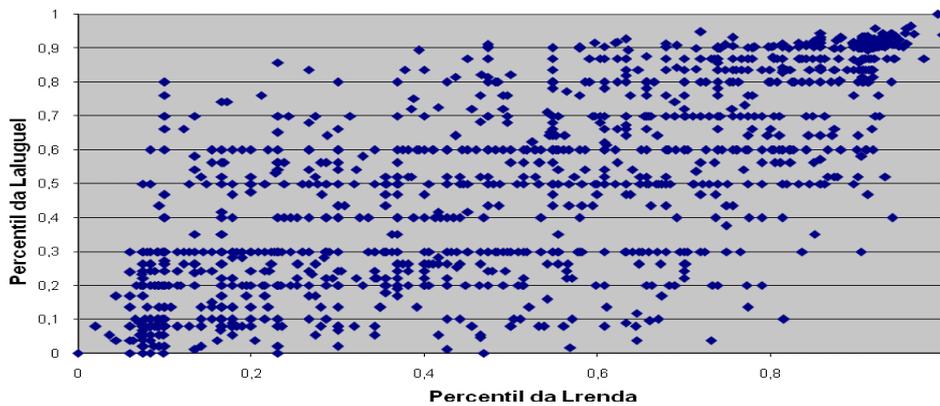


Figura 6 - Distribuição uniforme bidimensional dos percentis p_Y e p_Z , sem “indicação de ajuste” linear.

A figura 7 mostra o arquivo (X, Y, Z) com registros estejam classificados em $L=8$ classes, segundo o número de cômodos.

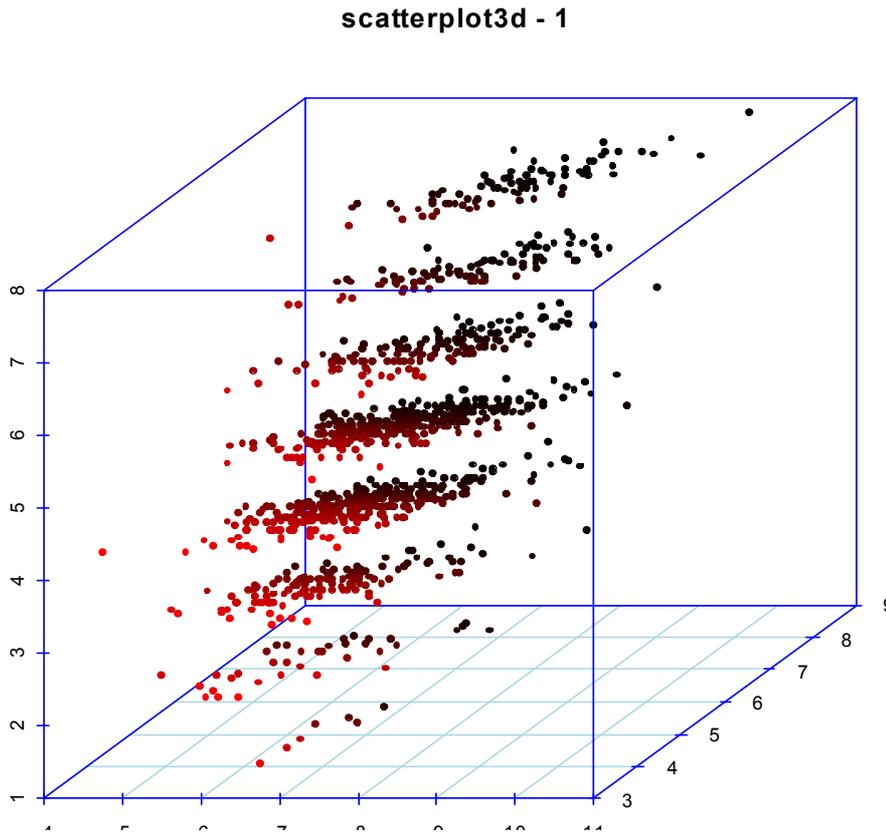


Figura 7 - Distribuição das variáveis da Lrenda e Laluguel por número de cômodos

Na figura 7, por exemplo, observa-se que, em cada classe, as variáveis de renda e aluguel são correlacionadas positivamente, apresentando uma tendência linear. E as variáveis percentis de renda e aluguel seguem a mesma tendência.

O conceito de afiliação estocástica, para as variáveis observadas Y e Z é respaldada pela teoria usada na Economia (aluguel crescente com a renda). As variáveis percentis p_Y e p_Z mantêm a reprodução da afiliação estocástica de (Y,Z) , e têm distribuições uniformes similares entre 0 e 1.

O arquivo composto pela quintupla ordenada (X,Y,Z,p_Y,p_Z) , é apresentado na figura 8.

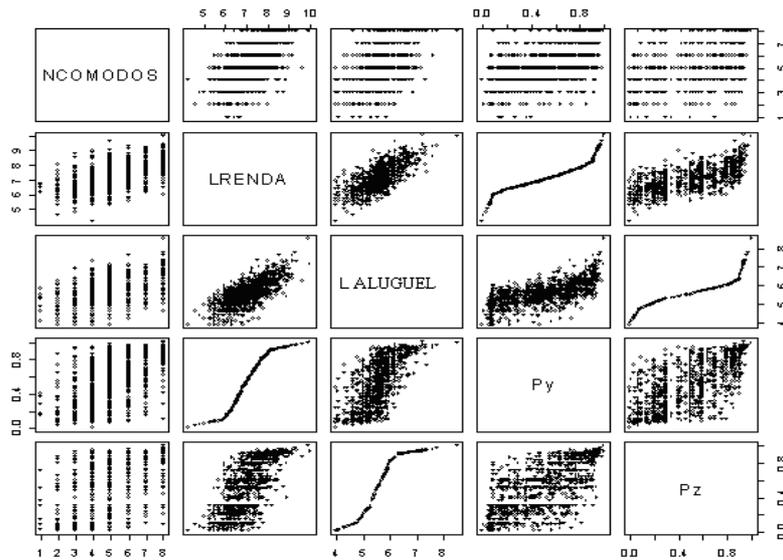


Figura 8 - Distribuição das variáveis da renda e aluguel e seus percentis

Na base completa (X, Y, Z, p_Y, p_Z) , calculam-se os resíduos, como a diferença entre os percentis p_{Y_k} e p_{Z_k} , mantidos em uma base de resíduos:

$$residuo_k = p_{Y_k} - p_{Z_k} \quad k = 1, \dots, n_A + n_B$$

A figura 9, obviamente não linear, ilustra a distribuição ordenada dos resíduos, por número de cômodos.

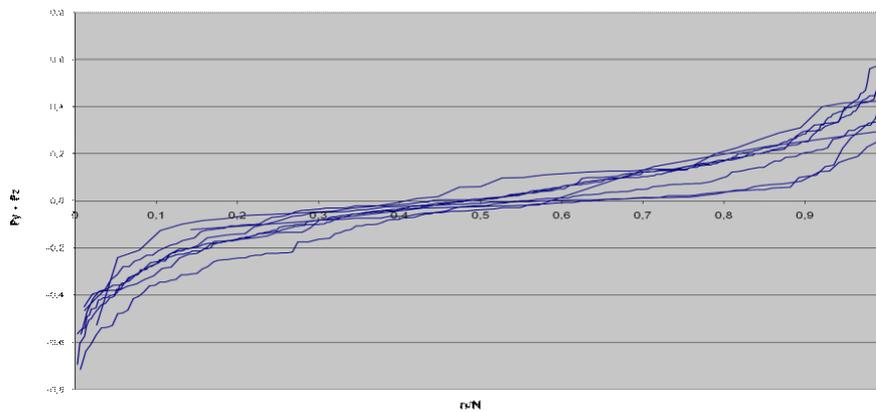


Figura 9 - Resíduos por número de cômodos

A divisão da base de dados original completa realizada por amostragem sistemática simples gera o arquivo A , (X, Y, p_Y) , que possui $n_A = 603$ registros e B , (X, Z, p_Z) , $n_B = 603$ registros e estão descritos na tabela 8, onde prevalecem os domicílios de 3 a 6 cômodos.

Número de cômodos	arquivo A	arquivo B	Total
1	3	3	6
2	19	19	38
3	68	68	136
4	173	173	346
5	175	175	350
6	76	76	152
7	46	46	92
8	43	43	86
Total	603	603	1206

Tabela 8 - Número de registros nas amostras, dos arquivos, por número de cômodos

Pode-se contar o número de registros nas amostras, dos arquivos, por número de cômodos, de acordo com faixas de renda e de aluguel respectivos. O arquivo de classe $L = 5$, número de cômodos, está representado na tabela 9, os valores dessa tabela 9 são os limites inferiores dos intervalos para os domicílios. Denota-se por (*) o limite inferior da classe de Lrenda, sendo (**) o limite inferior da classe de Laluguel. A célula (1,1) da tabela 9, de valor igual a 3, representa os domicílios com Lrenda de até 5,560682e Laluguel de até 3,91202. A figura 10 mostra a distribuição tridimensional dos dados da tabela 9, da renda e do aluguel para domicílios de 5 cômodos.

Lrenda *	Laluguel **					
	3,91202	4,33998	4,76795	5,19591	5,62387	6,05183
5,560682	7	15	9	11	0	0
6,391917	5	11	14	11	0	0
6,659294	0	10	11	13	5	5
5,703782	5	5	17	13	5	5
7,185387	0	5	14	15	6	6
7,473069	0	7	8	17	10	0
7,718685	0	0	8	13	13	7
9,615805	6	7	13	12	6	0

Tabela 9 - Frequência de Lrenda e Laluguel para domicílios de 5 cômodos.

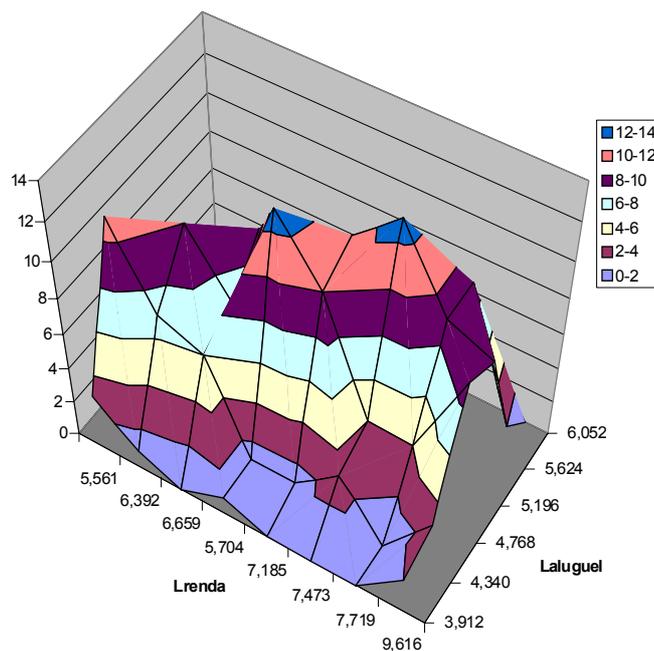


Figura 10 – Distribuição de renda e aluguel para domicílios de 5 cômodos.

Nessa base real de dados (X, Y, Z) , os resultados originais para as matrizes de covariância e de correlação estão representadas na tabela 10, na seção 6.1.2.

6.1.2. Resultados

Nesta seção, investiga-se se a aplicação de quatro procedimentos de emparelhamento estatístico conduz a resultados inconsistentes com a Teoria Econômica, primeiro sob o pressuposto da CIA, vide seção 6.2.2.1, e depois usando a informação auxiliar, com os métodos regressão, vide seção 6.2.2.2, TNRIr e TNRIo, vide seção 6.2.2.3.

Matriz de covariâncias de (X, Y, Z)			Matriz de correlação de (X, Y, Z)		
2,1579	0,7200	0,4424	1	0,5801	0,4896
0,7200	0,7140	0,3823	0,5801	1	0,7355
0,4425	0,3824	0,3784	0,4896	0,7355	1

Tabela 10 - Valores originais das covariâncias e correlações

As tabelas, 11 e 12, resumam os resultados dos quatro procedimentos de emparelhamento estatístico em X , número de cômodos, dos respectivos arquivos sintéticos resultantes da doação dos arquivos A e B . Estimam-se as covariâncias e as correlações das 500 amostras concatenadas, e calculam-se as suas médias, de forma a compará-las com as correlações e covariâncias do arquivo original.

As tabelas 13 e 14 no apêndice D apresentam os resultados dos dois procedimentos de emparelhamento estatístico propostos nesta tese, quando somente B ou A é o arquivo doador nos procedimentos TNRIr e TNRIo.

B e A doadores – Sob CIA e Transformação não-paramétrica relacional					
Matriz de covariância					
		cov(X,Y)	cov(X,Z)	cov(Y,Z)	
	Valores originais	0,7200021	0,4424501	0,3823579	
M É T O D O S	Regressão		0,7223736	0,4354639	0,5068815
		Viés	0,0023715	0,0069862	0,1245236
		EQM	0,0000056	0,0000488	0,0155061
	CIA		0,7195290	0,4428400	0,1505950
		Viés	0,0004731	0,0003899	0,2317629
		EQM	0,0000002	0,0000002	0,0537140
	TNR_{Ir}		0,6580620	0,4221790	0,4352290
		Viés	0,0619401	0,0202711	0,0528711
		EQM	0,0038366	0,0004109	0,0027954
	TNR_{Io}		0,6607830	0,4208740	0,4276920
		Viés	0,0592191	0,0215761	0,0453341
		EQM	0,0035069	0,0004655	0,0020552

Tabela 11 - Estimativas das covariâncias do emparelhamento estatístico segundo os diferentes métodos.

B e A doadores – Sob CIA e Transformação não-paramétrica relacional

		Matriz de correlação		
		cor(X,Y)	cor(X,Z)	cor(Y,Z)
Valores originais		0,5800470	0,4896053	0,7355417
M É T O D O S	Regressão	0,5807365	0,4844695	0,9782821
	Viés	0,0006895	0,0051358	0,2427404
	EQM	0,0000005	0,0000264	0,0589229
	CIA	0,5812860	0,4904530	0,2906160
	Viés	0,0012390	0,0008477	0,4449257
	EQM	0,0000015	0,0000007	0,1979589
	TNRir	0,5286260	0,4718120	0,7430430
	Viés	0,0717020	0,0670840	0,0099740
	EQM	0,0051410	0,0045000	0,0000990
	TNRio	0,5292010	0,4794010	0,7419510
	Viés	0,0726270	0,0551040	0,0087410
	EQM	0,0052750	0,0030360	0,0000760

Tabela 12 - Estimativas das correlações do emparelhamento estatístico segundo os diferentes métodos.

Cada procedimento de emparelhamento estatístico dos arquivos *A* e *B* são apresentados nas seções 6.1.2.1, 6.1.2.2 e 6.1.2.3.

Note, também que, na prática, a partir dos arquivos *A* e *B*, os fatores que podem ser estimados para os mesmos são as suas covariâncias e correlações: $cov(X, Y)$, $cov(X, Z)$, $cor(X, Y)$ e $cor(X, Z)$.

As funções R usadas para a análise estão descritas no apêndice C.

6.1.2.1.

O emparelhamento estatístico pressupondo a CIA

A suposição de que Y (Lrenda) e Z (Laluguel), do domicílio, são independentes, dado X (número de cômodos) equivale a pressupor a CIA.

Se a CIA for válida então $\Sigma_{YZ.X} = 0$, e um arquivo sintético ampliado pode ser obtido, usando uma das técnicas *hot deck* de emparelhamento (seção 3.2).

As tabelas 11 e 12 mostram que, sob o pressuposto da CIA, as estimativas da covariância e da correlação, foram ruins, porque as estimativas de ambas diminuíram de valor, em relação ao valor original de covariância e de correlação.

O arquivo síntese apresenta uma estimativa da covariância igual a 0,150595 que se compara com a do valor original, avaliada em 0,3823579. Também o valor da estimativa da correlação 0,2906160 do arquivo síntese, comparada com a do valor original, avaliada em 0,735542, mostra um resultado menor.

Aqui, o efeito de supor a CIA conduz a resultados inconsistentes, pois a correlação de Y e Z não é preservada após a concatenação, porque a suposição de independência entre renda e aluguel dado o número de cômodos, na teoria econômica não é válida. Fato ratificado pela matriz da covariância (correlação) e pelos valores do EQM respectivos. Ressalte-se que a simulação de reconstruir o arquivo original sob a CIA foi realizado via um procedimento *random hot deck*, descrito na seção 3.2.1.

Nas próximas seções, usa-se a informação auxiliar, com os métodos de regressão (sem adição de resíduos), TNRIr e TNRIo. E os três métodos serão comparados nas tabelas 11 e 12.

6.1.2.2.

Transformação não-paramétrica relacional

A transformação não-paramétrica relacional, sem adição de resíduos que equivale à regressão não-paramétrica, teve suas estimativas de covariância e de correlação aumentadas, ou seja, não foram preservadas.

Nas tabelas, 11 e 12, a estimativa da covariância de Σ_{YZ} ficou igual a 0,5068815, no arquivo síntese e tem valor de covariância original, avaliada em 0,3823579. A estimativa da correlação ficou igual a 0,978282 e comparada com a do valor original, avaliada em 0,735542.

Na aplicação do emparelhamento baseado na distância dos vizinhos mais próximos, considerando a distância absoluta, para o método proposto obteve o pior resultado, dado que as estimativas aumentaram de valor quando o método de regressão sem resíduos foi usada e apresentaram os maiores EQM.

6.1.2.3.

Transformação não-paramétrica relacional intervalar e com interpolação

Baseada na suposição de afiliação estocástica, e na transformação percentil, pode-se usar as variáveis (X_j, Y_j, p_{Y_j}) do arquivo A , e (X_i, Z_i, p_{Z_i}) do arquivo B . Relembre que a suposição de ordenamento dos domicílios de acordo com a sua renda, que com uma transformação monotônica identificou o

par (ρ_Y, ρ_Z) , é uma indicação confiável de afiliação estocástica, segundo a teoria econômica.

Conforme mencionado na descrição dos métodos TNRIr e TNRIo, o uso da transformação monotônica percentil para a renda e o aluguel ordenados pode conter uma preciosa informação para o emparelhamento estatístico de A e B .

O uso desses percentis como informação auxiliar, que advém das próprias amostras a serem emparelhadas, é realizado segundo os métodos TNRIr e TNRIo de emparelhamento estatístico, descritos nas seções 5.2.1 e 5.2.2.

Com 8 classes, X , se refere ao número de cômodos, sendo possível estimar $cov(Y,Z)$ ou $cor(Y,Z)$.

Nos dois procedimentos propostos parte-se dos percentis, aos quais são somados um componente aleatório, cuja predição nos conduz a identificação de um intervalo, cujos limites inferior e superior são usados obter as estimativas dos dois procedimentos.

- **TNRIr**

A estimativa relativa ao método TNRIr utiliza o limite observado mais próximo do valor predito \tilde{Z}_j .

Usa-se o procedimento *hot deck* que busca o vizinho mais próximo com doação nas classes, também conhecido como método de imputação não-paramétrica simples. Em cada classe, o arquivo A agora é uma dupla ordenada (X,Y) e o arquivo B é dupla ordenada (X,Z) . Usa-se um procedimento de imputação não-paramétrico simples, onde \tilde{Z}_j é o valor observado de “ Z_j mais próximo de Y_j ” selecionado no arquivo doador B em relação ao arquivo

receptor A . A predição \tilde{Z}_j é uma estimativa que representa o a -ésimo valor observado de Z imputado em A .

Um procedimento simétrico de imputação *hot deck* é usado quando o arquivo A for o doador.

Dentre esses dois valores de p_Z , escolhe-se o valor mais próximo de \hat{p}_{Z_j} para ser o \hat{Z}_j doador de A .

- **TNRlo**

A estimativa, utilizando o método TNRlo, escolhe o valor mais próximo de \hat{p}_{Z_j} entre os valores observados $p_{Z_{(k-1)}}$ e $p_{Z_{(k)}}$, do intervalo, que será a referência para a doação. Para cada intervalo que contém esse percentil estimado \hat{p}_{Z_j} , seu limite inferior e superior dos percentis observados, tem valores observados correspondentes $Z_{(k-1)}$ e $Z_{(k)}$, respectivos, no arquivo B . O cálculo da interpolação, \hat{Z}_j , utiliza os percentis dos limites do intervalo com seus valores observados correspondentes.

Comparando os quatro métodos, as tabelas 11 e 12, mostram o sumário de 500 simulações, usando o método de transformação não-paramétrica relacional intervalar e com interpolação, onde os arquivos A e B , são usados como doadores. Os métodos de regressão e sob o pressuposto de CIA também são incluídos. São mostradas as suas respectivas covariâncias e correlações estimadas.

Note que com o uso da suposição de afiliação estocástica, da teoria econômica, a base de dados sintética resultante mostrou maior consistência,

porque foi recuperada parte da correlação entre renda e aluguel, a partir do uso da informação auxiliar formalizada pelo percentil.

Na simulação, os métodos não-paramétricos TNRIr e TNRIo, com reposição, apresentaram covariâncias e correlações, que são consideradas as médias das covariâncias e das correlações estimadas, dos arquivos sínteses resultantes, em 500 simulações, entre o logaritmo de renda e aluguel e apresentam valores consistentes com a hipótese de suposição de afiliação estocástica, consistente com a teoria econômica, que gerou a informação auxiliar utilizada, vide tabelas 11 e 12. Onde uma comparação das estatísticas sumárias, para covariâncias e correlações, de renda e aluguel calculados nos arquivos sínteses obtidos via regressão sem resíduos, no caso da CIA (*random hot deck*) e os dois métodos propostos, com respeito às estimativas correspondentes, sem considerar o peso, , com B e A doadores, usando X número de cômodos , como classe.

Pode ser percebido que o resultado, do método TNRIo ao se usar interpolação é ligeiramente superior ao método TNRIr, considerando-se o EQM tanto da $cov(Y,Z)$ quanto da $cor(Y,Z)$.

Ambos os métodos propostos apresentaram resultados consistentes na simulação para reproduzir o arquivo original.

Conforme citado na seção 3.2.3, a questão sobre a qualidade do arquivo síntese e sua habilidade de preservar a distribuição da variável imputada, isto é, Z dado X deve ser sempre discutida. E a figura 11 ilustra essa habilidade ao comparar o histograma da distribuição de Z original e a obtida no arquivo síntese \hat{Z} com B doador.

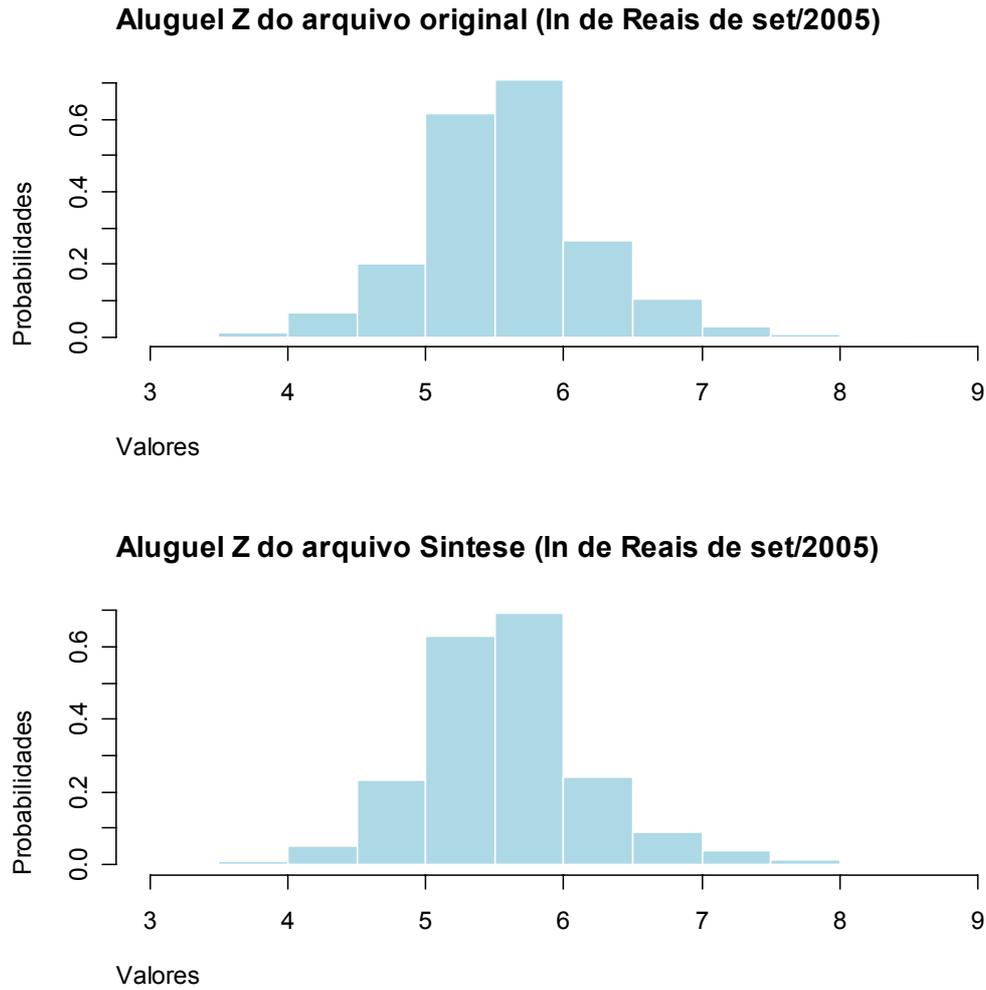


Figura 11 – Distribuição do aluguel antes e depois do emparelhamento

Equivalente a figura 11, a figura 12 ilustra a habilidade ao comparar o histograma da distribuição de Y original e a obtida no arquivo síntese \hat{Y} com A doador.

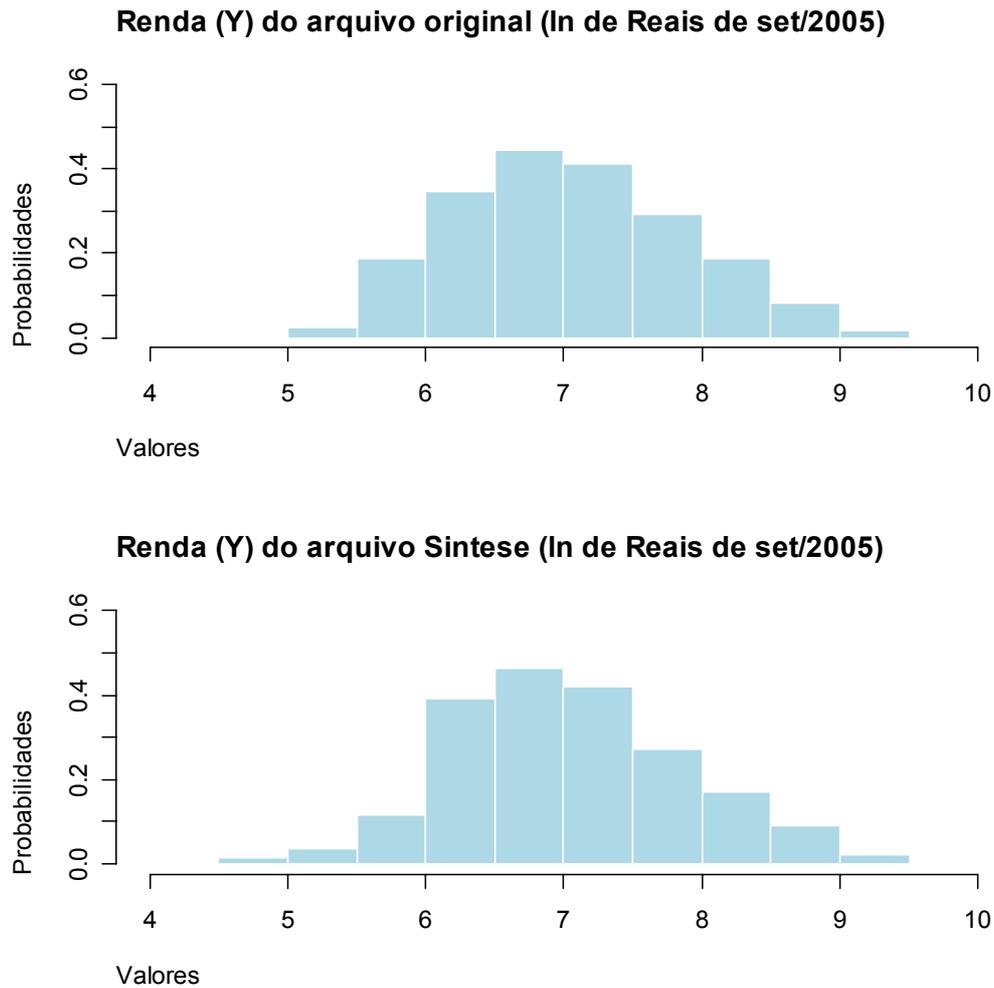


Figura 12 - Distribuição do aluguel antes e depois do emparelhamento

Pode ser percebido que o resultado, do método TNRIo ao se usar interpolação é ligeiramente superior ao método TNRIr, considerando-se o EQM tanto da $cov(Y,Z)$ quanto da $cor(Y,Z)$.

Ambos os métodos propostos apresentaram resultados consistentes na simulação para reproduzir o arquivo original.