

RT-MAE-9118

RECUPERAÇÃO DE SALÁRIOS BRUTOS ATRAVÉS DE
SALÁRIOS LÍQUIDOS

by

Wilton de Oliveira Bussab

and

Rinaldo Artes

Palavras Chaves: imputação, regressão isotônica, método deixar-um-de-fora
(Key words) ("leave-one-out")

Classificação AMS: 62P25, 62J99, 62D05
(AMS Classification)

Recuperação de Salários Brutos Através de Salários Líquidos

Wilton Oliveira Bussab

Rinaldo Artes

IME/USP

Resumo

Em levantamentos amostrais, informações sobre salários são de difícil obtenção. Em muitos casos, quando há resposta, o entrevistado não sabe informar seu salário bruto, declarando apenas o valor do salário líquido. O salário bruto entretanto é um indicador largamente utilizado. Assim utilizando-se informações sobre o salário líquido do entrevistado, sugere-se três procedimentos de imputação do salário bruto: imputação funcional, linear e isotônica. Utilizando-se dados fornecidos pela Fundação SEADE, os procedimentos são comparados através da técnica "deixar-um-de-fora" ("leave-one-out").

1 Introdução

Em levantamentos amostrais, as questões mais sujeitas a distorções por parte dos informantes são aquelas relacionadas a rendimentos. É comum o erro de arredondamento e o de informar apenas o salário líquido ou o bruto.

Neste trabalho são investigados três métodos para imputação do salário bruto a partir do conhecimento do salário líquido, ilustrando-os com dados da Pesquisa Emprego e Desemprego (PED) do DIEESE-SEADE.

Na seção 2 são abordados as razões que motivaram este trabalho.

Na seção 3 são propostas três possíveis alternativas para a imputação do salário bruto a partir do líquido: imputação funcional, linear e isotônica.

Na seção 4 são utilizados dados reais para comparação dos possíveis efeitos dos três métodos usando-se o critério de "deixar um de fora" ("leave-one-out"), DUF, utilizado por Rao(1987).

Finalmente, na seção 5, são apresentadas as recomendações dos autores, baseadas nos resultados obtidos na seção 4.

2 O Problema

A Pesquisa Emprego e Desemprego (PED) é um levantamento mensal da Região Metropolitana de São Paulo, patrocinado pelo DIEESE e SEADE. Ela vem sendo realizada desde outubro de 1984. Uma das informações levantadas pela PED refere-se a salários (líquido e bruto). Por tradição, a PED divulga somente informações relativas a salários brutos, em especial o salário bruto médio. Acontece que muitos dos entrevistados só fornecem informações sobre o salário líquido, o que faz com que uma parte desses dados seja descartada da amostra.

As figuras 1 e 2 ilustram bem a necessidade de se estudar métodos que visem recuperar a informação sobre o salário bruto no âmbito da PED. A figura 1 representa a porcentagem de questionários descartados do cálculo do salário bruto médio devido a não resposta. É interessante constatar a evolução desse percentual. Verifica-se que em determinados períodos há uma queda acentuada desse índice, o que pode estar indicando uma maior preocupação por parte da equipe de crítica em recuperar essa informação, no entanto, até o momento deste trabalho, a recuperação não vinha sendo feita segundo uma metodologia uniforme. Esta é uma das razões desta pesquisa.

Na figura 2 está representado o comportamento mensal da razão entre o salário líquido médio dos questionários utilizados e dos não utilizados na estimativa do salário bruto médio. Exetuando-se alguns poucos meses, essa quantidade tem sido sempre superior a um e apresenta um comportamento crescente. Isso significa que as observações descartadas possuem, em média, um salário líquido maior do que as utilizadas. É natural esperar que o mesmo ocorra com o salário bruto. Pode-se, portanto, concluir que o salário bruto médio vem sendo subestimado.

3 Métodos de Imputação

3.1 Imputação Funcional

Este método de imputação baseia-se na simples observação do sistema de descontos oficiais que incidem sobre o salário bruto. Dois descontos são efetuados mensalmente: imposto de renda (IR) e seguro social (IAPAS). O percentual do desconto varia conforme o valor do salário, pode ser de 8%, 9% ou 10%, no caso do IAPAS e de 0%, 10% ou 25%

Figura 1: Evolucao da Porcentagem de Nao

Declaracao do Salario Bruto

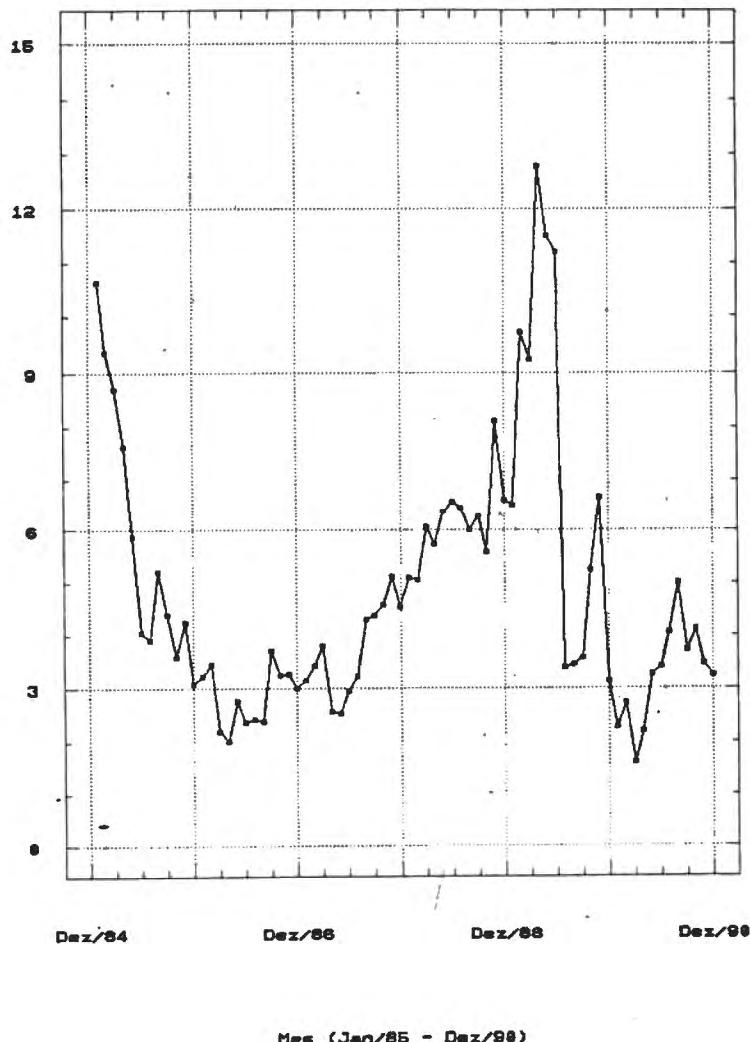
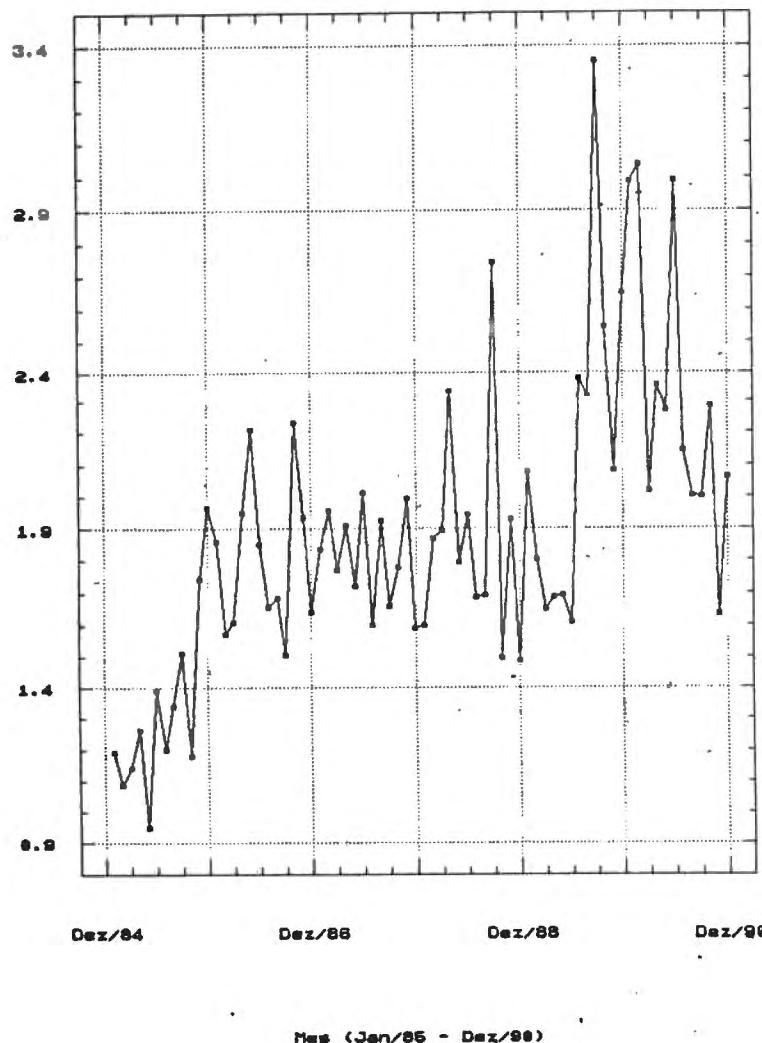


Figura 2: Razao Entre o SL Medio dos que

Nao Declararam o SB e dos que Declararam



no caso do IR, sendo que esse sob o sistema de efeito cascata. Na figura 3 encontram-se as faixas de desconto para o mês de realização da pesquisa (Janeiro de 90).

Combinando-se esses dois percentuais cria-se seis diferentes faixas salariais, figura 3, homogêneas quanto ao desconto. Para cada uma dessas faixas é fácil calcular a equação que expressa o valor do salário bruto a partir do salário líquido. A relação expressa pelas equações só será verdadeira para aqueles salários que não sofrem nenhum outro tipo de desconto e que não tem abatimentos no IR. Essas equações são as utilizadas para a imputação funcional de salários.

3.2 Imputação Linear

Por imputação linear entende-se o ajuste de um modelo de regressão linear simples (Draper & Smith(1981)) tendo por variável resposta o salário bruto e por preditora o líquido.

As principais restrições quanto ao uso dessa metodologia é que a técnica exige a suposição de uma relação linear entre o salário bruto e líquido, o que nem sempre pode ser garantido. Além disto, pressupõe-se a homocedasticidade dos resíduos, que também não pode ser garantida, pois sabe-se que a variância dos salários brutos aumenta conforme aumenta o salário líquido. Visando minimizar estes problemas, decidiu-se estimar um modelo diferente para cada uma das faixas salariais definidas no tópico anterior.

3.3 Imputação Isotônica

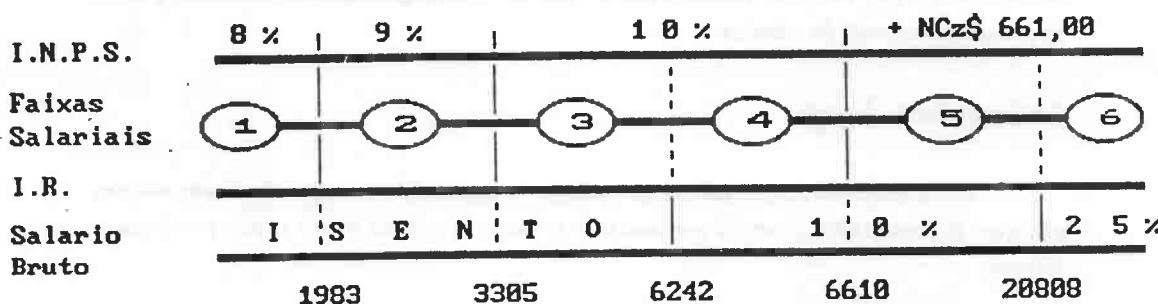
Em problemas usuais de regressão, fixa-se uma série de valores de uma variável preditora e para cada um verifica-se o comportamento de uma variável de interesse y . O objetivo é estimar os parâmetros de uma função $g(x)$, onde $g(x) = E[y/x]$.

A regressão isotônica (Barlow et al(1972)), tem os mesmos objetivos da regressão usual, com a restrição de $g(x)$ ter um comportamento monótono; ou seja, se $x_1 > x_2 \Rightarrow g(x_1) > g(x_2)$.

Na verdade, por definição, uma função $g^*(x)$ será a regressão isotônica de y em x , ponderada por $w(x)$, se minimizar a quantidade

$$\sum_{(x,y)} [y - g(x)]^2 w(x)$$

Figura 3: Descontos para o mes de Dezembro de 1989 - Pesquisa de Janeiro de 1990



sujeita à restrição $x_1 > x_2 \Rightarrow g(x_1) > g(x_2)$.

É razoável supor que se o salário líquido de um grupo de indivíduos é superior ao de um outro grupo, o mesmo acontecerá com os salário bruto. Essa suposição sugere a adoção da regressão isotônica.

Como nos casos anteriores, decidiu-se estimar um modelo para cada diferente faixa salarial. A nível de ilustração, a figura 4 reproduz a regressão isotônica da faixa 2. Note que geometricamente essa regressão é uma linha poligonal.

4 Comparação dos Métodos

4.1 Critério de Comparação

As comparações dos diferentes métodos de imputação são realizadas através do estudo do seu efeito no cálculo do salário médio e também através de uma adaptação do erro quadrático médio (EQM) de cada modelo estimado.

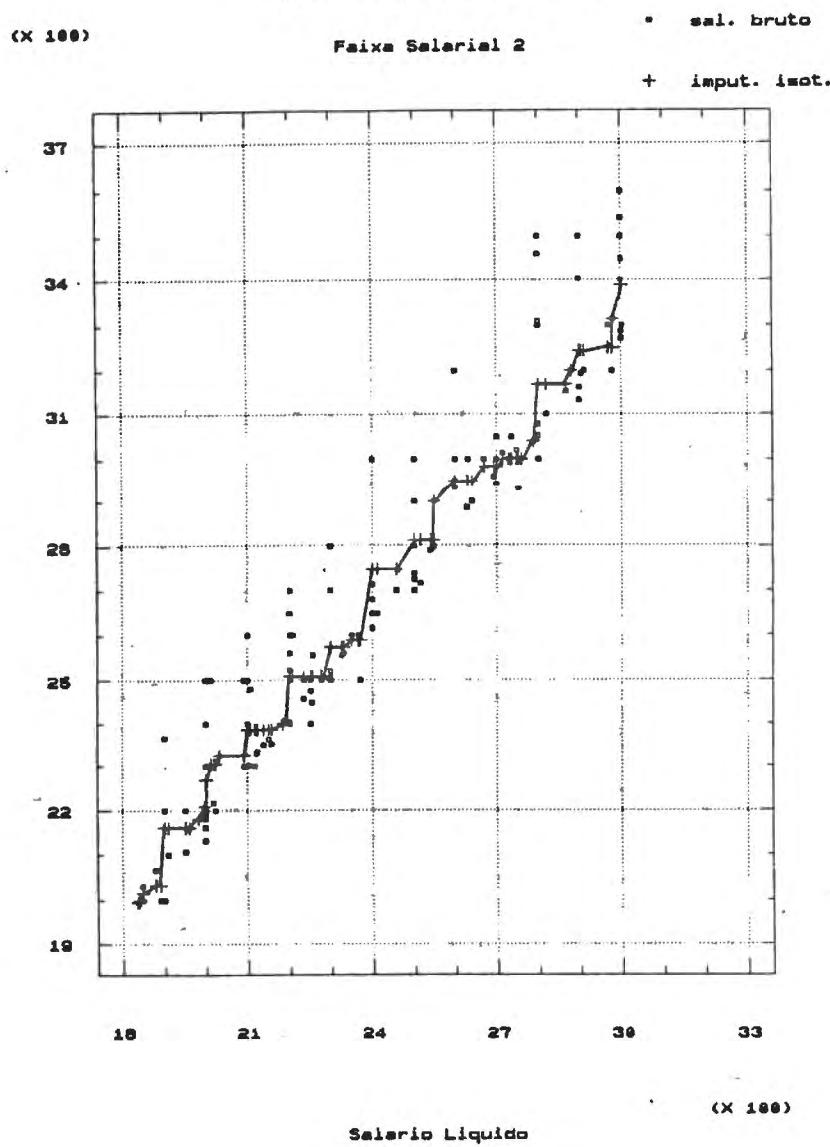
Um dos problemas ao se utilizar o EQM estimado é que seu valor é fortemente influenciado por observações desajustadas. Para atenuar esse efeito, Rao(1987) utiliza o método DUF que também avalia a qualidade preditiva dos modelos.

Considere x_j o salário líquido do indivíduo j e y_j o seu salário bruto. A amostra completa é dada por $(x_j, y_j), j = 1, 2, \dots, n$. Para aplicação do método DUF é necessário, inicialmente, que cada modelo de imputação seja estimado n vezes, sendo que para a i -ésima estimativa deve ser utilizado $(x_j, y_j), j = 1, 2, \dots, i-1, i+1, \dots, n$ como amostra. Considere $\hat{y}_j(i)$ a previsão do salário j obtida a partir da estimativa do modelo sem a observação i . Define-se então como estimador do EQM do modelo a quantidade

$$E\hat{Q}M = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{y}_i(i)]^2.$$

Calculada essa estimativa para cada diferente modelo de imputação, o escolhido será aquele que possuir menor $E\hat{Q}M$.

Figura 4: Imputação Isotônica



4.2 Amostra

Na amostra da PED do mês de janeiro de 1990, fez-se um acompanhamento detalhado do trabalho da equipe de crítica em relação ao levantamento dos rendimentos dos assalariados, principalmente quanto a recuperação dos salários brutos a partir dos salários líquidos. A figura 5 resume os resultados observados.

A amostra de trabalho, utilizada para testar o efeito dos três métodos de estimação descritos, foi obtida excluindo dos 1970 casos ocorridos naquele mês:

- (i) 247 casos que só declararam renda líquida e tiveram os salários brutos imputados pela crítica;
- (ii) 58 casos que só declararam a renda líquida e não foram aproveitados pela crítica;
- (iii) 22 casos que só declararam a renda bruta e
- (iv) 161 casos que declararam ambas as rendas e foram considerados valores aberrantes por mostrarem grande discrepância entre os dois salários.

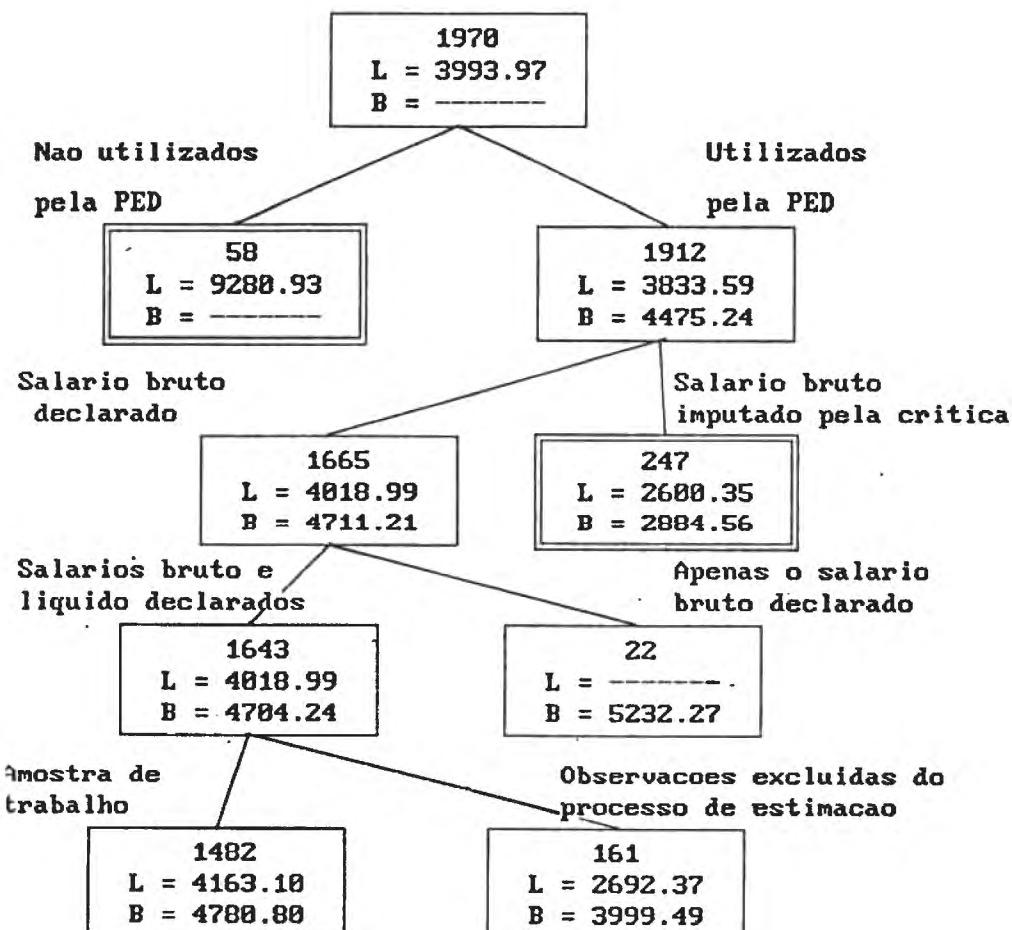
Desse modo, a amostra de trabalho foi formada por 1482 observações.

4.3 Resultados Empíricos

A tabela 1 apresenta o resultado da aplicação do método DUF aos modelos de imputação. Nota-se que o erro quadrático médio estimado para a imputação isotônica foi inferior aos da funcional e linear, com exceção da faixa salarial 4, onde tem-se apenas 10 observações. Logo, quanto à qualidade do ajuste neste exemplo, o melhor modelo de imputação é o isotônico, seguido pelo linear e funcional.

Na figura 6 apresenta-se o desempenho dos três métodos de imputação quanto ao valor final do salário médio. Os métodos foram utilizados na imputação dos 58 questionários descartados e dos 247 que haviam sido imputados pela crítica. Como era previsto, o procedimento atual subestima o rendimento bruto em cerca de 4%, o valor exato dependerá do método de imputação. Esse viés é devido principalmente a apenas 53 casos, 2,9% da amostra de declarantes, que reforça a idéia de que quanto maior o rendimento

**Figura 5 – Dados sobre o salario
medio de assalariados – PED
(JAN/90)**

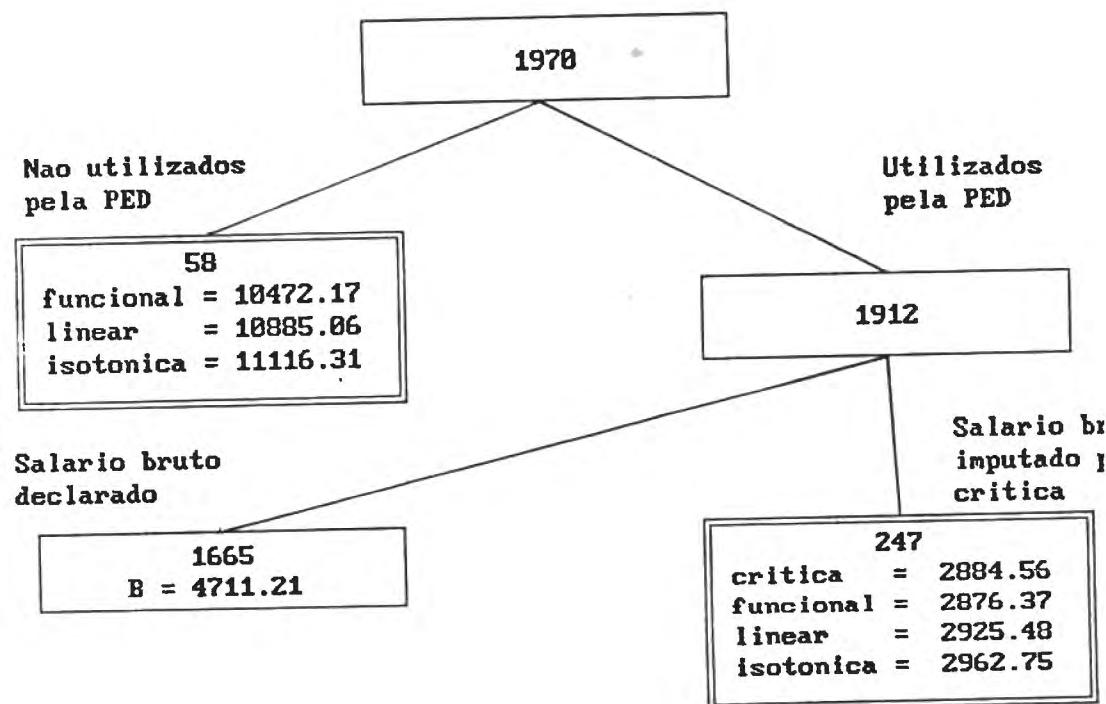


Legenda

#
L =
B =

#: numero de observacoes
L: salario liquido medio
B: salario bruto medio

Figura 6 - Efeito das imputações no salário bruto médio



Efeito Final das imputações

1970		Acréscimo em relação a média atual
Atual	= 4475.24	
funcional	= 4650.77	3.92 %
linear	= 4669.08	4.33 %
isotonica	= 4680.56	4.59 %

menos pessoas lembram-se de seus salários brutos. A crítica também tem um procedimento viesado: recupera mais os rendimentos menores.

5 Conclusão

A imputação isotônica mostrou-se superior em termos do EQM, porém da figura 6 observa-se que o efeito dos três métodos sobre o cálculo do salário bruto é muito semelhante do ponto de vista prático. Do ponto de vista operacional, o procedimento funcional é o de mais simples implantação, pois a imputação já pode ser feita na etapa da crítica. Assim, a tendência dos autores é recomendar a escolha deste como método adequado à pesquisa.

Embora a aplicação tenha sido restrita a uma pesquisa particular, desconfia-se do mesmo padrão para levantamentos similares, porém a magnitude do viés deve variar em função da região da pesquisa.

6 Bibliografia

BARLOW, R.E., BARTHOLOMEW, D.J., BREMER, J.M. & BRUNK, H.D. (1972). *Statistical inference under order restriction*. New York, Wiley.

DRAPE, N.R. & SMITH, H. (1981). *Applied regression analysis*. 2nd ed. New York, Wiley.

RAO, C.R. (1987) Prediction of future observations in growth curve models. *Statistical Science*. 2(4):434-71.

TABELA 1. EQM estimado pelo método DUF

Faixa salarial	Imputação		
	funcional	linear	isotônica
1	4337.01	3360.70	2940.41
2	12042.70	10332.57	9336.26
3	22132.88	20403.03	19068.60
4	825.00	4024.36	4000.00
5	999458.90	808988.31	742157.74
6	3563521.40	3610360.94	3326559.03

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1990

- 9001 - JR., HODGES, J.L. RAMSEY, P.H. & WECHSLER, S.; Improved Significance Probabilities of the Wilcoxon Test, 30p.
- 9002 - PAULA, G.A.; Bias Correction for Exponential Family Nonlinear Models, 10p.
- 9003 - SCHONMANN, R.H. & TANAKA, N.I.; One Dimensional Caricature of Phase Transition, 16p.
- 9004 - ZACKS, S. & BOLFARINE, H.; Maximum Likelihood Prediction in Finite Populations, 15p.
- 9005 - ZACKS, S. & BOLFARINE, H.; Equivariant Prediction of the Population Variance Under Location-Scale Superpopulation Models, 8p.
- 9006 - BOLFARINE, H.; Finite Population Prediction Under Error in Variables Superpopulation Models, 14p.
- 9007 - BOLFARINE, H.; Ratio and Regression Estimators Under Error-In-Variables Superpopulation, 16p.
- 9008 - BOTTER, D.A. & MOTTA, J.M.; Experiments With Three-Treatment Three-Period Crossover Design: Analysis Through the General Linear Models, 15p.
- 9009 - SCHONMANN, R.H.; Finite Size Scaling Behavior of a Biased Majority Rule Cellular Automaton, 11p.
- 9010 - BUENO, V.C.; A Martingale Version of Multistate Monotone Systems, 5p.
- 9011 - CORDEIRO, G.M. & FERRARI, S.L.P.; A Modified Score Tests Statistic Having Chi-Squared Distribution to Order, 12p.
- 9012 - NEVES, E.J. & SCHONMANN, R.H.; Behavior of Droplets for a Class a Glauber Dynamics at very Low Temperature, 33pg.
- 9013 - FERRARI, P.A.; Shock Fluctuations in Asymmetric Simple Exclusion, 26p.
- 9014 - MAGALHAES, M.N. & TANAKA, N.I.; Reverse Process in M/M/1 Queues, 7p.

- 9015 - CORDEIRO, G.M. & PAULA, G.A.; Estimation Large Sample Parametric Tests and Diagnostic for Non-Exponential Family Nonlinear Models, 31p.
- 9016 - CORDEIRO, G.M., FERRARI, S.L.P. & PAULA, G.A.; Improved Score Tests for Generalized Linear Models, 25p.
- 9017 - ESTON, V.R., GALVES, A., JACOBI, C.H., & LANGENVIN, R.; Dominance Switch Between Two Interacting Species and Large Deviations, 8p.
- 9018 - IRONY, T.Z., PEREIRA, C.A.B. & BARLOW, R.E.; Bayesian Models for Quality Assurance, 22p.
- 9019 - BOLFARINE, H. & ZACKS, S., Bayes and Minimax Prediction in Finite Populations, 16p.
- 9020 - BOLFARINE, H.; RODRIGUES, J. & ZACKS, S.: Some Asymptotic Results in Finite Populations, 13p.
- 9021 - BOLFARINE, H., CORDANI, L.K.; Estimation of a Structural Linear Regression Model With a Known Reliability Ratio, 9pg.
- 9022 - PAULINO, C.D.M., PEREIRA, C.A.B.; Bayesian Analysis of Categorical Data Informatively Censored. São Paulo, 18pg
- 9023 - FERRARI, P.A., MARTINEZ, S. & PICCO, P.; Some Properties of Quasi Stationary Distributions in the Birth and Death Chains, 11pg.
- 9024 - BOLFARINE, H., SANDOVAL, M.C.; Prediction of the finite population distribution function under Gaussian super population models, 13 pg.
- 9025 - BUENO, V.C. & NORROS, I.; Component importance through compensator transforms, 7p

1991

- 9101 - RAVISHANKAR, K.; Convergence of a random motion to a point, 11p.
- 9102 - FERRARI, P.A. & KAVISHANKAR, R.; Shocks in asymmetric exclusion automata, 17p.
- 9103 - ANDRE, C.D.S. & PERES, C.A.; An Algorithm for the Mse Estimation of the Multistage Dose-Response model, 14p.

9104 - FERRARI, P.A.; Microscopic shocks in one dimensional driven systems, 21p.

9105 - BOLFARINE, H., LEITE, J.G. & RODRIGUES, J.; On the Estimation of the Size of a Finite and Closed Population, 17p.

9106 - IGLESIAS, P., SANDOVAL, M.C. & PEREIRA, C.A.B.; Predictive Likelihood in Finite Populations, 17p.

9107 - CORDEIRO, G.M., PAULA, G.A. & BUTTER, D.A.; Improved Likelihood Ratio Tests for Dispersion Models, 32p.

9108 - WERKEMA, M.C.C. & TANAKA, N.I.; Alguns Aspectos sobre Filtragens em Filas, M/M^{1,1}/N com "Feedback", 35p.

9109 - WECHSLER, S.; Bayesian Note on Game Theory, 3p.

9110 - WECHSLER, S.; Rationality in Rubin's Sense, 16p.

9111 - BOLFARINE, H., ELIAN, S.N., RODRIGUES, J. & ZACKS, S.; Optimal Prediction of the Finite Population Regression Coefficient in Finite Populations, 13p.

9112 - PAULA, G.A.; Assessing Local Influence in Restricted Regression Models, 20p.

9113 - RODRIGUES, F.W. & WECHSLER, S.; A Discrete Bayes Explanation of a Failure Rate Paradox, 03p.

9114 - FERRARI, P.A., MARTINEZ, S. & PICCO, P. Existence of Quasi Stationary Distributions. A Renewal Dynamical Approach, 13p.

9115 - FERRARI, P.A., MARTINEZ, S. & PICCO, P. Existence of Non Trivial Quasi Stationary Distributions in the Birth and Death Chain, 20p.

9116 - FERRARI, P.A.; Shock in the Burgers Equation and the Asymmetric Simple Exclusion Process, 40p.

9117 - BOLFARINE, H., NASCIMENTO, J.A. & RODRIGUES, J. On the Estimation of the True x Values in a Simple Regression Model When Both Variables Have Error, 15p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01498 - São Paulo, Brasil.

