DIFERENCIAIS DE SALÁRIO: QUESTÕES METODOLÓGICAS E ILUSTRAÇÕES

Ricardo Paes de Barros*
Jaime Pereira de Pontes
Simone Varandas

Resumo

Este texto possui tres objetivos. Primeiro procura-se enfatizar a importância da sensibilidade dos diferenciais ao conjunto de variáveis de controle utilizado. Em segundo lugar, investiga-se a natureza desta sensibilidade, isto é, como propriedades das variáveis de controle se relacionam com a magnitude e direção da sua influência sobre os diferenciais de salários estimados. Finalmente, apresenta-se um estudo do efeito da posse da carteira de trabalho sobre o nível dos salários, com exemplo e contra exemplo, sendo este último um caso de como um aumento no número de variáveis de controle não induz, necessariamente, a redução do diferencial de salários.

Abstract

The objective of this paper is threefold. First, it seeks to stress the importance of the sensitivity of wage differentials to the set of control variables used. Secondly, the nature of this sensitivity is investigated, that is, how the properties of the control variables are related to the magnitude and direction of their influence on the estimated wage differentials. Finally, the paper presents a study of the effect of possession of work documents on the wage level, with an example and counter-example, the latter being a case of how an increase in the number of control variables does not necessarily lead to a reduction of the wage differential

1. Introdução

A avaliação de diferenciais de salário tem ocupado papel de destaque na literatura empírica referente à determinação de salários na

* Os autores pertencem ao Instituto de Planejamento Econômico e Social — IPEA.

economia brasileira. Alguns exemplos são as análises dos diferenciais de salário entre ramos de atividade, Cunha e Bonelli (1978) entre homens e mulheres, Camargo e Serrano (1983) entre empresas públicas e privadas, Macedo (1985), entre trabalhadores urbanos e rurais, Rezende (1985) e entre empregados com e sem carteira de trabalho assinada, Barros e Varandas (1987).

Um aspecto controverso nessas avaliações é a sensibilidade dos diferenciais de salário ao conjunto de variáveis de controle utilizado. Por exemplo, Macedo (1985, tabela 12) quando analisa diferenciais de salário entre empresas públicas e privadas observa que ao controlar por idade, educação, experiência e sexo, os diferenciais se reduzem a cerca de 1/4 em relação ao diferencial global obtido antes do controle das variáveis. Nada garante, contudo, que os diferenciais estimados variem de forma monotônica à medida que o número de variáveis de controle se eleva. A tabela 2 demonstra este fato quando se compara os salários dos empregados com carteira com os salários dos trabalhadores por conta própria. Neste caso, à medida que um maior número de variáveis de controle é introduzido, o diferencial de salário não varia numa mesma direção, além do que seu sinal se reverte. Por outro lado, os diferenciais de salário entre empregados com e sem carteira se comportam conforme o esperado. De fato, estes diferenciais reduzem-se monotonicamente à medida que o conjunto de variáveis de controle é expandido.

Nesta nota investiga-se a natureza desta sensibilidade dos diferenciais de salário ao conjunto de variáveis de controle utilizado. Particularmente, procura-se estudar como propriedades das variáveis de controle se relacionam com a magnitude e direção de sua influência sobre os diferenciais de salário estimados. A análise, contudo, restringe-se ao caso em que todas as variáveis de controle são de natureza binária. Os resultados obtidos decorrem e, portanto, dependem precisamente desta natureza binária das variáveis de controle.

De acordo com o que tem sido enfatizado na literatura¹, a introdução de uma variável de controle só afetará o estimador do efeito de um dado atributo sobre o salário, caso a variável de controle considerada, condicionada às variáveis previamente incluídas, esteja correlacionada simultaneamente com (a) o nível salarial e (b) a incidência do atributo. Por exemplo, no caso particular da avaliação do impacto da carteira de trabalho, a introdução de uma variável de controle somente terá influência se esta variável estiver simultaneamente correlacionada com (a) o nível salarial e (b) a posse da carteira de trabalho. Na próxima seção deriva-se uma expressão para o caso de variáveis de controle binárias que além de demonstrar este fato, expõe exatamente como

¹Ver Anderson et alii (1980) ou Judd and Kenny (1981).

estas duas correlações interagem para determinar a direção e magnitude da sensibilidade do diferencial de salário estimado.

A terceira seção desta nota enfatiza e ilustra a utilidade dos resultados obtidos no contexto de um estudo do efeito da posse da carteira de trabalho sobre o nível dos salários, com base em informações da PNAD-85. Na última seção são apresentadas algumas conclusões e um sumário dos resultados.

2. Teoria

Seja $(Y, b, a_1, \dots, a_m)$ um conjunto de variáveis aleatórias onde $b, a_1, \dots, a_m$ são variáveis binárias podendo apenas assumir os valores 0 ou 1. A variável b é o atributo cujo efeito sobre o salário Y se deseja estudar. As variáveis $a_1, \dots, a_m$ formam o conjunto de variáveis de controle observadas.

Suponha que se decida estimar o efeito do atributo b sobre o salário Y controlando-se apenas pelo subconjunto de variáveis $(a_1, \dots, a_k)$, para algum $k \leq m$.² Dada a natureza binária das variáveis, existem 2^k possíveis valores que o vetor $A_k \equiv (a_1, \dots, a_k)$ pode assumir. Seja P_k o conjunto formado por estes 2^k valores e denote por p_k um elemento genérico deste conjunto. Portanto, $p_k = (0, 1)^k$ e p_k é uma sequência de k valores zeros e uns, comumente chamada de uma célula. Denota-se por $B_k^*(p_k)$ o efeito do atributo b sobre a média de Y na célula p_k .

$$B_k^*(p_k) \equiv E[Y \mid b = 1, A_k = p_k] - E[Y \mid b = 0, A_k = p_k].$$

Em geral, $B_k^*(p_k)$ varia com p_k . Assim, torna-se conveniente definir o "efeito médio de b sobre o salário", B_k , via³

$$B_k \equiv \sum_{p_k} B_k^*(p_k) \cdot P[A_k = p_k].$$

note que, se $(a_1, \dots, a_m)$ e b forem independentes, então

$$B_k = B_0 = E[Y \mid b = 1] - E[Y \mid b = 0]$$

para todo $k \leq m$,

² É importante incluir na análise que se segue o caso em que $k = 0$, neste caso o conjunto de variáveis de controle efetivamente utilizado é vazio.

³ O somatório $\sum_{p_k}$ é entendido como sendo a soma sobre todas as células $p_k \in P_k$.

isto é, B_k , o efeito médio do atributo b sobre Y , não varia com k , o número de variáveis de controle incluídas. Caso contrário, B_k em geral depende de k . Considere, portanto, a questão de determinar como B_k varia quando decide-se aumentar o número de variáveis de controle de k para $k + 1$. Ou seja, considere a diferença $D_k \equiv B_k - B_{k+1}$.⁴

Proposição 1:

$$D_k = \sum_{p_k} D_k^*(p_k) \cdot P[A_k = p_k].$$

onde:

$$D_k^*(p_k) \equiv B_k^*(p_k) - \bar{B}_k(p_k)$$

e

$$\begin{aligned} \bar{B}_k(p_k) &\equiv B_{k+1}^*(p_k, 0) \cdot P[a_{k+1} = 0 | A_k = p_k] + \\ &+ B_{k+1}^*(p_k, 1) \cdot P[a_{k+1} = 1 | A_k = p_k]. \end{aligned}$$

$\bar{B}_k(p_k)$ pode ser apropriadamente interpretado como o efeito de b sobre a média de Y na célula p_k , quando controla-se não só pelo conjunto de variáveis $(a_1, \dots, a_k)$ como também pela variável a_{k+1} . Assim, $D_k^*(p_k)$ mede a sensibilidade do efeito de b sobre a média de Y , na célula p_k , com respeito à inclusão da variável de controle a_{k+1} .

Sob que condições a respeito da variável a_{k+1} , $D_k^*(p_k)$ é nulo? Para responder a esta indagação é importante notar que:

Proposição 2:

$$D_k^*(p_k) = r_k(p_k) \cdot \bar{C}_k(p_k)$$

onde

$$\begin{aligned} r_k(p_k) &\equiv P[a_{k+1} = 1 | b = 1, A_k = p_k] - \\ &- P[a_{k+1} = 1 | b = 0, A_k = p_k], \\ \bar{C}_k(p_k) &\equiv C_k^0(p_k) \cdot P[b = 1 | A_k = p_k] + \end{aligned}$$

⁴ A demonstração de todas as proposições desta seção encontram-se no apêndice.

$$+ C_k^1(p_k) \cdot P[b = 0 | A_k = p_k,$$

$$C_k^0(p_k) \equiv E[Y | b = 0, A_k = p_k, a_{k+1} = 1] -$$

$$- E[Y | b = 0, A_k = p_k, a_{k+1} = 0],$$

e

$$C_k^1(p_k) \equiv E[Y | b = 1, A_k = p_k, a_{k+1} = 1] -$$

$$- E[Y | b = 1, A_k = p_k, a_{k+1} = 0].$$

Note que $C_k^0(p_k)$ e $C_k^1(p_k)$ representam o impacto da variável a_{k+1} sobre o nível médio dos salários, na célula p_k , entre aqueles cujo atributo $b = 0$ e aqueles cujo atributo $b = 1$, respectivamente, $\bar{C}_k(p_k)$ é uma média ponderada destes dois impactos com os pesos revertidos.⁵ Assim, $\bar{C}_k(p_k) > 0$ significa que o atributo a_{k+1} tende a elevar o nível dos salários, ao menos na célula p_k .

A proposição 2 implica que $D_k^*(p_k) = 0$ se e somente se $r_k(p_k) = 0$ ou $\bar{C}_k(p_k) = 0$. Por definição, $r_k(p_k) = 0$ se e somente se a variável a_{k+1} é independente de b na célula p_k . Por outro lado, uma condição suficiente para que $\bar{C}_k(p_k) = 0$ é que $C_k^0(p_k) = C_k^1(p_k) = 0$, ou seja, que o efeito de a_{k+1} sobre a média de Y na célula p_k seja nulo tanto entre aqueles cujo atributo $b = 0$ como entre aqueles cujo atributo $b = 1$.

Na próxima seção procura-se qualificar e ilustrar a utilidade da proposição 2. É oportuno, contudo, mencionar aqui que esta proposição possui uma interpretação bastante simples e intuitiva. Note que $\bar{C}_k$ é uma medida do "impacto da variável a_{k+1} sobre o nível dos salários", enquanto que r_k mede "em que porcentagem a incidência de indivíduos tipo $a_{k+1} = 1$ é mais comum no grupo $b = 1$ do que no grupo $b = 0$ " (veja figura 1). Assim, a proposição 2 estabelece que (i) se a incidência de indivíduos "tipo $a_{k+1} = 1$ " é $\alpha\%$ maior no "grupo $b = 1$ " do que no "grupo $b = 0$ " e (ii) se a_{k+1} eleva em média os salários em λ , então quando a variável a_{k+1} é omitida da análise o impacto de b sobre os salários encontrar-se-á superestimado em $\alpha\%$ de λ com relação ao que se obteria caso esta variável, a_{k+1} , fosse introduzida.

⁵ Esta reversão de pesos é um resultado também comumente encontrado em interpolação linear. Seja q um ponto entre q_0 e q_1 . Além disso, seja $d(q_0, q)$, $d(q, q_1)$ e $d(q_0, q_1)$ as distâncias entre estes pontos. Defina $\alpha_0 \equiv d(q_0, q) / d(q_0, q_1)$ e $\alpha_1 \equiv d(q, q_1) / d(q_0, q_1)$ então $q = \alpha_1 \cdot q_0 + \alpha_0 \cdot q_1$. Note a reversão dos pesos nesta expressão para q .

DIAGRAMA 1
E[15.0]
(em salários mínimos)

DIAGRAMA 2

P L. 1

(%)

EMPREGADOS COM CARTEIRA
VERSUS
EMPREGADOS SEM CARTEIRA

EMPREGADOS COM CARTEIRA
VERSUS
TRABALHADORES POR CONTA PRÓPRIA

DIAGRAMA 3

8^a (1, 00454

(SALÁRIOS MÍNIMOS)

EMPREGADOS COM CARTEIRA
EMPREGADOS SEM CARTEIRAEMPREGADOS COM CARTEIRA
VERSUS
TRABALHADORES POR CONTA PRÓPRIA

DIAGRAMA 4

$$D_k^*(.), \quad 0 \leq k \leq 4$$

(SALÁRIOS MÍNIMOS)

EMPREGADOS COM CARTEIRA
VERSUS
EMPREGADOS SEM CARTEIRA

EMPREGADOS COM CARTEIRA
VERSUS
TRABALHADORES POR CONTA PRÓPRIA

DIAGRAMA 5

$$\bar{C}_k(\cdot), 0 \leq k \leq 4$$

(SALÁRIOS MÍNIMOS)

EMPREGADOS COM CARTEIRA
VERSUS
EMPREGADOS SEM CARTEIRA

EMPREGADOS COM CARTEIRA
VERSUS
TRABALHADORES POR CONTA PRÓPRIA

DIAGRAMA 6

$r_k(1.)$, $0 \leq k \leq 4$

(%)

EMPREGADOS COM CARTEIRA
VERSUS
EMPREGADOS SEM CARTEIRA

EMPREGADOS COM CARTEIRA
VERSUS
TRABALHADORES POR CONTA PRÓPRIA

As proposições 3 e 4 abaixo apresentam interpretações alternativas para $r_k(p_k)$, $C_k^0(p_k)$ e $C_k^1(p_k)$.

Proposição 3:

$$r_k(p_k) = \text{COV}[b, a_{k+1} | A_k = p_k] / \text{Var}[b | A_k = p_k].$$

Proposição 4:

$$C_k^1(p_k) = \text{COV}[Y, a_{k+1} | b = i, A_k = p_k] / \text{Var}[a_{k+1} | b = i, A_k = p_k]$$

$$i \in (0, 1).$$

A proposição 3 demonstra que $r_k(p_k)$ é o coeficiente da regressão de a_{k+1} em b na célula p_k , enquanto que a proposição 4 demonstra que $C_k^1(p_k)$ é o coeficiente da regressão de Y em a_{k+1} condicional a que $b = i$ na célula p_k , $i \in (0, 1)$. Fica, portanto, demonstrado que $D_k = 0$ se, para toda célula $p_k \in P_k$, (i) a_{k+1} não estiver correlacionado com b e/ou (ii) a_{k+1} não estiver correlacionada com Y tanto condicional a que $b = 0$ quanto condicional a que $b = 1$.

3. Ilustrações: O caso do efeito da carteira de trabalho sobre o nível dos salários

Nesta seção, com base em dois exemplos, procura-se ilustrar a análise teórica desenvolvida na seção anterior. Nos exemplos compara-se os níveis salariais dos empregados com carteira com aqueles dos empregados sem carteira e dos trabalhadores por conta própria respectivamente.⁶ O conjunto das variáveis de controle e a ordem de entrada destas variáveis na análise são idênticas nos dois exemplos. As variáveis de controle, por ordem de entrada na análise são: sexo, região de residência, condição na família e ocupação. Uma definição precisa das variáveis independentes utilizadas é apresentada na Tabela 1.

⁶O salário medido em salários mínimos foi obtido da seguinte forma:

$$S = (R / 333.120) / (H / 48)$$

onde:

S — salário em salários mínimos

R — rendimentos mensais em dinheiro e em mercadorias normalmente recebidos na ocupação principal

H — número de horas normalmente trabalhadas por semana na ocupação principal

Cz\$ 333.120 — valor do salário mínimo vigente em todo o território nacional no período da pesquisa da PNAD.

Tabela 1
Variáveis utilizadas

Variável	Símbolo	Valor = 0	Valor = 1
Posição na ocupação I	b	Empregado sem carteira	Empregado com carteira
Posição na ocupação II	b	Trabalhador por conta própria	Empregado com carteira
Sexo	a ₁	Mulher [M]	Homem [H]
Região de residência	a ₂	Rural [R]	Urbana [U]
Condição na família	a ₃	Não chefe [NC]	Chefe [C]
Ocupação	a ₄	Ligado à produção [LP]	Não ligado à produção [$\bar{N}LP$] ¹

Nota: 1) Foram classificados como ocupações não ligadas à produção as ocupações técnicas, científicas e assemelhadas, auxiliares técnico-científicas, e administrativas de acordo com os conceitos e definições de ocupação utilizados no Censo Demográfico de 1980.

Tabela 2
Diferenciais de salário e variações nos diferenciais de salário
segundo o número de variáveis de controle utilizadas
(salários mínimos)

Número de Variáveis de Controle (k)	Empregados com Carteira versus Empregados sem Carteira		Empregados com Carteira versus Trabalhadores por conta própria	
	Diferencial (B _k)	Variação (D _k)	Diferencial (B _k)	Variação (D _k)
0	2,59	0,05	1,34	-0,03
1	2,54	0,35	1,37	0,62
2	2,19	0,26	0,75	-0,14
3	1,93	0,75	0,88	1,68
4	1,19	—	-0,80	—

A base empírica deriva-se da Pesquisa Nacional por Amostra de Domicílios-85. Na análise apenas foram incluídas (1) as pessoas ocupadas,⁷ (2) as que declararam sua renda e jornada de trabalho na ocupação principal, (3) as que não possuem ocupação na administração pública,⁸ (4) as que declararam sua posição na ocupação como sendo empregado ou trabalhador por conta própria,⁹ (5) e as que em caso de serem empregados declararam se possuíam ou não carteira de trabalho assinada. Passaram por estes filtros e, portanto, foram utilizados na análise, 79.115 empregados com carteira, 58.551 empregados sem carteira e 47.440 trabalhadores por conta própria.

O Diagrama 1 apresenta as médias salariais para os três grupos: empregados com carteira, empregados sem carteira e trabalhadores por conta própria. Para cada uma das comparações, obteve-se valores para B_k e D_k , $P[.]$, $B_k^*(.)$, $D_k^*(.)$, $\bar{C}_k(.)$ e $r_k(.)$, veja Tabela 2 e Diagramas 2 a 6.

Os valores obtidos para B_k e $B_k^*(.)$ (Tabela 2 e Diagrama 3), indicam que o salário médio dos empregados com carteira é nitidamente superior ao dos empregados sem carteira. Comparando empregados com carteira e trabalhadores por conta própria também se observa que o salário dos empregados com carteira é mais elevado, com exceção do caso em que controla-se por ocupação ($k = 4$).

Como o objetivo principal deste estudo é analisar a sensibilidade dos diferenciais de salário em relação ao número de variáveis de controle incluídas, é fundamental observar o comportamento de B_k e D_k , quando k varia de 0 a 4. Conforme demonstrado, as variações de B_k e

⁷ São consideradas pessoas ocupadas aquelas que tinham trabalho na semana de referência da pesquisa. Estas pessoas podem ter realmente trabalhado ou não durante a semana de referência. Inclui-se, portanto, como ocupadas pessoas em férias, de doença, etc. . .

⁸ Especificamente, eliminou-se da amostra aqueles cujo ramo de atividade era a administração pública, além daqueles cuja ocupação era alguma das seguintes: ocupações ligadas à defesa nacional e segurança pública, os técnicos e fiscais de tributação e arrecadação, os inspetores do trabalho, os magistrados e, os procuradores, promotores e curadores públicos. Estes trabalhadores foram excluídos porque, apesar de, em geral, não possuírem carteira de trabalho assinada, a determinação de seus salários se assemelha à determinação dos salários dos empregados com carteira.

⁹ A categoria "empregados" é entendida como incluindo os membros da força de trabalho que declararam sua posição na ocupação como sendo empregados, parceiros empregados ou trabalhadores agrícolas volantes. A categoria "conta própria" é definida como aqueles que declararam sua posição na ocupação como conta própria ou parceiros conta própria.

¹⁰ Os diagramas referentes às mulheres foram omitidos, mas podem ser obtidos com os autores.

D_k são determinadas a partir dos valores para $\bar{C}_k (.)$ e $r_k (.)$. Uma vez que os salários tendem a ser maiores: (1) entre os homens, (2) na área urbana, (3) entre aqueles que são chefes de famílias, e (4) nas ocupações não ligadas diretamente à produção, tem-se que $\bar{C}_k (.) > 0$ em todas as células. Deste fato decorre então que o sinal de $D_k^* (.)$ é determinado pelo sinal de $r_k (.)$. O comportamento de $r_k (.)$ é diferente nas duas ilustrações realizadas, portanto deve considerá-las separadamente.

3.1 – Empregados com carteira versus empregados sem carteira

Nesta comparação, $r_k (.) > 0$, $0 \leq k < 4$, isto é, a incidência da carteira é maior (1) entre os homens, (2) na área urbana, (3) entre aqueles que são chefes de família, e (4) nas ocupações não ligadas diretamente à produção. Portanto, a incidência da carteira tende a ser maior exatamente nos grupos que possuem mais altos salários. Decorre desta conjunção de fatores que a introdução de qualquer uma das variáveis de controle tende sempre a reduzir o diferencial estimado, ou seja, $B_0 > B_1 > B_2 > B_3 > B_4$. Em particular, nota-se que a introdução da variável sexo reduz o diferencial estimado em $B_0 - B_1 = 0,04$ salários mínimos, que equivale a $r_0 = 8,9\%$, do diferencial de salário entre homens e mulheres, $\bar{C}_0 = 0,54$ salários mínimos.

3.2 – Empregados com carteira versus trabalhadores por conta própria

Nesta comparação, os diferenciais estimados, B_k , $0 \leq k \leq 4$, não decrescem monotonicamente (Tabela 2). Esta característica decorre do fato de que a incidência relativa de empregados com carteira não é mais sistematicamente maior nos subgrupos com salários mais elevados. De fato, a participação de trabalhadores por conta própria é relativamente maior entre os homens do que entre as mulheres, assim como maior entre os chefes de família do que entre aqueles que não são chefes. Conseqüentemente, os diferenciais de salários estimados tendem a elevar-se quando as variáveis que indicam sexo ou condição na família são introduzidas na análise, isto é, $D_k < 0$ para $k = 0, 2$ (Tabela 2). Em particular, quando o controle por sexo é introduzido, o diferencial médio estimado se eleva de $B_1 - B_0 = -D_1 = 0,03$ salários mínimos que correspondem a $-r_0 = 2,7\%$ do diferencial de salários entre homens e mulheres, $\bar{C}_0 = 1,21$ salários mínimos. Note que esta comparação ilustra o fato de que a introdução de uma variável de controle pode tanto reduzir como elevar os diferenciais de salário estimados.

4. Conclusões

Verifica-se nesta nota como a estimação de diferenciais de salário é sensível às variáveis de controle utilizadas. Parece haver uma expectativa entre diversos investigadores de que à medida que o número de variáveis de controle aumenta os diferenciais estimados devam diminuir. Entretanto, esta expectativa se cumpre somente quando (1) o nível da variável de controle a ser incluída é maior (menor) entre aqueles que possuem o atributo cujo efeito sobre o nível salarial se deseja estimar e (2) esta variável de controle tende a elevar (reduzir) o nível salarial. Caso contrário a inclusão da variável de controle tenderá a elevar, ao invés de reduzir, os diferenciais de salário inicialmente estimados.

O comportamento dos diferenciais de salário entre empregados com e sem carteira segue o padrão esperado. À medida que variáveis de controle adicionais são incluídas na análise, os diferenciais de salário se reduzem. A comparação entre empregados com carteira e trabalhadores por conta própria oferece um interessante contra-exemplo a esta norma. Neste caso, em duas oportunidades a inclusão de variáveis de controle (sexo e posição na família) levaram a uma elevação nos diferenciais de salário.

Há ainda diversas outras conseqüências que se seguem dos resultados obtidos neste estudo: Por exemplo, uma variável de controle só afetará o estimador do efeito de um dado atributo sobre o salário caso a variável de controle considerada esteja relacionada simultaneamente com o nível salarial e a incidência do atributo cujo efeito sobre salário se pretende avaliar. É possível também predizer a magnitude e direção da sensibilidade dos diferenciais de salário a partir das propriedades da variável de controle a ser incluída. Na ilustração sobre o impacto da carteira de trabalho demonstra-se que o controle por sexo reduz o diferencial médio de salário entre empregados com carteira e sem carteira em 0,04 salários mínimos. Os resultados obtidos nesta nota mostram que esta superestimação do diferencial de salário que se obtém quando não se controla por sexo deve-se a dois fatos: primeiro, a incidência da carteira ser 9% maior entre os homens do que entre as mulheres e segundo, o salário entre os homens tende a ser 0,54 salários mínimos mais elevado entre os homens do que entre as mulheres.

APÊNDICE

Neste apêndice constam as demonstrações dos teoremas apresentados na seção 2.

Proposição 1:

$$D_k = \sum_{p_k} D_k^*(p_k) \cdot P[A_k = p_k] \quad (1)$$

onde:

$$D_k^*(p_k) \equiv B_k^*(p_k) - \bar{B}_k(p_k)$$

e

$$\begin{aligned} \bar{B}_k(p_k) &\equiv B_{k+1}^*(p_k, 0) \cdot P[a_{k+1} = 0 \mid A_k = p_k] + \\ &B_{k+1}^*(p_k, 1) \cdot P[a_{k+1} = 1 \mid A_k = p_k] \end{aligned}$$

Demonstração:

Dado que $P_{k+1} = P_k \times \{0, 1\}$ tem-se que

$$\begin{aligned} B_{k+1} &= \sum_{p_k} B_{k+1}^*(p_k, 0) \cdot P[A_k = p_k, a_{k+1} = 0] + \\ &\sum_{p_k} B_{k+1}^*(p_k, 1) \cdot P[A_k = p_k, a_{k+1} = 1] = \\ &\sum_{p_k} \{ B_{k+1}^*(p_k, 0) \cdot P[a_{k+1} = 0 \mid A_k = p_k] + \\ &+ B_{k+1}^*(p_k, 1) \cdot P[a_{k+1} = 1 \mid A_k = p_k] \} \cdot P[A_k = p_k] \end{aligned}$$

Logo

$$B_{k+1} = \sum_{p_k} \bar{B}_k(p_k) \cdot P[A_k = p_k]$$

e

$$D_k \equiv B_k - B_{k+1} = \sum_{p_k} \{ B_k^*(p_k) - \bar{B}_k(p_k) \} \cdot P[A_k = p_k]$$

Proposição 2:

$$D_k^*(p_k) = r_k(p_k) \cdot \bar{C}_k(p_k) \quad (2)$$

onde:

$$r_k(p_k) \equiv P[a_{k+1} = 1 | b = 1, A_k = p_k] - \\ P[a_{k+1} = 1 | b = 0, A_k = p_k], \\ C_k^0(p_k) \equiv E[Y | b = 0, A_k = p_k, a_{k+1} = 1] - \\ E[Y | b = 0, A_k = p_k, a_{k+1} = 0].$$

e

$$C_k^1(p_k) \equiv E[Y | b = 1, A_k = p_k, a_{k+1} = 1] - \\ E[Y | b = 1, A_k = p_k, a_{k+1} = 0].$$

Demonstração:

Seja

$$M_{ij} \equiv E[Y | b = i, A_k = p_k, a_{k+1} = j] \quad i, j \in \{0, 1\}, \\ q_i P[a_{k+1} = 1 | b = i, A_k = p_k] \quad i \in \{0, 1\},$$

e

$$\alpha \equiv P[b = 1 | A_k = p_k].$$

Portanto,

$$C_k^0(p_k) = M_{01} - M_{00}, \quad (3)$$

$$C_k^1(p_k) = M_{11} - M_{10}, \quad (4)$$

$$\bar{C}(p_k) = (M_{01} - M_{00}) \cdot \alpha + (M_{11} - M_{10}) \cdot (1 - \alpha), \quad (5)$$

e

$$r_k(p_k) = q_1 - q_0. \quad (6)$$

Note que

$$\begin{aligned}
B_k^*(p_k) &\equiv E[Y \mid b=1, A_k=p_k] - E[Y \mid b=0, A_k=p_k] = \\
&= E[Y \mid b=1, A_k=p_k, a_{k+1}=1] \cdot P[a_{k+1}=1 \mid b=1, A_k=p_k] + \\
&\quad E[Y \mid b=1, A_k=p_k, a_{k+1}=0] \cdot P[a_{k+1}=0 \mid b=1, A_k=p_k] - \\
&\quad \{E[Y \mid b=0, A_k=p_k, a_{k+1}=1] \cdot P[a_{k+1}=1 \mid b=0, A_k=p_k] + \\
&\quad E[Y \mid b=0, A_k=p_k, a_{k+1}=0] \cdot P[a_{k+1}=0 \mid b=0, A_k=p_k]\} = \\
&= M_{11} \cdot q_1 + M_{10} \cdot (1 - q_1) - \{M_{01} \cdot q_0 + M_{00} \cdot (1 - q_0)\}.
\end{aligned}$$

Além disso, demonstrando (5)

$$\begin{aligned}
\bar{B}_k(p_k) &\equiv B_{k+1}^*(p_k, 0) \cdot P[a_{k+1}=0 \mid A_k=p_k] + \\
&\quad + B_{k+1}^*(p_k, 1) \cdot P[a_{k+1}=1 \mid A_k=p_k] = \\
&= (M_{10} - M_{00}) \cdot P[a_{k+1}=0 \mid A_k=p_k] + \\
&\quad + (M_{11} - M_{01}) \cdot P[a_{k+1}=1 \mid A_k=p_k].
\end{aligned}$$

Tem-se também que

$$\begin{aligned}
P[a_{k+1}=0 \mid A_k=p_k] &= \\
&= P[a_{k+1}=0 \mid b=0, A_k=p_k] \cdot P[b=0 \mid A_k=p_k] + \\
&\quad + P[a_{k+1}=0 \mid b=1, A_k=p_k] \cdot P[b=1 \mid A_k=p_k] = \\
&= (1 - q_0) \cdot (1 - \alpha) + (1 - q_1) \cdot \alpha.
\end{aligned}$$

Analogamente, demonstrando (6)

$$P[a_{k+1}=1 \mid A_k=p_k] = q_0 \cdot (1 - \alpha) + q_1 \cdot \alpha.$$

Logo,

$$\begin{aligned}
\bar{B}_k(p_k) &= (M_{10} - M_{00}) \left\{ (1 - q_0) \cdot (1 - \alpha) + (1 - q_1) \cdot \alpha \right\} + \\
&\quad (M_{11} - M_{01}) \cdot \left\{ q_0 \cdot (1 - \alpha) + q_1 \cdot \alpha \right\}.
\end{aligned}$$

Conseqüentemente,

$$\begin{aligned} D_k^*(p_k) &\equiv B_k^*(p_k) - \bar{B}_k(p_k) = M_{11} \cdot (q_1 - q_0) \cdot (1 - \alpha) - \\ &M_{10} \cdot (q_1 - q_0) \cdot (1 - \alpha) + M_{01} \cdot (q_1 - q_0) \cdot \alpha - M_{00} \cdot (q_1 - q_0) \cdot \alpha = \\ &= (q_1 - q_0) \cdot \left\{ M_{01} - M_{00} \right\} \cdot \alpha + (M_{11} - M_{10}) \cdot (1 - \alpha) \Big\}. \end{aligned}$$

Finalmente (2) segue de (5) e (6).

Proposição 3:

$$r_k(p_k) = \frac{\text{COV}[b, a_{k+1} \mid A_k = p_k]}{\text{VAR}[b \mid A_k = p_k]}$$

Demonstração:

$$\begin{aligned} \text{COV}[b, a_{k+1} \mid A_k = p_k] &= P[b = 1, a_{k+1} = 1 \mid A_k = p_k] - \\ &P[b = 1 \mid A_k = p_k] \cdot P[a_{k+1} = 1 \mid A_k = p_k] = \\ &= P[a_{k+1} = 1 \mid b = 1, A_k = p_k] \cdot P[b = 1 \mid A_k = p_k] - \\ &P[b = 1 \mid A_k = p_k] \cdot \left\{ P[a_{k+1} = 1 \mid b = 1, A_k = p_k] \cdot \right. \\ &\left. \cdot P[b = 1 \mid A_k = p_k] + P[a_{k+1} = 1 \mid b = 0, A_k = p_k] \cdot \right. \\ &\left. \cdot P[b = 0 \mid A_k = p_k] \right\} = q_1 \cdot \alpha - \left\{ q_1 \cdot \alpha + q_0 \cdot (1 - \alpha) \right\} \alpha = \\ &= \alpha \cdot (1 - \alpha) \cdot (q_1 - q_0). \end{aligned}$$

Além disso,

$$\begin{aligned} \text{VAR}[b \mid A_k = p_k] &= P[b = 1 \mid A_k = p_k] \cdot P[b = 0 \mid A_k = p_k] = \\ &= \alpha \cdot (1 - \alpha). \end{aligned}$$

Logo

$$\frac{\text{COV}[b, a_{k+1} \mid A_k = p_k]}{\text{VAR}[b \mid A_k = p_k]} = q_1 - q_0$$

Proposição 4:

$$q_k^i(p_k) = \frac{\text{COV}[Y, a_{k+1} \mid b=i, A_k=p_k]}{\text{VAR}[a_{k+1} \mid b=i, A_k=p_k]} \quad i \in \{0, 1\} \quad (8)$$

Demonstração:

$$\begin{aligned} \text{COV}[Y, a_{k+1} \mid b=i, A_k=p_k] &= E[Y \cdot a_{k+1} \mid b=i, A_k=p_k] - \\ &\quad E[Y \mid b=i, A_k=p_k] \cdot E[a_{k+1} \mid b=i, A_k=p_k] = \\ &= E[Y \mid b=i, A_k=p_k, a_{k+1}=1] \cdot P[a_{k+1}=1 \mid b=i, A_k=p_k] - \\ &\quad \{E[Y \mid b=i, A_k=p_k, a_{k+1}=1] \cdot P[a_{k+1}=1 \mid b=i, A_k=p_k] + \\ &\quad E[Y \mid b=i, A_k=p_k, a_{k+1}=0] \cdot P[a_{k+1}=0 \mid b=i, A_k=p_k]\} \cdot \\ &\quad P[a_{k+1}=1 \mid b=i, A_k=p_k] = \\ &= M_{i1} \cdot q_i - \{M_{i1} \cdot q_i + M_{i0} \cdot (1 - q_i)\} \cdot q_i \\ &= (M_{i1} - M_{i0}) \cdot q_i \cdot (1 - q_i). \end{aligned}$$

Como

$$\begin{aligned} \text{VAR}[a_{k+1} \mid b=i, A_k=p_k] &= P[a_{k+1}=1 \mid b=i, A_k=p_k] \cdot \\ &\quad P[a_{k+1}=0 \mid b=i, A_k=p_k] = q_i \cdot (1 - q_i). \end{aligned}$$

Então (8) segue de (3) e (4).

Bibliografia

- ANDERSON, Sharon et alii. **Statistical methods for comparative studies**. New York, John Wiley & Sons, 1980.
- BARROS, Ricardo Paes de & VARANDAS, Simone. A carteira de trabalho e as condições de trabalho e remuneração dos chefes de família no Brasil. **Revista da ANPEC**, Rio de Janeiro, 10 (12): 15-9, dez. 1987.

- CAMARGO, José Márcio & SERRANO, Franklin. Os dois mercados: homens e mulheres na indústria brasileira. **R. Bras. Econ.**, Rio de Janeiro, 37 (4): 435-48, out./dez. 1983.
- CUNHA, Paulo Vieira da & BONELLI, Regis. Estrutura de salários industriais no Brasil: um estudo sobre a distribuição de salários médios em 1970. **Pesq. e Plan. Econ.**, Rio de Janeiro, 8 (1): 117-68, abr. 1978.
- JUDD, Charles M. & KENNY, David. A. **Estimating the effects of social interventions**. Cambridge, Cambridge University Press, 1981.
- MACEDO, Roberto Brás Matos. Os salários nas empresas estatais. São Paulo, Nobel; Brasília, Ministério do Trabalho, 97p. 1985.
- RESENDE, Gervásio Castro. "Interação entre mercados de trabalho e razão entre salários rurais e urbanos no Brasil". **Estudos Econômicos**, vol. 15, n. 1: 47-67, jan./abr. 1985.