

## UM MODELO ESCOLHA DISCRETA PARA AVALIAÇÃO DE DESEMPENHO EM CONCURSOS PÚBLICOS.

**Ivan Castelar** - Pós-Doutorado - University of Aberdeen (GB); Professor Titular do Depto. de Economia Aplicada(UFC).

**Alexandre Weber Aragão** - Mestre em economia (CAEN/UFC) / Assessor da Secretaria das Cidades do Estado do Ceará.

**Roberto Tatiwa Ferreira** - Doutorado pelo CAEN/UFC; Professor do Curso de Pós-Graduação em Economia (CAEN/UFC)

**Ilton Soares** - Mestrando- EPGE-FGV.

### Resumo:

A principal preocupação deste trabalho foi a de investigar os determinantes da aprovação de um candidato num concurso público. Para tanto usou-se um estudo de caso com base em dados e resultados de um concurso para provimento de cargo administrativo de nível médio para o Banco do Nordeste do Brasil, do qual participaram 232.771 candidatos. Utilizou-se o modelo *logit* de escolha binária com estimação baseada no método de máxima verossimilhança. Nos testes de significância estatística de cada um dos parâmetros estimados somente uma variável, referente ao estado civil, não se mostrou estatisticamente significativa. As que se mostraram significativas apresentaram os sinais esperados quanto ao efeito marginal sobre a chance de aprovação. Quanto à magnitude dos coeficientes, o dado relevante é que a condição do candidato ter o grau de escolaridade requerido para a função a ser exercida (nível médio) foi o que apresentou maior impacto (negativo) na chance do candidato ser aprovado. As principais conclusões do trabalho apontam o ensino médio público como um possível *locus* de ação e políticas que atuem para reduzir as desigualdades de condições iniciais de cada candidato a concursos públicos. Por outro lado, o preenchimento de postos de trabalho que requerem apenas uma educação de nível médio com detentores de título universitário, e até mesmo com pós-graduação, aponta para uma má alocação dos recursos públicos conjugada com uma possível deficiência estrutural do mercado de trabalho regional.

**Palavras Chaves:** concurso público; avaliação; logit; probit.

### Abstract:

The main objective of this thesis was to investigate the factors determining success in an aptitude test for admission in a public bank in the Northeast of Brazil. The data set is based in a questionnaire for 232,771 Candidates which is composed of the result of the test and a number of questions about some attributes of the candidate. Attributes selected were personal monthly income; household monthly income; education level; type of school attended (public or private); marital status; occupation and age. A *logit* model was used to estimate a regression of success in the test against the attributes defined above. Results of the estimated model show that all the coefficients but marital status were significant and with the expected sign. The marginal effects inform that a high school level (a requirement to take the test) had the strongest and negative

impact on the probability of success. Specification tests for functional form and unobserved heterogeneity show that the model was correctly specified.

**Key Words:** test; public sector admission; logit; probit

## 1. INTRODUÇÃO

A instituição do concurso público para provimento de cargos do aparelho do Estado Brasileiro tem sido um forte atrator de recursos humanos desde meados do século passado. Assim, uma elite egressa da classe média projetou-se na conformação da burocracia estatal através do concurso público.

A razão para esta atração pode ser explicada por alguns fatores. Um deles é o de que o setor público, em muitos níveis e regiões, vem pagando salários melhores do que os praticados na iniciativa privada. Existem ainda, atrelados ao emprego público, benefícios indiretos tais como planos de saúde e previdência diferenciada.

Outra explicação para esta atração pode ser o fato de que estar alocado no setor público proporciona um grau de segurança maior quanto à permanência no emprego, dada a legislação que garante certa estabilidade ao funcionário público, o que não se observa no setor privado. Ou mesmo a vontade das pessoas em seguir uma carreira pública, pois é característico destas funções a existência de planos de cargos e salários diferenciados para alguns concursos.

Portanto, a questão primordial a ser colocada é: o que faz determinadas pessoas serem aprovadas em detrimento de outras? Será que existem fatores socioeconômicos que aumentam as chances de uma pessoa ter um melhor desempenho? Outras questões subseqüentes a esta são: maior renda implica em maiores chances de passar?; Ter nível de educação acima do exigido no concurso implica em maiores chances de passar? Assim, o objetivo principal deste trabalho é investigar os determinantes da aprovação em um determinado concurso público.

Para tanto, fez-se um estudo de caso com base em dados e resultados de um concurso público para provimento de cargo administrativo de nível médio para o Banco do Nordeste do Brasil (BNB), realizado no ano de 2003.

Quanto ao caso estudado nesse trabalho, tem-se a particularidade de ter sido um concurso para uma instituição localizada em uma região de menor renda, quando comparada às regiões Sudeste e Sul do Brasil. Isso reflete ainda mais a necessidade de se entender o que explica a aprovação de um candidato. Pois, de forma patente, a região Nordeste se ressentia de ofertas de vagas de trabalho atrativas. Assim, por exemplo, se os candidatos provenientes de famílias de melhor estrutura financeira tiverem maiores chances de serem aprovados, poderá existir um círculo vicioso de concentração de renda.

Na seqüência do trabalho estas questões serão devidamente desenvolvidas seguindo a estrutura descrita adiante. Além desta introdução, na seção 2 expõe-se a metodologia a ser adotada para buscar a constatação empírica dos determinantes da aprovação do candidato no concurso. Nesta metodologia discutir-se-á, particularmente, um modelo microeconômico de escolha e as técnicas econométricas de estimação de modelos de escolha discreta e sua pertinência a este estudo. Na seção 3 são apresentados os resultados empíricos e, finalmente, na seção 4 são apresentadas as considerações finais.

## 2. METODOLOGIA

### 2.1. O MODELO TEÓRICO

O modelo teórico utilizado neste trabalho segue aquele proposto por Fair (1978), o qual é bastante flexível para se adequar a diferentes tipos de escolha. Considerando um indivíduo que está estudando para obter aprovação em um concurso público, este agente econômico tem que alocar o seu tempo em três atividades diferentes; a saber, lazer ( $t_1$ ), estudo ( $t_2$ ) e trabalho ( $t_3$ ). Definindo  $U$  como a utilidade total do indivíduo tal que,

$$(1) U = U_1 + U_2$$

Onde,

$$(2) U_1 = f(t_1, X_1, E_1)$$

$$(3) U_2 = f(t_2, X_2, E_2)$$

A utilidade em  $U_1$  é função do tempo gasto em lazer ( $t_1$ ), das unidades de bens consumidas com essa escolha por lazer ( $X_1$ ) e do vetor de outras variáveis ( $E_1$ ).  $U_2$ , por sua vez, depende do tempo gasto com estudo ( $t_2$ ), dos bens consumidos com essa escolha ( $X_2$ ) e do vetor de outras variáveis.

Sendo  $T$  o tempo disponível total do indivíduo; isto é,  $T = t_1 + t_2 + t_3$ ;  $p$  o preço dos bens;  $W$  a taxa salarial e  $V$  a renda que independe do trabalho. A restrição orçamentária individual pode ser formada como,

$$(4) W(T - t_1 - t_2) + V = p(X_1 + X_2)$$

Desta forma o problema individual de maximização de utilidade sujeita a restrição orçamentária é dada pelo seguinte Lagrangeano,

$$(5) L = U + I[W(T - t_1 - t_2) + V - p(X_1 + X_2)]$$

As condições de primeira ordem desse problema de maximização são listadas abaixo.

$$(6a) \frac{\partial f}{\partial t_1} - IW = 0$$

$$(6b) \frac{\partial g}{\partial t_2} - IW$$

$$(6c) \frac{\partial f}{\partial X_1} - Ip = 0$$

$$(6d) \frac{\partial f}{\partial X_2} - Ip = 0$$

$$(6e) W(T - t_1 - t_2) + V - p(X_1 + X_2) = 0$$

Considerando que as primeiras derivadas de  $f$  e  $g$  são positivas, bem como as derivadas parciais cruzadas e, que as derivadas segundas de  $f$  e  $g$  são negativas, tem-se que:

$$(7) \frac{dt_2}{dV} > 0$$

Este resultado de estática comparativa, mostra que um aumento na renda que independe do trabalho, permite que o indivíduo trabalhe menos e aumente o seu tempo dedicado ao estudo. Portanto, nestas condições, o indivíduo terá maior chance de ser aprovado no concurso público.

Outro resultado interessante para este estudo, relaciona-se aos efeitos de  $E_1$  em  $t_2$ ,

$$(8) \frac{dt_2}{dE_1} < 0$$

Desta forma, qualquer variável que tenha um efeito positivo na utilidade do lazer  $U_1$ , afeta negativamente a escolha por tempo de estudo, o que reduz as chances de aprovação do indivíduo.

Continuando a análise de estática comparativa, os seguintes efeitos apresentam ambiguidades em seus sinais,

$$(9) \frac{dt_2}{dW}; \frac{dt_2}{dp} \text{ e } \frac{dt_2}{dE_2}$$

Em relação à  $\frac{dt_2}{dE_2}$ , espera-se que este efeito seja positivo, pois variáveis que afetam positivamente  $U_2$ , devem aumentar a escolha por  $t_2$ .

No que se refere a base de dados utilizada neste estudo, o tempo dedicado ao lazer e ao estudo são variáveis não observadas. Observa-se apenas se o indivíduo passou ou não. Considerando que  $Y=1$  quando o indivíduo é aprovado no concurso e que  $U_2 > U_1$  e  $Y=0$  caso contrário. Considerando ainda que,

$$(10) U_2 = x'b_2 + e_2$$

$$(11) U_1 = x'b_1 + e_1$$

Portanto,

$$(12) \text{Pr ob}[Y = 1 | x] = \text{Pr ob}[U_2 > U_1 | x] = \text{Pr ob}[e < x'b | x]$$

Onde,

$$(13) \mathbf{b} = \mathbf{b}_2 - \mathbf{b}_1 \text{ e } \mathbf{e} = \mathbf{e}_2 - \mathbf{e}_1$$

## 2.2. Especificação do Modelo Econométrico

### 2.2.1 O MODELO DE REGRESSÃO

Os resultados empíricos deste trabalho serão obtidos a partir da estimação de modelos probit e logit. No modelo probit, a probabilidade de se observar uma aprovação ( $Y=1$ ), condicionada a um vetor de variáveis explicativas possui uma distribuição normal padrão; ou seja,

$$(14) \text{Pr ob}(Y = 1 | x) = \int_{-\infty}^{x'b} f(t) dt = \Phi(x'b)$$

No caso do logit, a distribuição normal é substituída pela distribuição logística,

$$(15) \text{Pr ob}(Y = 1 | x) = \frac{e^{x'b}}{1 + e^{x'b}} = \Lambda(x'b)$$

Quanto à questão de qual dessas distribuições seria a mais adequada, a distribuição logística é bastante similar à distribuição normal, sendo aquela um pouco mais densas nas caudas (se aproximando de uma  $t$  com 7 graus de liberdade). Portanto, para valores intermediários de  $x'b$ , as duas distribuições tendem a fornecer probabilidades similares.

A distribuição logística tende a dar probabilidades maiores para  $Y=0$  quando  $x'b$  é extremamente pequeno (e probabilidades menores para  $Y=0$  quando  $x'b$  é muito grande) em relação a distribuição normal.

Em muitas aplicações, não parece fazer muita diferença a utilização de um modelo ou outro. Mas, como em certas aplicações estas diferenças são perceptíveis, discute-se adiante uma maneira mais segura de optar entre um destes dois modelos e até mesmo de considerar uma terceira alternativas.

Para se obter os efeitos marginais de cada variável nos modelos probit, deve-se proceder da seguinte maneira,

$$(16) \frac{\partial E[y | x]}{\partial x} = f(x'b)b$$

Nos modelos logit, estes efeitos são obtidos através do cálculo abaixo,

$$(17) \frac{\partial E[y | x]}{\partial x} = \Lambda(b \cdot x)[1 - \Lambda(b \cdot x)]b$$

### 2.2.2. ESTIMAÇÃO E INFERÊNCIA EM MODELOS DE ESCOLHA BINÁRIA

A estimação dos modelos probit e logit se baseia, usualmente, no método de máxima verossimilhança, onde cada observação é tratada como um único resultado de uma distribuição de *Beunoulli*.

Supondo que as observações são independentes e reescrevendo (12) como,

$$(18) \text{Pr ob}(Y = 1 | x) = \text{Pr ob}[e < x' \mathbf{b} | x] = F(x' \mathbf{b})$$

Então,

$$(19) \text{Pr ob}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{y_i=0} [1 - F(x_i' \mathbf{b})] \prod_{y_i=1} F(x_i' \mathbf{b})$$

Ou ainda,

$$(20) L = \prod_{i=1}^n [F(x_i' \mathbf{b})]^{y_i} [1 - F(x_i' \mathbf{b})]^{1-y_i}$$

Aplicando o logaritmo em (20), obtém-se a função de verossimilhança logaritimizada,

$$(21) \ln L = \sum_{i=1}^n [y_i \ln F(x_i' \mathbf{b}) + (1 - y_i) \ln(1 - F(x_i' \mathbf{b}))]$$

Greene (1993) apresenta os detalhes do processo de maximização da função de verossimilhança e, portanto, dos procedimentos de estimação dos parâmetros dos modelos logit e probit. Há problemas em modelos de escolha binária semelhantes aos encontrados em modelos de regressões em geral. No entanto, estes problemas podem, em alguns casos, ter conseqüências mais graves com escolha discreta do que em modelos tradicionais.

Particularmente, existem quatro tipos básicos de erros de especificação que podem ocorrer neste tipo de modelo. O primeiro é o erro de especificação sobre a presença ou omissão de variáveis explicativas, no qual se inclui o problema da heterogeneidade não observada. O segundo tipo de erro é sobre a estrutura de variância da disturbância; a qual, em princípio, pode ser heterocedástica. O terceiro tipo de erro é sobre a determinação simultânea das variáveis do modelo; ou seja, variáveis do lado direito da equação requerem equações adicionais para explicá-las. Finalmente, o quarto tipo de erro é sobre a especificação funcional do modelo; ou seja, a função de probabilidade escolhida não é correta.

O problema de heterogeneidade não observada pode ser melhor ilustrado por uma regressão de dados em painel do tipo:

$$(22) Y_{it} = a + bX_{it} + C_i + e_{it}$$

Onde  $C_i$  é uma variável não observada. Em equações de determinação de salário, por exemplo,  $X$  é um vetor de atributos do trabalhador e  $C_i$  é um efeito específico da unidade observacional, diferenciando-a das demais.  $C_i$  pode representar a habilidade individual de cada trabalhador. No entanto, se o vetor  $X$  contém variáveis de escolaridade é muito provável que  $C_i$  seja correlacionado com  $X$ , levando a estimativas inconsistentes dos parâmetros do modelo (Hsiao, 2003). Em dados em painel esta questão é subjacente à escolha entre tratamento de efeitos fixos e tratamento de efeitos aleatórios.

No modelo a ser especificado mais adiante, onde o sucesso em um concurso é modelado como função de um vetor de atributos, o qual inclui variáveis de grau de instrução, não é possível observar a habilidade específica de cada candidato. Mesmo se fosse possível observá-la seria difícil mensurá-la, embora o Quociente de Inteligência (QI) seja uma boa *proxy* de habilidade.

Em modelos de dados em painel o teste de *Hausman* é a ferramenta mais usual para discriminar se a heterogeneidade é correlacionada com o vetor  $X$ . Com dados transversais é impossível trabalhar-se com uma modelagem de efeitos fixos se as diferenças individuais são correlacionadas com as demais variáveis independentes, pois o teste de *Hausman* é baseado em contrastes de MQO e MQG.

No entanto, como heterogeneidade não observada cai na categoria mais geral de erro de especificação, existe uma estratégia para testar a sua existência ou, mais importante, testar se a sua ocorrência é danosa para a estimação. *White* (1980) alerta que o seu teste de heterocedasticidade é também um teste para erro de especificação.

Um outro tipo de erro de especificação é a possibilidade de endogeneidade das variáveis do lado direito do modelo. No modelo a ser definido posteriormente, a renda do candidato, a renda da família, o nível educacional e o estado civil podem ser simultaneamente determinados. A questão, no entanto, é inteiramente similar a da heterogeneidade negligenciada. Ou seja, o problema fundamental a ser observado é se os erros das equações auxiliares de determinação destas variáveis são correlacionadas ou não com o erro do modelo originalmente proposto (*Sargan*, 1975). Necessita-se, portanto, de um teste de exogeneidade; neste caso, um teste de *River-Wuong* (*Wooldridge*, 2002).

Infelizmente, para este teste precisa-se de instrumentos, os quais inexistem para este problema específico. No questionário base apenas a variável tipo de lazer preferido foi excluída do modelo, e esta nem é apropriada nem suficiente para fazer as vezes de instrumento neste tipo de estimação.

Na ausência de instrumentos para testar a endogeneidade e, caso esta se confirme, proceder a uma estimação simultânea, não se pode estar plenamente confiante sobre a consistência dos estimadores. Resta, no entanto, uma argumentação razoável sobre a possível ausência de correlação entre os erros de um modelo de determinação simultânea das variáveis a serem tratadas mais adiante.

Existe a possibilidade de que a renda familiar, a renda do indivíduo, o grau de instrução e até mesmo o estado civil, sejam simultaneamente determinados. Contudo, é extremamente improvável que os erros das equações que determinam estas variáveis sejam correlacionados com o erro da equação determinante do sucesso no concurso público. A razão para isto é que cada variável citada acima é determinada em diferentes mercados, como é o caso das rendas; ou por diferentes mecanismos de formação, como são os casos de grau de instrução e estado civil. Portanto, o que se presume, é que exista alguma correlação entre estas variáveis, mas não a existência de vies de simultaneidade.

O último tipo de erro de especificação a ser discutido é o do formato da escolha funcional. Neste estudo, este problema se traduz na escolha entre os modelos probit e logit. Embora estes modelos forneçam resultados muito similares, nem por isso deve-se negligenciar um teste sobre a escolha funcional, principalmente pelo motivo da amostra nesse problema apresentar um número de valores nulos de  $Y$  excessivamente grande com relação ao número de 1's. Discriminar entre estes dois formatos requer um teste na categoria de modelos não entrelaçados. *Silva* (2001) mostra que, sob condições adequadas de regularidade, hipóteses alternativas sobre forma funcional em modelos binários podem ser entrelaçada em um modelo dado por:



$$(23) L_r(Y | X, r, a, b, g) = \frac{[(1-a)L_1(Y | X, b)^r + aL_2(Y | X, g)^r]^{1/r}}{\int [(1-a)L_1(Z | X, b)^r + aL_2(Z | X, g)^r]^{1/r} dZ}$$

Onde L denota máxima verossimilhança. Para  $r = 1$ ,

$$(24) L(Y | X, a, b, g) = (1-a)L_1(Y | X, b) + aL_2(Y | X, g)$$

que é uma combinação linear convexa das formas funcionais alternativas e a decisão sobre a melhor hipótese recai sobre o valor de  $a$ .

### 3. RESULTADOS EMPÍRICOS

#### 3.1. SÍNTESE DA DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS

Na Tabela 1 é definido o conjunto de variáveis *dummies* que comporá o vetor de variáveis independentes, com suas especificações e o sinal esperado de cada uma quanto ao efeito marginal sobre o desempenho no concurso:

**TABELA 10** – ESPECIFICAÇÃO DAS VARIÁVEIS EXPLICATIVAS E SINAIS ESPERADOS

VARIÁVEIS		ESPECIFICAÇÃO		SINAL ESPERADO
		1	0	
Grau de Escolaridade	EMED	Ensino Médio	Outros	-
	ESUPO	Superior Completo ou com Pós-graduação	Outros	+
Estado Civil	SOLT	Solteiro	Outros	+
Renda Mensal da Família	YFALT	Maior que 10 Salários Mínimos (SM)	Menor ou igual a 10 SM.	+
Renda Pessoal Mensal	YPBAI	Menor ou igual a 2 SM	Maior que 2 SM.	-
Escola em que cursou o Ensino Médio	PRIV	Todo em Escola Privada	Outros	+
Tipo de Curso do Ensino Médio	TIPC	Ensino comum	Outros	+
Situação no Mercado de Trabalho	DEST	Estagiário ou Desempregado	Outros	+
Faixa Etária	IDBAI	De 18 a 22 anos	Acima de 22 anos	+
Local de Origem	RMETR	Oriundos de Regiões Metropolitanas	Outras	+

Inicialmente escolheu-se um modelo de escolha binária logit, cujo método de estimação utilizado foi o de Máxima Verossimilhança.

O algoritmo de maximização utilizado precisou de 10 interações para convergir, e o valor da função log-verossimilhança no ponto máximo é apresentado na Tabela 2, bem como o valor da estatística de razão de verossimilhança para testar a restrição de que todos os regressores (nove) fossem iguais a zero.

O ponto crítico para, uma qui-quadrado com 9 graus de liberdade, dado um nível de significância de 5%, é de **16,91**. O valor de  $-2\ln(\lambda) = 453,30$  aponta para a rejeição da hipótese nula. O *p-valor* representa a probabilidade de ocorrência de um erro do tipo I. O *p-valor* da estatística da razão de verossimilhança é inferior a 0,05 (Tabela 2), o que indica que a hipótese nula é rejeitada ao nível de significância de 5%.

Os resultados do modelo apresentam ainda duas medidas de ajustamento. Uma é o  $R^2$  de *McFadden*, o qual é extremamente baixo. A outra medida é o percentual de valores corretamente preditos (Tabela 2), o qual fornece o percentual corretamente predito de 1's, quando a probabilidade é maior do que certo valor crítico; e de zeros, quando a probabilidade é menor do que um certo valor crítico.

Usualmente este valor crítico é estabelecido como 0.5. No entanto, neste problema específico, como a proporção de zeros é substancialmente maior do que os valores de 1's, usou-se como valor crítico a probabilidade empírica de passar no concurso; ou seja, 0.00266. Os resultados deste novo cálculo são mostrados na Tabela 3. Assim a proporção de valores corretamente preditos para  $Y=1$  é de 61.26% e a proporção de valores corretamente preditos para  $Y=0$  é de 74.27%, como pode ser visto na Tabela 3. Tais resultados levam a um valor ponderado de predições corretas de 61.29%. No entanto, é razoável ter-se em mente que medidas de ajustamento em modelos de escolha binária não são importantes nem do ponto de vista estatístico nem do ponto de vista econômico para avaliar a significância das variáveis explicativas (*Wooldridge*, 2002). O teste da razão de máxima verossimilhança e os testes individuais de significância é tudo que se precisa para avaliar o modelo.

A razão entre o coeficiente estimado e o erro padrão (Tabela 2) representa uma estatística distribuída assintoticamente como uma normal padrão (Estatística *z*). Esses valores são utilizados para a realização do teste de hipótese sobre a significância estatística de cada um dos parâmetros estimados. Como se pode observar nas estatísticas de teste, somente a variável referente ao estado civil (SOLT) não se mostrou estatisticamente significativa (0,03), ao nível de significância de 5% (valor crítico de **1,96**).

Quanto ao *p-valor*, excetuando-se a variável SOLT, todas as demais variáveis apresentaram *p-valor* inferior a 0,05, o que significa dizer que a hipótese nula (coeficiente ser igual a zero) pode ser rejeitada ao nível de significância de 5%. Assim, a exceção da variável que representa o estado civil (SOLT), todas as demais são estatisticamente significantes. Vale lembrar, que os coeficientes das variáveis explicativas no modelo logit não expressam os efeitos marginais sobre a probabilidade da variável dependente assumir o valor 1.

Em seguida discute-se as razões dos sinais apresentados pelos valores estimados dos coeficientes da regressão. Para identificar um elemento comum que seja relevante quanto à condição de determinante de desempenho por parte das variáveis explicativas utilizadas, far-se-á observações que indiquem pontos de contato entre essas variáveis.

**TABELA 2 – RESULTADO DO MODELO *LOGIT***

<b>COVARIADAS</b>	<b>COEFICIENTE ESTIMADO</b>	<b>ERRO PADRÃO</b>	<b>ESTATÍSTICA Z</b>	<b>P-VALOR</b>
<b>C</b>	- 6,353	0,132	-47,98	0,0000
<b>EMED</b>	- 1,320	0,118	-11,17	0,0000
<b>ESUPO</b>	0,289	0,100	2,88	0,0039
<b>SOLT</b>	0,004	0,109	0,03	0,9696
<b>YFALT</b>	0,545	0,098	5,54	0,0000
<b>YPBAI</b>	-0,182	0,082	-2,21	0,0266
<b>PRIV</b>	0,250	0,088	2,84	0,0045
<b>DEST</b>	0,319	0,087	3,67	0,0002
<b>IDBAI</b>	0,325	0,108	3,00	0,0026
<b>RMETR</b>	0,503	0,088	5,69	0,0000
Média da Variável Dependente	0,002660	Critério de informação de Akaike		0,034995
S.D. da Variável Dependente	0,051509	Critério de Schwarz		0,035441
Erro Padrão da Regressão	0,051451	Critério de Hannan-Quinn		0,035126
Soma do quadrado dos resíduos	614,9497	R <sup>2</sup> McFadden		0,052937
Log-verossimilhança				-4.054,8530
Log-verossimilhança restrita				-4.281,5040
Estatística de Razão de verossimilhança [ln( $\lambda$ )] - 9 graus de liberdade				453,3016
p-valor de $\lambda$				0,0000
Obs. com y=0	231.690	Total de Observações		232.308
Obs. com y=1	618			

Nota: Dados gerados através do aplicativo *EViews 4.1*<sup>®</sup>

**TABELA 3 – PODER DE PREDIÇÃO DA ESTIMAÇÃO DO MODELO LOGIT**

	EQUAÇÃO ESTIMADA			PROBABILIDADE CONSTANTE		
	Y=0	Y=1	Total	Y=0	Y=1	Total
<b>P(Y=1)≤C</b>	141928	159	142087	0	0	0
<b>P(Y=1)&gt;C</b>	89762	459	90221	231690	618	232308
<b>Total</b>	231690	618	232308	231690	618	232308
<b>Correto</b>	141928	459	142387	0	618	618
<b>Correto (%)</b>	61.26	74.27	61.29	0.00	100.00	0.27
<b>Incorreto (%)</b>	38.74	25.73	38.71	100.00	0.00	99.73
<b>Acerto Total</b>	61.26	-25.73	61.03			
<b>Acerto Percentual</b>	61.26	NA	61.19			

Nota 1: Resultados gerados através do aplicativo *EViews 4.1*<sup>®</sup>

Nota 2: Corte para sucesso (Y=1), C = 0.00266026

Com o propósito de organizar a linha de raciocínio a ser desenvolvida, condensar-se-á as considerações em três blocos de variáveis; ou seja, aquelas diretamente ligadas à educação formal; as que se referem à renda; e as que representam determinações mais gerais.

Inicialmente tratar-se-á das determinações chamadas gerais; passando em seguida para as referentes à renda; e, por fim, para as que dizem respeito à educação formal.

## **DETERMINANTES GERAIS**

### **A) Local de Origem**

Os resultado apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos originários de regiões metropolitanas. Pode-se afirmar que o elemento diferencial desse determinante é a maior quantidade e melhor qualidade dos serviços ligados à educação e cultura nos grandes centros urbanos, *vis-à-vis*, os municípios do interior. Nesse sentido, o elemento fundamental que está por trás desse determinante é o acesso a um maior volume de serviços educacionais e culturais de melhor qualidade.

## **B) Faixa Etária**

No quesito faixa de idade os resultados apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos dentro da faixa mais baixa (18 a 22 anos). O resultado apresentado é considerado procedente com base na argumentação de melhor ritmo de estudo dos candidatos mais jovens, por estarem ainda freqüentando aulas ou por terem parado de estudar há pouco tempo. Mais uma vez, o elemento fundamental do determinante de desempenho está ligado a questões educacionais.

## **C) Situação no Mercado de Trabalho**

Um efeito marginal positivo sobre a chance de aprovação dos candidatos que estão desempregados ou são estagiários é tido como esperado, pois essa condição no mercado de trabalho possibilita uma maior disponibilidade de tempo desses para se dedicarem aos estudos. Novamente o estudo se apresenta como elemento central.

Adicionalmente, cabe observar que na pergunta do questionário sócio-econômico que trata da condição no mercado de trabalho não se observa uma alternativa referente a estar desocupado; ou seja, não estar procurando emprego.

Assim, dentre os que se denominaram desempregados, é possível ter uma parcela que tenha optado por se dedicar aos estudos antes de entrar no mercado de trabalho, ou para retornar a esse mercado com melhores condições de colocação ou remuneração.

Então, uma maior quantidade de tempo dedicada à obtenção de educação e cultura, e a boa qualidade desses serviços, representa um elemento constituinte que perpassa os determinantes: Local de Origem, Faixa de Idade e Situação no Mercado de Trabalho.

## **DETERMINANTES RELACIONADOS À RENDA**

### **A) Renda Pessoal Mensal**

Ter renda pessoal baixa apontou para um efeito marginal negativo sobre a chance de aprovação. Um resultado que está em conformidade com a argumentação de serem piores as condições de se preparar para o concurso para os que têm baixos rendimentos, visto que essa situação é limitante do acesso a bens materiais de estudo e cursinhos de qualidade. Destaca-se a relevância do acesso a produtos e serviços que reforcem, ou complementem, o aprendizado já adquirido, ou que iniciem os candidatos em temas desconhecidos, como os exigidos em provas específicas.

### **B) Renda Familiar Mensal**

Possuir uma renda familiar alta redundava num efeito marginal positivo sobre a chance de aprovação. Todo o background existente entre os que são de famílias de maior renda, conforme descrito anteriormente, redundava num perfil de uma pessoa que cursou ensino formal de qualidade, com acesso a materiais didáticos e para-didáticos de qualidade, ambiente propício ao desenvolvimento cognitivo, maior volume de informações em educação e cultura, acesso a atividades extracurriculares, inclusão digital, dentre outros inúmeros elementos materiais e intangíveis.

É nítida a melhor e mais adequada condição de acesso à educação e à cultura quando se fala de renda pessoal ou familiar. Base educacional e cultural é novamente o elemento central.

## **DETERMINANTES RELACIONADOS À EDUCAÇÃO FORMAL**

### **A) Ser Graduado ou Pós-Graduado (“super-educados”)**

Os candidatos “super-educados” apresentam efeito marginal positivo sobre a chance de aprovação. Isso reflete fundamentalmente o baixo nível geral do ensino médio. Os que têm melhores possibilidades vão no sentido de avançar nos estudos para adquirir mais conhecimento e sinalizar ao mercado de trabalho que eles seriam mais produtivos por terem nível superior.

Contudo, não se pode desprezar o fato de que num concurso que exige nível médio, no qual participaram praticamente 80.000 candidatos universitários, mais de 35.000 graduados e aproximadamente 5.600 pós-graduados, 28,5% das vagas ofertadas ficaram ociosas.

Ou seja, de um universo de cerca de 120.600 candidatos que são pelo menos universitários, somente 519 candidatos cumpriram os requisitos mínimos para serem aprovados no concurso, os demais perderam para si mesmos. Isto é uma forte evidência de que o baixo nível da qualidade do ensino tem transbordado para o nível superior.

Um dos componentes no estímulo à disseminação de cursos superiores de baixa qualidade é a incipiente formação de uma quantidade expressiva dos vestibulandos, fruto de um ensino médio de baixa qualidade.

Então, os que frequentaram cursos de ensino médio de baixa qualidade, para melhor se inserir no mercado de trabalho, buscaram cursar ensino superior em instituições de menor tradição e qualidade, que têm preços mais acessíveis e vestibulares menos exigentes. Contudo, a baixa qualidade desses cursos pode não garantir essa melhor inserção, nem produzir um melhor desempenho em concursos públicos.

### **B) Ter Cursado Apenas Curso de Nível Médio**

Ter cursado apenas o nível médio apresentou efeito marginal negativo sobre a chance de aprovação. Resultado também esperado, considerando a qualidade geral do ensino médio e que a concorrência com candidatos de grau universitário ou até mesmo com pós-graduação levaria a uma competição desigual principalmente em provas que tratavam de temas que não são nem mesmo discutidos no ensino médio.

### C) Ter Cursado o Ensino Médio em Escola Privada

Este item, na verdade, complementa o anterior. Ter cursado o ensino médio em escola privada apresentou efeito marginal positivo na chance de aprovação.

Considerando o sucateamento do ensino público brasileiro de nível médio e a ascensão da escola privada, onde são ofertadas metodologias de ensino, ambiente de estudo e acesso a materiais didáticos complementares incomparáveis com os disponíveis na rede pública, não se constitui em grande surpresa este resultado.

Quanto à magnitude dos coeficientes, como observado na metodologia, por se tratar de uma regressão não linear, os coeficientes apresentados na Tabela 12 não representam os efeitos marginais. O efeito marginal de cada variável independente para a distribuição logística pode ser obtido através da seguinte expressão,

$$(25) \frac{e^{x'b}}{1 + e^{x'b}} \left( 1 - \frac{e^{x'b}}{1 + e^{x'b}} \right) b$$

O produto  $x'b$  representa o vetor de coeficientes multiplicado por um vetor que contenha valores para as variáveis independentes.

É óbvio que o efeito marginal vai variar de acordo com os valores atribuídos às variáveis contidas em X. Sendo assim, utilizou-se os valores médios das variáveis explicativas para efetuar os cálculos dos efeitos marginais, os quais são apresentados na Tabela 4.

**TABELA 4 – EFEITOS MARGINAIS SOBRE A CHANCE DE SER APROVADO NO CONCURSO**

VARIÁVEIS EXPLICATIVAS		COEFICIENTE ESTIMADO (B)	VALOR MÉDIO (X)	(B)x(X)	EFEITO MARGINAL
Intercepto	C	-6,3538	-0,0013	0,00829	
Ensino Médio	EMED	-1,3209	0,4815	-0,63599	<b>-3,9%</b>
Nível Superior Completo ou com Pós-graduação	ESUPO	0,2897	0,1536	0,04450	<b>0,8%</b>
Estado Civil – Solteiro	SOLT	0,0042	0,7259	0,00303	<b>0,0%</b>
Renda Familiar superior a 10 Salários Mínimos	YFALT	0,5459	0,1012	0,05525	<b>1,6%</b>
Renda Individual abaixo de 2 Salários Mínimos	YPBAI	-0,1829	0,6382	-0,11672	<b>-0,5%</b>
Toda em escola Privada	PRIV	0,2503	0,3156	0,07901	<b>0,7%</b>
Estagiário ou Desempregado	DEST	0,3199	0,4914	0,15717	<b>0,9%</b>
Oriundos de Regiões Metropolitanas	RMETR	0,5034	0,4377	0,22034	<b>1,5%</b>
Idade de 18 a 22	IDBAI	0,3259	0,6636	0,21624	<b>1,0%</b>
$\beta'X$				<b>0,03113</b>	



O dado relevante quanto à magnitude dos efeitos marginais apresentados na Tabela 4 é que, das variáveis explicativas utilizadas no modelo, a condição do candidato ter o grau de escolaridade adequado, nível médio, para a função a ser exercida foi o que apresentou maior impacto (o qual é negativo) na chance do candidato ser aprovado.

Para avaliar problemas de especificação do modelo dois tipos de teste foram realizados. O primeiro foi o teste de heterocedasticidade, como descrito na metodologia. Vale lembrar que este teste foi usado não como um detector de heterocedasticidade, pois o método de estimação implicitamente aplica uma ponderação de mínimos quadrados não linear.

O objetivo do teste é detectar erro de especificação para o caso de heterogeneidade não observada. Como a suspeita sobre o erro de especificação recai sobre a heterogeneidade não observada, ou mais especificamente sobre a possível ausência de uma medida de habilidade para cada candidato na equação, optou-se por definir o vetor  $Z$  na variância do erro como constituído apenas das variáveis de instrução.

Os resultados deste teste mostram que o valor calculado para o teste de *Lagrange* é  $LM=0,623415$ . Já o valor crítico da distribuição *qui-quadrado*, a um nível de significância de 5% e com dois graus de liberdade é 5,99. O valor-p é de 0,7321. Portanto, aceita-se a hipótese nula de homocedasticidade, o que aqui significa ausência de erro de especificação. Este resultado pode parecer surpreendente à primeira vista; no entanto, é em essência o mesmo encontrado por Griliches e Mason (1972), os quais não encontraram vestígios de correlação entre habilidade e escolaridade.

O segundo teste referente a erro de especificação realizado foi o da forma funcional, também descrito na metodologia. Apesar da argumentação de alguns autores (*Train* (2003), por exemplo) de que os modelos *Probit* e *Logit* são empíricamente indistinguíveis, existem situações em que eles podem divergir, requerendo portanto um teste formal de escolha.

Assim optou-se por um teste de hipótese para discriminar entre os modelos logit e probit. Para tanto o modelo probit foi estimado e seus resultados são apresentados no Apêndice B. Já o resultado do teste é apresentado no Apêndice C. Como pode ser visto, como o valor do coeficiente da variável  $Z = F_{2i} - F_{1i}$  é estatisticamente insignificante opta-se pelo modelo logit.

Optar pelo modelo logit tendo como alternativa o modelo probit não leva à conclusão que o modelo logit é o mais apropriado. Realizou-se então o teste tendo como hipótese alternativa uma distribuição gompit como representativa da classe assimétrica.

O modelo de regressão usando a distribuição gompit é definido como,

$$(26) P(Y=1|X) = \exp(-\exp(-x'\beta))$$

Os resultados da estimação deste modelo podem ser visto no Apêndice D. O resultado do teste para discriminar entre o logit e o gompit é mostrado no Apêndice E.

Vê-se novamente que o coeficiente estimado de  $Z$  é estatisticamente insignificante ao nível usual de 5%. Portanto, pode-se ter uma razoável confiança de que o modelo escolhido foi bem especificado.

## 4. CONCLUSÕES

No presente estudo de caso de avaliação de desempenho em concurso público foram destacados, na introdução, alguns exemplos de questões primordiais a serem respondidas:

- O que faz determinadas pessoas serem aprovadas em detrimento de outras?
- Existem fatores sócio-econômicos que aumentam as chances de uma pessoa ter um melhor desempenho?
- Maior renda implica em maiores chances de passar?
- Ter nível de educação acima do exigido no concurso implica em maiores chances de aprovação?

Após a especificação de um modelo econométrico do tipo *Logit* onde o sucesso no concurso é definido como função de um conjunto de variáveis explicativas constituído por renda familiar e pessoal, grau de instrução, tipo de escola freqüentada, estado civil, região onde habita, situação atual no mercado de trabalho e idade. Constatou-se, após a estimação do modelo, que renda, escolaridade acima do ensino médio, ser oriundo de região metropolitana e estar em baixa faixa de idade, contribuem positivamente para aumentar a probabilidade de passar no concurso. Por outro lado, ser detentor apenas de escolaridade de ensino médio contribui negativamente para aprovação no concurso, e o coeficiente dessa variável representa o maior efeito marginal em termos absolutos, o que dramatiza a situação dos candidatos de mais baixa escolaridade.

Delineia-se assim um perfil de um candidato jovem, tipicamente pertencente a uma classe média de renda elevada e que teve acesso à instrução e meios culturais de boa qualidade, como elemento típico de quem obteve sucesso no concurso. Reitera-se então, através dos resultados obtidos, o círculo vicioso de concentração de renda onde candidatos provenientes de famílias de melhor estrutura financeira e maiores informações culturais têm maiores chances de aprovação e acesso a postos de trabalho com maior remuneração média.

Em termos de políticas públicas que contrabalancem estes efeitos perversos nota-se que todos os determinantes do modelo apontam, em última instância, para a necessidade de uma base educacional e cultural de qualidade. O ensino de nível médio parece ser um gargalo no processo de disseminação desse importante condicionante, o qual dentro do escopo desse trabalho, revela-se fundamental não só ao bom desempenho no concurso objeto de estudo, como para quaisquer outros concursos.

Entende-se que o ensino médio público pode ocupar lugar estratégico quanto à criação de condições mais igualitárias na preparação e no desempenho em concursos. É no âmbito do ensino médio que se pode ter elementos que se não suprem todas as carências dos candidatos menos afortunados, podem atenuar a falta de condição no âmbito familiar, ou a deficiência estrutural de uma região pobre. Portanto, identifica-se o ensino médio público como um *locus* de ação e políticas que atuem para reduzir as desigualdades de condições iniciais de cada candidato.

Por outro lado, preencher postos de trabalho que requerem apenas uma educação de nível médio com detentores de título universitário, e até mesmo com pós-graduação, aponta para uma má alocação dos recursos públicos conjugada a uma possível deficiência estrutural do mercado de trabalho regional.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL, República Federativa do. (2004) **ENEM 2003: Relatório Final**. DACC/MEC. Brasília.

FAIR, R.C. (1978) A Theory of Extramarital Affairs. **Journal of Political Economy**, vol. 86, pp. 45-61.

GREENE, W.H. (1993) **Econometric Analysis**. Prentice Hall, New Jersey.

GRILICHES, Z. e MANSON, W. M. (1972). Education, Income and Ability, **Journal of Political Economy**, 80:3, part 2 may/june.

HSIAO, C. (2003) **Analysis of Panel Data**. Cambridge University Press Cambridge.

RUUD, P. A. (2004) **An Introduction to Classical Econometric Theory**. Oxford University Press. New York.

SILVA, A. S. S. (2001) A Score test for non-nested Hipoteses with Applications to Discrete Data Model. **Journal of Applied Econometrics**, 16, pp. 577-97.

SARGAN, D. (1975) A Comment. In: G. A Renton (Ed.). **Modelling the Economy**.

SPECTOR L.C. e MAZZEO M.A. (1980) Probit Analysis and Economic Education, **Journal of Economic Education**, pp. 37-44.

TRAIN, K. (2003) **Discrete Choice Methods with Simulation**. Cambridge University Press. Cambridge.

WHITE, H. (1980) A Heteroskedasticity-Consistente Covariance Matrix Estimation and Direct Test for Heteroskedasticity. **Econometrica**, vol. 48, n. 4, pp. 817-38.

WOOD, W. C. e DOYLE, J. M. (2002) Economic Literacy Among Corporate Employees. **Journal of Economic Education** (Summer) pp. 195-205.

WOOLDRIDGE, J. M. (2002) **Econometric Analysis of Cross Section and Panel Data**. The MIT Press. Cambridge.