Lista de Exercícios #2 – Econometria I

Dados em Painel · Modelos de Escolha Discreta

PPGEco/2026

Questão 1 Dados em Painel: POLS, FE e FD

Objetivo. Comparar os estimadores de OLS agrupado (POLS), efeitos fixos (FE) e primeiras diferenças (FD) em um painel pequeno, e compreender quando cada um é adequado na presença de heterogeneidade individual não observada.

Considere o modelo:

$$y_{it} = \beta_1 + \beta_2 x_{it} + c_i + \varepsilon_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

em que c_i é um efeito individual não observado, invariante no tempo, e ε_{it} é o erro idiossincrático.

Use o painel abaixo para todos os cálculos.

i	t	x_{it}	y_{it}
1	1	2	5
1	2	4	9
1	3	6	11
2	1	1	4
2	2	3	6
2	3	5	10

(a) POLS. Empilhe as $nT = 6$ observações ($\hat{\beta}_1^{\text{POLS}}, \hat{\beta}_2^{\text{POLS}}$) à mão, usando as fórmulas usuais de OLS. Mostre as somas utilizadas ou a forma matricial do cálculo.

(b) Estimador de efeitos fixos (FE).

i. Calcule as médias individuais $\bar{x}_i$ e $\bar{y}_i$ para cada i e construa as variáveis demeaned $\tilde{x}_{it} = x_{it} - \bar{x}_i$ e $\tilde{y}_{it} = y_{it} - \bar{y}_i$.

ii. Estime $\hat{\beta}_2^{\text{FE}}$ regredindo $\tilde{y}_{it}$ em $\tilde{x}_{it}$ sem constante.

iii. Recupere os efeitos fixos estimados $\hat{c}_1$ e $\hat{c}_2$.

(c) Estimador em primeiras diferenças (FD).

i. Calcule $\Delta y_{it} = y_{it} - y_{i,t-1}$ e $\Delta x_{it} = x_{it} - x_{i,t-1}$ para $t = 2, 3$.

ii. Estime $\hat{\beta}_2^{\text{FD}}$ regredindo Δy_{it} em Δx_{it} sem constante.

(d) Comparação dos estimadores. Compare numericamente $\hat{\beta}_2^{\text{POLS}}$, $\hat{\beta}_2^{\text{FE}}$ e $\hat{\beta}_2^{\text{FD}}$. Explique por que eles podem diferir mesmo quando calculados a partir da mesma base de dados.

(e) Variável omitida e endogeneidade. Suponha que $\text{Corr}(c_i, x_{it}) \neq 0$.

i. Explique por que o estimador POLS é viesado e inconsistente nesse caso.

ii. Explique por que FE e FD eliminam o viés causado por c_i, mesmo sem observá-lo.

(f) Limites de FE e FD. Quais hipóteses sobre ε_{it} ainda são necessárias para que FE e FD sejam consistentes? Dê um exemplo de situação em que FE e FD *não* resolvem o problema de endogeneidade.

Questão 2 Variável Dependente Binária

Objetivo. Mostrar como um modelo de variável latente gera modelos binários como logit e probit. O foco está na verossimilhança, na normalização da escala do erro, na interpretação dos coeficientes, nos efeitos marginais e na inferência baseada em máxima verossimilhança.

Considere o modelo de variável latente:

$$y_i^* = \beta_1 + \beta_2 x_i + u_i, \qquad y_i = \mathbf{1}[y_i^* > 0],$$

com amostra i.i.d. $\{(y_i, x_i)\}_{i=1}^n$ e $u_i \sim F(\cdot)$, independente de x_i. Suponha que F é uma cdf contínua, estritamente crescente e simétrica em torno de zero: $F(z) = 1 - F(-z)$.

Parte A Probabilidade, verossimilhança e log-verossimilhança

A.1 Mostre que $E(y_i \mid x_i) = \Pr(y_i = 1 \mid x_i) = F(\beta_1 + \beta_2 x_i)$.

A.2 Defina $z_i = \beta_1 + \beta_2 x_i$ e $F_i = F(z_i)$. Escreva a verossimilhança conjunta $L(\beta)$ para as n observações. Explique por que a hipótese de amostra i.i.d. é fundamental para essa construção.

A.3 Mostre que a log-verossimilhança é:

$$\ln L(\beta) = \sum_{i=1}^{n} \big[y_i \ln F_i + (1 - y_i) \ln(1 - F_i) \big].$$

A.4 Explique por que se maximiza $\ln L(\beta)$ em vez de $L(\beta)$ diretamente.

A.5 Mostre que $\ln L(\beta) \leq 0$ para todo β. Em que situação $\ln L(\beta) = 0$? Essa situação é normalmente atingida em amostras finitas?

Parte B Gradiente, Hessiano e maximização numérica

Defina $f_i = F'(z_i)$.

B.1 Calcule o gradiente $\nabla_\beta \ln L(\beta)$ e o Hessiano $H(\beta) = \nabla^2_\beta \ln L(\beta)$.

B.2 Escreva a regra de iteração de Newton-Raphson:

$$\beta^{(k+1)} = \beta^{(k)} - \big[H(\beta^{(k)}) \big]^{-1} \nabla_\beta \ln L(\beta^{(k)}).$$

B.3 Discuta possíveis escolhas para o ponto inicial $\beta^{(0)}$ e defina um critério de convergência baseado em $\|\beta^{(k+1)} - \beta^{(k)}\|$.

B.4 Explique por que o algoritmo pode travar quando o Hessiano está em uma região quase singular. O que fazer nesse caso?

Parte C Logit, Probit e normalizações do erro latente

C.1 *Logit.* Tome $F(z) = \Lambda(z) = \frac{1}{1+e^{-z}}$. Mostre que $f(z) = \Lambda(z)[1 - \Lambda(z)]$ e escreva o gradiente e o Hessiano para esse caso específico.

C.2 *Probit.* Tome $F(z) = \Phi(z)$, de modo que $f(z) = \phi(z)$. Escreva o gradiente e o Hessiano correspondentes.

C.3 Explique intuitivamente por que existe a relação aproximada $\hat{\beta}^{\text{logit}} \approx 1{,}6\,\hat{\beta}^{\text{probit}}$.

C.4 Explique por que, para fins de cálculo de probabilidades marginais, frequentemente importa pouco se o modelo estimado é logit ou probit.

C.5 No logit, mostre que o log da razão de chances satisfaz:

$$\ln \frac{\Pr(y_i = 1 \mid x_i)}{\Pr(y_i = 0 \mid x_i)} = \beta_1 + \beta_2 x_i.$$

Explique como isso permite interpretar $e^{\hat{\beta}_2}$ como a razão entre razões de chance associada a um incremento unitário em x_i. Em que contextos essa interpretação é vantajosa em relação ao probit?

C.6 *Normalização da média.* Suponha $E(u_i) = \mu$. Mostre que

$$y_i^* = (\beta_1 + \mu) + \beta_2 x_i + (u_i - \mu),$$

e explique por que, na presença de constante, μ não é separadamente identificado de β_1. Em que sentido fixar $E(u_i) = 0$ é apenas uma normalização de localização?

C.7 *Normalização da variância.* A presença de constante é suficiente para fixar $\text{Var}(u_i) = 1$? Explique por que não.

C.8 Suponha $u_i = \sigma\varepsilon_i$, com F_ε sendo a cdf de ε_i. Mostre que

$$\Pr(y_i = 1 \mid x_i) = F_\varepsilon\!\left(\frac{\beta_1 + \beta_2 x_i}{\sigma}\right).$$

Conclua que, em modelos binários, a escala do erro latente não é identificada separadamente dos coeficientes, e relacione esse resultado às normalizações usuais do logit e do probit.

Parte D Exercício numérico: logit com regressor binário

Contexto. Uma amostra com $n = 100$ indivíduos. A variável $Y \in \{0, 1\}$ indica fracasso ou sucesso; $X \in \{0, 1\}$ indica grupo de controle ou grupo tratado. Frequências observadas:

	$Y = 0$	$Y = 1$
$X = 0$	30	10
$X = 1$	20	40

Considere o modelo logit $\Pr(Y_i = 1 \mid X_i) = \Lambda(\beta_0 + \beta_1 X_i)$, onde $\Lambda(z) = e^z/(1 + e^z)$.

D.1 Defina $p_0 = \Pr(Y_i = 1 \mid X_i = 0)$ e $p_1 = \Pr(Y_i = 1 \mid X_i = 1)$. Mostre que $p_0 = \Lambda(\beta_0)$ e $p_1 = \Lambda(\beta_0 + \beta_1)$.

D.2 Escreva a verossimilhança e a log-verossimilhança com base nas contagens da tabela.

D.3 Como X é binária, mostre que as estimativas de máxima verossimilhança coincidem com as proporções amostrais:

$$\hat{p}_0 = \frac{10}{40} = 0{,}25 \qquad \text{e} \qquad \hat{p}_1 = \frac{40}{60} = \tfrac{2}{3}.$$

D.4 Use a função $\text{logit}(p) = \ln\big(p/(1-p)\big)$ para obter os coeficientes estimados:

$$\hat{\beta}_0 = \text{logit}(\hat{p}_0), \qquad \hat{\beta}_1 = \text{logit}(\hat{p}_1) - \text{logit}(\hat{p}_0).$$

Calcule numericamente $\hat{\beta}_0$ e $\hat{\beta}_1$.

D.5 Interprete o sinal de $\hat{\beta}_1$: o tratamento está associado a maior ou menor probabilidade de sucesso? Interprete também $e^{\hat{\beta}_1}$ como razão de chances (*odds ratio*).

D.6 Calcule o efeito marginal discreto do tratamento:

$$\widehat{\Delta} = \hat{p}_1 - \hat{p}_0.$$

Interprete em pontos percentuais. Por que, com X binária, faz mais sentido falar em efeito marginal discreto do que em derivada em relação a X?

D.7 Calcule o pseudo-R^2 de McFadden:

$$\widetilde{R}^2 = 1 - \frac{\ln \hat{L}}{\ln \hat{L}_0},$$

usando $\hat{p} = 50/100 = 0{,}5$ no modelo restrito (só constante). Comente por que esse pseudo-R^2 não deve ser interpretado como o R^2 da regressão linear.

D.8 Calcule o AIC do modelo com X ($k = 2$) e do modelo só com constante ($k = 1$):

$$\text{AIC} = -2 \ln \hat{L} + 2k.$$

Qual dos dois apresenta menor AIC? O que isso sugere?

D.9 Calcule a matriz de variância-covariância de $(\hat{\beta}_0, \hat{\beta}_1)'$ usando a matriz de informação observada:

$$\widehat{\mathcal{I}} = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} \begin{pmatrix} 1 & X_i \end{pmatrix}.$$

Mostre que, usando as contagens da tabela:

$$\widehat{\mathcal{I}} = 40\, \hat{p}_0(1-\hat{p}_0) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 60\, \hat{p}_1(1-\hat{p}_1) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Em seguida, obtenha $\widehat{\text{Var}}(\hat{\beta}) = \widehat{\mathcal{I}}^{-1}$.

D.10 Usando os erros-padrão do item anterior, calcule as estatísticas de Wald:

$$z_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}, \quad j = 0, 1.$$

As estimativas são estatisticamente diferentes de zero aos níveis usuais de significância?

D.11 Explique para que serve o método Delta e por que ele é necessário para calcular erros-padrão de quantidades como $e^{\hat{\beta}_1}$, $\hat{p}_1 - \hat{p}_0$ ou efeitos marginais. Por que esses erros-padrão *não* podem ser obtidos simplesmente aplicando a mesma transformação aos erros-padrão dos coeficientes?

Questão 3 Modelos Multinomial e Ordinal: Conceitos e Intuição

Objetivo. Discutir modelos de escolha discreta com mais de duas alternativas. O foco é consolidar a intuição econométrica e econômica dos modelos multinomiais e ordinais, sem exigir derivações formais longas.

Parte A Logit multinomial e utilidade aleatória

Considere um agente que escolhe entre J alternativas. A utilidade da alternativa j para o indivíduo i é:

$$U_{ij} = x_{ij}'\beta + \varepsilon_{ij},$$

onde ε_{ij} é um componente aleatório não observado pelo econometrista.

A.1 Explique, em suas próprias palavras, o que é um modelo de utilidade aleatória (RUM). Por que o econometrista modela a utilidade como aleatória se, do ponto de vista do agente, a escolha é determinística?

A.2 A fórmula do logit multinomial é:

$$\Pr(y_i = j \mid X_i) = \frac{\exp(x_{ij}'\beta)}{\sum_{k=1}^{J} \exp(x_{ik}'\beta)}.$$

Explique intuitivamente (a) por que as probabilidades são positivas e somam 1, e (b) por que a distribuição dos erros ε_{ij} importa para a forma funcional do modelo.

A.3 O que é a hipótese IIA (*Independence of Irrelevant Alternatives*)? Use o exemplo do ônibus vermelho/azul para mostrar quando ela é economicamente implausível.

Parte B Modelos ordinais: ologit e oprobit

Considere uma variável dependente ordinal $y_i \in \{1, 2, \ldots, J\}$ gerada pelo modelo latente:

$$y_i^* = x_i'\beta + u_i, \qquad y_i = j \iff \kappa_{j-1} < y_i^* \leq \kappa_j,$$

com limiares $-\infty = \kappa_0 < \kappa_1 < \cdots < \kappa_J = +\infty$ e $u_i \sim F(\cdot)$ independente de x_i.

B.1 Explique intuitivamente o papel dos limiares κ_j: como eles traduzem uma variável latente contínua y_i^* nas categorias discretas observadas $y_i \in \{1, \ldots, J\}$?

B.2 Por que um modelo ordinal é preferível a:

(a) tratar y_i como variável contínua e aplicar MQO?

(b) ignorar a ordenação e aplicar um logit multinomial irrestrito?

B.3 Como se interpreta um coeficiente $\hat{\beta}_2 > 0$ em termos das probabilidades das categorias mais altas e mais baixas? Por que o efeito marginal de x_i pode ter sinais opostos para categorias diferentes?

B.4 Suponha que y_i representa o nível de escolaridade em três categorias: fundamental, médio e superior. Justifique a escolha entre logit multinomial e ologit para modelar essa variável.

Gabarito – Lista de Exercícios #2

Questão 1 Dados em Painel: POLS, FE e FD

(a) POLS

Empilhando as seis observações, temos:

$$(x, y) = (2, 5), (4, 9), (6, 11), (1, 4), (3, 6), (5, 10).$$

As médias amostrais são:

$$\bar{x} = \frac{2+4+6+1+3+5}{6} = \frac{21}{6} = 3{,}5,$$

e

$$\bar{y} = \frac{5+9+11+4+6+10}{6} = \frac{45}{6} = 7{,}5.$$

O estimador de inclinação do OLS agrupado é:

$$\hat{\beta}_2^{\text{POLS}} = \frac{\sum_{i,t}(x_{it} - \bar{x})(y_{it} - \bar{y})}{\sum_{i,t}(x_{it} - \bar{x})^2}.$$

Calculando as somas:

$$\sum_{i,t}(x_{it} - \bar{x})(y_{it} - \bar{y}) = 26{,}5,$$

e

$$\sum_{i,t}(x_{it} - \bar{x})^2 = 17{,}5.$$

Logo,

$$\hat{\beta}_2^{\text{POLS}} = \frac{26{,}5}{17{,}5} = \frac{53}{35} \approx 1{,}514.$$

O intercepto é:

$$\hat{\beta}_1^{\text{POLS}} = \bar{y} - \hat{\beta}_2^{\text{POLS}}\bar{x} = 7{,}5 - 1{,}514(3{,}5) = 2{,}2.$$

Portanto:

$$\boxed{\hat{y}_{it} = 2{,}2 + 1{,}514x_{it}}.$$

(b) Estimador de efeitos fixos (FE)

Para o indivíduo 1:

$$\bar{x}_1 = \frac{2+4+6}{3} = 4,$$

e

$$\bar{y}_1 = \frac{5+9+11}{3} = \frac{25}{3} \approx 8{,}333.$$

Para o indivíduo 2:

$$\bar{x}_2 = \frac{1+3+5}{3} = 3,$$

e

$$\bar{y}_2 = \frac{4+6+10}{3} = \frac{20}{3} \approx 6{,}667.$$

As variáveis demeaned são:

i	t	$\tilde{x}_{it}$	$\tilde{y}_{it}$	$\tilde{x}_{it}\tilde{y}_{it}$
1	1	-2	$-10/3$	$20/3$
1	2	0	$2/3$	0
1	3	2	$8/3$	$16/3$
2	1	-2	$-8/3$	$16/3$
2	2	0	$-2/3$	0
2	3	2	$10/3$	$20/3$

Assim:

$$\sum_{i,t} \tilde{x}_{it}\tilde{y}_{it} = \frac{20+16+16+20}{3} = 24,$$

e

$$\sum_{i,t} \tilde{x}_{it}^2 = 4+0+4+4+0+4 = 16.$$

Logo:

$$\hat{\beta}_2^{\text{FE}} = \frac{24}{16} = 1{,}5.$$

Portanto:

$$\boxed{\hat{\beta}_2^{\text{FE}} = 1{,}5}.$$

Para recuperar os efeitos fixos individuais, calcula-se:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_2^{\text{FE}}\bar{x}_i,$$

onde $\alpha_i = \beta_1 + c_i$ é o intercepto específico do indivíduo. Assim:

$$\hat{\alpha}_1 = \frac{25}{3} - 1{,}5(4) = \frac{25}{3} - 6 = \frac{7}{3} \approx 2{,}333,$$

e

$$\hat{\alpha}_2 = \frac{20}{3} - 1{,}5(3) = \frac{20}{3} - \frac{9}{2} = \frac{13}{6} \approx 2{,}167.$$

Assim:

$$\boxed{\hat{\alpha}_1 = \frac{7}{3} \qquad \text{e} \qquad \hat{\alpha}_2 = \frac{13}{6}}.$$

Como o modelo contém uma constante comum e efeitos individuais, β_1 e c_i não são separadamente identificados sem uma normalização adicional. O estimador FE recupera os interceptos individuais $\alpha_i = \beta_1 + c_i$.

(c) Estimador em primeiras diferenças (FD)

Para o indivíduo 1:

$$\Delta x_{12} = 4-2 = 2, \qquad \Delta y_{12} = 9-5 = 4,$$

e

$$\Delta x_{13} = 6-4 = 2, \qquad \Delta y_{13} = 11-9 = 2.$$

Para o indivíduo 2:

$$\Delta x_{22} = 3-1 = 2, \qquad \Delta y_{22} = 6-4 = 2,$$

e

$$\Delta x_{23} = 5-3 = 2, \qquad \Delta y_{23} = 10-6 = 4.$$

Logo, as quatro observações em primeiras diferenças são:

$$(\Delta x, \Delta y) = (2,4), (2,2), (2,2), (2,4).$$

O estimador FD, sem constante, é:

$$\hat{\beta}_2^{\text{FD}} = \frac{\sum \Delta x_{it}\Delta y_{it}}{\sum (\Delta x_{it})^2}.$$

Calculando:

$$\sum \Delta x_{it}\Delta y_{it} = 2(4)+2(2)+2(2)+2(4) = 24,$$

e

$$\sum (\Delta x_{it})^2 = 4+4+4+4 = 16.$$

Portanto:

$$\boxed{\hat{\beta}_2^{\text{FD}} = \frac{24}{16} = 1{,}5}.$$

(d) Comparação dos estimadores

Os estimadores encontrados foram:

$$\hat{\beta}_2^{\text{POLS}} = 1{,}514,$$
$$\hat{\beta}_2^{\text{FE}} = 1{,}5,$$

e

$$\hat{\beta}_2^{\text{FD}} = 1{,}5.$$

Neste exemplo, FE e FD coincidem, mas isso não é uma propriedade geral para painéis com $T > 2$. Eles coincidem aqui por causa da estrutura simples e simétrica dos dados. O POLS é ligeiramente diferente porque usa tanto a variação dentro dos indivíduos quanto a variação entre indivíduos. Já FE e FD exploram apenas a variação dentro dos indivíduos ao longo do tempo, eliminando o componente invariante c_i.

(e) Variável omitida e endogeneidade

Se:

$$\text{Corr}(c_i, x_{it}) \neq 0,$$

então o termo de erro composto do POLS,

$$v_{it} = c_i + \varepsilon_{it},$$

é correlacionado com x_{it}. Nesse caso:

$$E(v_{it} \mid x_{it}) \neq 0.$$

Logo, o estimador POLS viola a hipótese de exogeneidade e torna-se viesado e inconsistente.

O estimador FE elimina c_i pela transformação within:

$$y_{it} - \bar{y}_i = \beta_2(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i).$$

Como c_i é constante no tempo, ele desaparece:

$$c_i - \bar{c}_i = 0.$$

O estimador FD elimina c_i pela diferenciação:

$$y_{it} - y_{i,t-1} = \beta_2(x_{it} - x_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

Novamente, como c_i é invariante no tempo:

$$c_i - c_i = 0.$$

Assim, FE e FD resolvem o problema de endogeneidade causado por heterogeneidade individual não observada constante no tempo.

(f) Limites de FE e FD

Mesmo após eliminar c_i, é necessário que o erro idiossincrático satisfaça uma condição de exogeneidade apropriada. Para FE, uma hipótese usual é exogeneidade estrita:

$$E(\varepsilon_{it} \mid x_{i1}, x_{i2}, \ldots, x_{iT}, c_i) = 0 \quad \text{para todo } t.$$

Para FD, é necessário que:

$$E(\Delta\varepsilon_{it} \mid \Delta x_{it}) = 0,$$

ou condições mais fortes envolvendo a trajetória de x.

FE e FD não resolvem problemas de endogeneidade causados por variáveis omitidas que variam no tempo. Por exemplo, se um choque de produtividade, motivação, saúde ou demanda afeta simultaneamente x_{it} e y_{it}, então:

$$\text{Corr}(x_{it}, \varepsilon_{it}) \neq 0.$$

Nesse caso, mesmo após eliminar c_i, o estimador continua inconsistente.

Questão 2 Variável Dependente Binária

Parte A Probabilidade, verossimilhança e log-verossimilhança

A.1

Como:

$$y_i = \mathbf{1}[y_i^* > 0],$$

temos:

$$E(y_i \mid x_i) = \Pr(y_i = 1 \mid x_i).$$

Além disso:

$$y_i = 1 \iff y_i^* > 0 \iff \beta_1 + \beta_2 x_i + u_i > 0.$$

Logo:

$$\Pr(y_i = 1 \mid x_i) = \Pr(u_i > -(\beta_1 + \beta_2 x_i) \mid x_i).$$

Como u_i é independente de x_i:

$$\Pr(y_i = 1 \mid x_i) = 1 - F[-(\beta_1 + \beta_2 x_i)].$$

Como F é simétrica em torno de zero:

$$1 - F(-z) = F(z).$$

Portanto:

$$\boxed{E(y_i \mid x_i) = \Pr(y_i = 1 \mid x_i) = F(\beta_1 + \beta_2 x_i)}.$$

A.2

Defina:

$$z_i = \beta_1 + \beta_2 x_i, \qquad F_i = F(z_i).$$

A contribuição individual para a verossimilhança é:

$$\ell_i(\beta) = F_i^{y_i}(1 - F_i)^{1-y_i}.$$

Se $y_i = 1$, então:

$$\ell_i(\beta) = F_i.$$

Se $y_i = 0$, então:

$$\ell_i(\beta) = 1 - F_i.$$

Como a amostra é i.i.d., a verossimilhança conjunta é o produto das contribuições individuais:

$$\boxed{L(\beta) = \prod_{i=1}^{n} F_i^{y_i}(1 - F_i)^{1-y_i}}.$$

A hipótese i.i.d. é fundamental porque permite escrever a probabilidade conjunta como produto das probabilidades individuais.

A.3

Tomando logaritmo:

$$\ln L(\beta) = \ln \prod_{i=1}^{n} F_i^{y_i}(1 - F_i)^{1-y_i}.$$

Logo:

$$\boxed{\ln L(\beta) = \sum_{i=1}^{n} [y_i \ln F_i + (1 - y_i) \ln(1 - F_i)]}.$$

A.4

Maximiza-se $\ln L(\beta)$ em vez de $L(\beta)$ por três razões principais. Primeiro, a função logaritmo é estritamente crescente, então:

$$\arg\max_{\beta} L(\beta) = \arg\max_{\beta} \ln L(\beta).$$

Segundo, a log-verossimilhança transforma produtos em somas, o que facilita a derivação e a computação. Terceiro, trabalhar com logaritmos melhora a estabilidade numérica, pois produtos de muitas probabilidades pequenas podem gerar valores próximos de zero.

A.5

Como $0 < F_i < 1$, temos:

$$\ln F_i < 0 \qquad \text{e} \qquad \ln(1 - F_i) < 0.$$

Logo, cada termo da log-verossimilhança é menor ou igual a zero:

$$y_i \ln F_i + (1 - y_i) \ln(1 - F_i) \leq 0.$$

Portanto:

$$\boxed{\ln L(\beta) \leq 0}.$$

A igualdade $\ln L(\beta) = 0$ exigiria que cada observação recebesse probabilidade igual a 1. Ou seja, para cada i:

$$\Pr(y_i \mid x_i) = 1.$$

Isso não ocorre normalmente em modelos logit ou probit com parâmetros finitos, pois as probabilidades ficam estritamente entre 0 e 1. Pode ocorrer apenas como limite, por exemplo em situações de separação perfeita.

Parte B Gradiente, Hessiano e maximização numérica

B.1

Defina:

$$w_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}, \qquad z_i = w_i'\beta, \qquad F_i = F(z_i), \qquad f_i = F'(z_i).$$

A log-verossimilhança é:

$$\ln L(\beta) = \sum_{i=1}^{n} \left[y_i \ln F_i + (1 - y_i) \ln(1 - F_i) \right].$$

Derivando em relação a β:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{y_i}{F_i} - \frac{1 - y_i}{1 - F_i} \right] f_i w_i.$$

Logo:

$$\boxed{\nabla_\beta \ln L(\beta) = \sum_{i=1}^{n} \frac{(y_i - F_i) f_i}{F_i(1 - F_i)} w_i}.$$

Para o Hessiano, escreva:

$$q_i = \frac{(y_i - F_i) f_i}{F_i(1 - F_i)}.$$

Então:

$$\nabla_\beta \ln L(\beta) = \sum_{i=1}^{n} q_i w_i.$$

Como w_i não depende de β:

$$H(\beta) = \sum_{i=1}^{n} \frac{\partial q_i}{\partial z_i} w_i w_i'.$$

Se $f_i' = F''(z_i)$, então:

$$\frac{\partial q_i}{\partial z_i} = y_i \left[\frac{f_i'}{F_i} - \frac{f_i^2}{F_i^2} \right] - (1 - y_i) \left[\frac{f_i'}{1 - F_i} + \frac{f_i^2}{(1 - F_i)^2} \right].$$

Portanto:

$$\boxed{H(\beta) = \sum_{i=1}^{n} \left\{ y_i \left[\frac{f_i'}{F_i} - \frac{f_i^2}{F_i^2} \right] - (1 - y_i) \left[\frac{f_i'}{1 - F_i} + \frac{f_i^2}{(1 - F_i)^2} \right] \right\} w_i w_i'}.$$

B.2

A regra de Newton-Raphson é:

$$\boxed{\beta^{(k+1)} = \beta^{(k)} - [H(\beta^{(k)})]^{-1} \nabla_\beta \ln L(\beta^{(k)})}.$$

A ideia é aproximar a log-verossimilhança por uma expansão quadrática ao redor de $\beta^{(k)}$ e mover o vetor de parâmetros na direção que zera aproximadamente o gradiente.

B.3

Escolhas usuais para o ponto inicial incluem:

$$\beta^{(0)} = (0, 0)',$$

coeficientes de um modelo de probabilidade linear, ou valores obtidos em uma especificação mais simples.

Um critério de convergência baseado nos parâmetros é:

$$\|\beta^{(k+1)} - \beta^{(k)}\| < \epsilon.$$

Também é comum usar critério baseado na log-verossimilhança:

$$|\ln L(\beta^{(k+1)}) - \ln L(\beta^{(k)})| < \epsilon.$$

Valores típicos para ϵ são 10^{-6} ou 10^{-8}.

B.4

O algoritmo pode travar quando o Hessiano está quase singular porque a matriz $[H(\beta^{(k)})]^{-1}$ fica numericamente instável. Pequenas mudanças no gradiente podem gerar passos muito grandes. Isso pode ocorrer em casos de colinearidade forte, separação quase perfeita ou regiões muito planas da log-verossimilhança.

Possíveis soluções incluem: mudar o ponto inicial, usar passo reduzido, usar métodos quasi-Newton, acrescentar regularização numérica ou verificar se há separação perfeita.

Parte C Logit, Probit e normalizações do erro latente

C.1

No logit:

$$F(z) = \Lambda(z) = \frac{1}{1 + e^{-z}}.$$

A derivada é:

$$f(z) = \Lambda(z)[1 - \Lambda(z)].$$

Como $F_i = \Lambda(z_i)$ e $f_i = F_i(1 - F_i)$, o score genérico:

$$\sum_{i=1}^{n} \frac{(y_i - F_i)f_i}{F_i(1 - F_i)} w_i$$

se simplifica para:

$$\boxed{\nabla_\beta \ln L(\beta) = \sum_{i=1}^{n} (y_i - F_i) w_i}.$$

O Hessiano do logit é:

$$\boxed{H(\beta) = -\sum_{i=1}^{n} F_i(1 - F_i) w_i w_i'}.$$

Esse Hessiano é negativo semidefinido, pois $F_i(1 - F_i) > 0$ e $w_i w_i'$ é positivo semidefinido.

C.2

No probit:

$$F(z) = \Phi(z), \qquad f(z) = \phi(z).$$

Assim:

$$F_i = \Phi(z_i), \qquad f_i = \phi(z_i).$$

O score é:

$$\boxed{\nabla_\beta \ln L(\beta) = \sum_{i=1}^{n} \frac{(y_i - \Phi_i)\phi_i}{\Phi_i(1 - \Phi_i)} w_i}.$$

O Hessiano pode ser escrito na forma:

$$H(\beta) = \sum_{i=1}^{n} q_i' w_i w_i',$$

onde:

$$q_i' = y_i \left[\frac{-z_i \phi_i}{\Phi_i} - \frac{\phi_i^2}{\Phi_i^2} \right] - (1 - y_i) \left[\frac{-z_i \phi_i}{1 - \Phi_i} + \frac{\phi_i^2}{(1 - \Phi_i)^2} \right].$$

C.3

A relação aproximada:

$$\hat{\beta}^{\text{logit}} \approx 1{,}6 \hat{\beta}^{\text{probit}}$$

ocorre porque a distribuição logística e a normal padrão têm formatos muito parecidos, mas escalas diferentes. Uma forma simples de ver isso é comparar as inclinações das cdfs no ponto zero.

No logit:

$$\Lambda'(0) = \Lambda(0)[1 - \Lambda(0)] = \frac{1}{2}\frac{1}{2} = \frac{1}{4} = 0{,}25.$$

No probit:

$$\Phi'(0) = \phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0{,}399.$$

A razão é:

$$\frac{0{,}399}{0{,}25} \approx 1{,}6.$$

Portanto, para gerar variações semelhantes nas probabilidades, os coeficientes do logit costumam ser maiores que os coeficientes do probit.

C.4

Embora os coeficientes de logit e probit estejam em escalas diferentes, as probabilidades estimadas e os efeitos marginais frequentemente são muito parecidos. Isso ocorre porque as cdfs logística e normal padrão têm formatos semelhantes na região central, onde se concentra grande parte dos dados. Assim, após ajustar a escala dos coeficientes, os dois modelos tendem a produzir probabilidades próximas.

C.5

No logit:

$$p_i = \Pr(y_i = 1 \mid x_i) = \frac{e^{\beta_1+\beta_2 x_i}}{1 + e^{\beta_1+\beta_2 x_i}}.$$

Logo:

$$1 - p_i = \frac{1}{1 + e^{\beta_1+\beta_2 x_i}}.$$

A razão de chances é:

$$\frac{p_i}{1 - p_i} = e^{\beta_1+\beta_2 x_i}.$$

Tomando logaritmo:

$$\boxed{\ln \frac{\Pr(y_i = 1 \mid x_i)}{\Pr(y_i = 0 \mid x_i)} = \beta_1 + \beta_2 x_i}.$$

Portanto, se x_i aumenta uma unidade, a razão de chances é multiplicada por:

$$e^{\beta_2}.$$

Essa interpretação é particularmente vantajosa em áreas como epidemiologia, saúde, sociologia e avaliação de risco, nas quais *odds ratios* são frequentemente usados.

C.6

Se:

$$E(u_i) = \mu,$$

podemos escrever:

$$u_i = (u_i - \mu) + \mu.$$

Logo:

$$y_i^* = \beta_1 + \beta_2 x_i + u_i = \beta_1 + \beta_2 x_i + \mu + (u_i - \mu).$$

Assim:

$$\boxed{y_i^* = (\beta_1 + \mu) + \beta_2 x_i + (u_i - \mu)}.$$

Portanto, a média de u_i é absorvida pelo intercepto. Na presença de constante, não é possível separar β_1 de μ. Assumir $E(u_i) = 0$ é apenas uma normalização de localização.

C.7

A presença de constante não é suficiente para fixar:

$$\mathrm{Var}(u_i) = 1.$$

A constante desloca a localização da variável latente, mas não fixa sua escala. Em modelos binários, observamos apenas:

$$y_i = 1[y_i^* > 0],$$

ou seja, apenas se a variável latente está acima ou abaixo do limiar. Não observamos a escala de y_i^*.

C.8

Se:

$$u_i = \sigma\varepsilon_i,$$

então:

$$y_i^* = \beta_1 + \beta_2 x_i + \sigma\varepsilon_i.$$

Logo:

$$y_i = 1 \iff \beta_1 + \beta_2 x_i + \sigma\varepsilon_i > 0.$$

Portanto:

$$\varepsilon_i > -\frac{\beta_1 + \beta_2 x_i}{\sigma}.$$

Como a distribuição é simétrica:

$$\Pr(y_i = 1 \mid x_i) = F_\varepsilon\left(\frac{\beta_1 + \beta_2 x_i}{\sigma}\right).$$

Assim, os dados identificam apenas:

$$\frac{\beta_1}{\sigma} \qquad \text{e} \qquad \frac{\beta_2}{\sigma}.$$

Logo, β e σ não são separadamente identificados. No probit, impõe-se:

$$\mathrm{Var}(u_i) = 1.$$

No logit, a escala é fixada pela distribuição logística padrão, cuja variância é:

$$\frac{\pi^2}{3}.$$

Parte D Exercício numérico: logit com regressor binário

A tabela é:

	$Y = 0$	$Y = 1$
$X = 0$	30	10
$X = 1$	20	40

D.1

Defina:

$$p_0 = \Pr(Y_i = 1 \mid X_i = 0),$$

e

$$p_1 = \Pr(Y_i = 1 \mid X_i = 1).$$

No modelo logit:

$$\Pr(Y_i = 1 \mid X_i) = \Lambda(\beta_0 + \beta_1 X_i).$$

Se $X_i = 0$:

$$p_0 = \Lambda(\beta_0).$$

Se $X_i = 1$:

$$p_1 = \Lambda(\beta_0 + \beta_1).$$

Portanto:

$$\boxed{p_0 = \Lambda(\beta_0) \qquad \text{e} \qquad p_1 = \Lambda(\beta_0 + \beta_1)}.$$

D.2

Para o grupo controle, $X = 0$, há 10 sucessos e 30 fracassos. A contribuição para a verossimilhança é:

$$p_0^{10}(1-p_0)^{30}.$$

Para o grupo tratado, $X = 1$, há 40 sucessos e 20 fracassos. A contribuição é:

$$p_1^{40}(1-p_1)^{20}.$$

Logo:

$$\boxed{L(\beta_0, \beta_1) = p_0^{10}(1-p_0)^{30}p_1^{40}(1-p_1)^{20}}.$$

A log-verossimilhança é:

$$\boxed{\ln L(\beta_0, \beta_1) = 10\ln p_0 + 30\ln(1-p_0) + 40\ln p_1 + 20\ln(1-p_1)}.$$

D.3

Como X é binária, o modelo estima uma probabilidade para cada grupo. O EMV de cada probabilidade é a proporção amostral de sucessos no respectivo grupo.

Para $X = 0$:

$$\hat{p}_0 = \frac{10}{10+30} = \frac{10}{40} = 0{,}25.$$

Para $X = 1$:

$$\hat{p}_1 = \frac{40}{40+20} = \frac{40}{60} = \frac{2}{3} \approx 0{,}667.$$

Logo:

$$\boxed{\hat{p}_0 = 0{,}25 \qquad \text{e} \qquad \hat{p}_1 = \frac{2}{3}}.$$

D.4

A função logito é:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right).$$

Como:

$$p_0 = \Lambda(\beta_0),$$

temos:

$$\beta_0 = \text{logit}(p_0).$$

Logo:

$$\hat{\beta}_0 = \ln\left(\frac{\hat{p}_0}{1-\hat{p}_0}\right) = \ln\left(\frac{0{,}25}{0{,}75}\right) = \ln\left(\frac{1}{3}\right) \approx -1{,}099.$$

Além disso:

$$\beta_0 + \beta_1 = \text{logit}(p_1).$$

Portanto:

$$\beta_1 = \text{logit}(p_1) - \text{logit}(p_0).$$

Assim:

$$\hat{\beta}_1 = \ln\left(\frac{2/3}{1/3}\right) - \ln\left(\frac{1/4}{3/4}\right).$$

Como:

$$\frac{2/3}{1/3} = 2,$$

e

$$\frac{1/4}{3/4} = \frac{1}{3},$$

temos:

$$\hat{\beta}_1 = \ln(2) - \ln(1/3) = \ln(6) \approx 1{,}792.$$

Portanto:

$$\boxed{\hat{\beta}_0 \approx -1{,}099 \qquad \text{e} \qquad \hat{\beta}_1 \approx 1{,}792}.$$

D.5

O coeficiente $\hat{\beta}_1$ é positivo:

$$\hat{\beta}_1 = 1{,}792 > 0.$$

Logo, o tratamento está associado a maior probabilidade de sucesso.

A razão de chances é:

$$e^{\hat{\beta}_1} = e^{\ln(6)} = 6.$$

Portanto:

$$\boxed{e^{\hat{\beta}_1} = 6}.$$

Isso significa que a chance de sucesso no grupo tratado é 6 vezes a chance de sucesso no grupo controle.

De fato, no grupo controle:

$$\text{odds}_0 = \frac{0{,}25}{0{,}75} = \frac{1}{3}.$$

No grupo tratado:

$$\text{odds}_1 = \frac{2/3}{1/3} = 2.$$

Logo:

$$\frac{\text{odds}_1}{\text{odds}_0} = \frac{2}{1/3} = 6.$$

D.6

O efeito marginal discreto do tratamento é:

$$\widehat{\Delta} = \hat{p}_1 - \hat{p}_0.$$

Logo:

$$\widehat{\Delta} = \frac{2}{3} - \frac{1}{4} = \frac{8}{12} - \frac{3}{12} = \frac{5}{12} \approx 0{,}417.$$

Portanto:

$$\boxed{\widehat{\Delta} \approx 41{,}7 \text{ pontos percentuais}}.$$

Como X é binária, não faz sentido interpretar uma derivada infinitesimal em relação a X. O mais adequado é comparar a probabilidade prevista quando $X = 1$ com a probabilidade prevista quando $X = 0$.

D.7

A log-verossimilhança maximizada do modelo com X é:

$$\ln \hat{L} = 10 \ln(0{,}25) + 30 \ln(0{,}75) + 40 \ln(2/3) + 20 \ln(1/3).$$

Numericamente:

$$\ln \hat{L} \approx -60{,}684.$$

No modelo apenas com constante:

$$\hat{p} = \frac{50}{100} = 0{,}5.$$

Assim:

$$\ln \hat{L}_0 = 50 \ln(0{,}5) + 50 \ln(0{,}5) = 100 \ln(0{,}5).$$

Logo:

$$\ln \hat{L}_0 \approx -69{,}315.$$

O pseudo-R^2 de McFadden é:

$$\widetilde{R}^2 = 1 - \frac{\ln \hat{L}}{\ln \hat{L}_0}.$$

Substituindo:

$$\widetilde{R}^2 = 1 - \frac{-60{,}684}{-69{,}315} \approx 1 - 0{,}8755 = 0{,}1245.$$

Portanto:

$$\boxed{\widetilde{R}^2 \approx 0{,}125}.$$

Esse pseudo-R^2 mede o ganho relativo de log-verossimilhança em relação ao modelo apenas com constante. Ele não deve ser interpretado como a proporção da variância de Y explicada pelo modelo, como no R^2 da regressão linear.

D.8

Para o modelo com X, temos $k = 2$ e:

$$\text{AIC} = -2 \ln \hat{L} + 2k.$$

Logo:

$$\text{AIC}_X = -2(-60{,}684) + 2(2) = 121{,}368 + 4 = 125{,}368.$$

Para o modelo apenas com constante, $k = 1$:

$$\text{AIC}_0 = -2(-69{,}315) + 2(1) = 138{,}630 + 2 = 140{,}630.$$

Logo:

$$\boxed{\text{AIC}_X \approx 125{,}37 < \text{AIC}_0 \approx 140{,}63}.$$

O modelo com X apresenta menor AIC, sugerindo melhor ajuste mesmo após penalizar o parâmetro adicional.

D.9

A matriz de informação observada é:

$$\widehat{\mathcal{I}} = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} \begin{pmatrix} 1 & X_i \end{pmatrix}.$$

No grupo $X = 0$:

$$\hat{p}_0(1-\hat{p}_0) = 0{,}25(0{,}75) = 0{,}1875.$$

Como há 40 observações:

$$40\hat{p}_0(1-\hat{p}_0) = 40(0{,}1875) = 7{,}5.$$

No grupo $X = 1$:

$$\hat{p}_1(1-\hat{p}_1) = \frac{2}{3}\frac{1}{3} = \frac{2}{9}.$$

Como há 60 observações:

$$60\hat{p}_1(1-\hat{p}_1) = 60\frac{2}{9} = \frac{120}{9} = \frac{40}{3} \approx 13{,}333.$$

Assim:

$$\widehat{\mathcal{I}} = 7{,}5 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 13{,}333 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Logo:

$$\widehat{\mathcal{I}} = \begin{pmatrix} 20{,}833 & 13{,}333 \\ 13{,}333 & 13{,}333 \end{pmatrix}.$$

Em frações:

$$\widehat{\mathcal{I}} = \begin{pmatrix} 125/6 & 40/3 \\ 40/3 & 40/3 \end{pmatrix}.$$

A matriz de variância-covariância é a inversa:

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \widehat{\mathcal{I}}^{-1}.$$

Calculando:

$$\boxed{\widehat{\mathrm{Var}}(\hat{\beta}) = \begin{pmatrix} 0{,}133 & -0{,}133 \\ -0{,}133 & 0{,}208 \end{pmatrix}}.$$

Em frações:

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \begin{pmatrix} 2/15 & -2/15 \\ -2/15 & 5/24 \end{pmatrix}.$$

Logo:

$$se(\hat{\beta}_0) = \sqrt{0{,}133} \approx 0{,}365,$$

e

$$se(\hat{\beta}_1) = \sqrt{0{,}208} \approx 0{,}456.$$

D.10

As estatísticas de Wald são:

$$z_{\hat{\beta}_0} = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)} = \frac{-1{,}099}{0{,}365} \approx -3{,}01.$$

E:

$$z_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{1{,}792}{0{,}456} \approx 3{,}93.$$

Portanto:

$$\boxed{z_{\hat{\beta}_0} \approx -3{,}01 \qquad \text{e} \qquad z_{\hat{\beta}_1} \approx 3{,}93}.$$

Ambas as estimativas são estatisticamente diferentes de zero aos níveis usuais de significância. Em particular, $|\hat{z}_{\beta_1}| = 3{,}93 > 1{,}96$, logo rejeita-se $H_0 : \beta_1 = 0$ ao nível de 5%.

D.11

O método Delta serve para aproximar a variância de uma função não linear de estimadores. Se:

$$\hat{\theta}$$

é um estimador assintoticamente normal e queremos a variância de:

$$g(\hat{\theta}),$$

então o método Delta usa uma aproximação linear de primeira ordem:

$$g(\hat{\theta}) \approx g(\theta) + G(\theta)(\hat{\theta} - \theta),$$

onde $G(\theta)$ é o gradiente de g.

Assim:

$$\text{Var}[g(\hat{\theta})] \approx G(\hat{\theta})\widehat{\text{Var}}(\hat{\theta})G(\hat{\theta})'.$$

Ele é útil para calcular erros-padrão de objetos como:

$$e^{\hat{\beta}_1},$$

$$\hat{p}_1 - \hat{p}_0,$$

ou efeitos marginais, porque todos eles são funções não lineares dos coeficientes estimados.

Não se deve simplesmente aplicar a mesma transformação aos erros-padrão dos coeficientes porque:

$$se[g(\hat{\theta})] \neq g[se(\hat{\theta})].$$

A transformação deve ser aplicada ao estimador, enquanto a incerteza deve ser propagada usando a derivada da função e toda a matriz de variância-covariância dos parâmetros.

Questão 3 Modelos Multinomial e Ordinal

Parte A Logit multinomial e utilidade aleatória

A.1

Um modelo de utilidade aleatória parte da ideia de que o indivíduo escolhe a alternativa que lhe dá maior utilidade. Para o indivíduo, a escolha pode ser perfeitamente determinística: ele compara as alternativas e escolhe a melhor. Para o econometrista, entretanto, parte da utilidade não é observada. Por isso, escrevemos:

$$U_{ij} = x'_{ij}\beta + \varepsilon_{ij}.$$

O termo $x'_{ij}\beta$ representa a parte observável da utilidade, enquanto ε_{ij} representa fatores não observados pelo econometrista, como preferências individuais, informação privada, hábitos, conveniência ou características omitidas das alternativas.

Assim, a utilidade é aleatória do ponto de vista do econometrista, não necessariamente do ponto de vista do agente.

A.2

A fórmula do logit multinomial é:

$$\Pr(y_i = j \mid X_i) = \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^{J} \exp(x'_{ik}\beta)}.$$

As probabilidades são positivas porque:

$$\exp(x'_{ij}\beta) > 0$$

para qualquer valor de $x'_{ij}\beta$.

Além disso, elas somam 1 porque:

$$\sum_{j=1}^{J} \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^{J} \exp(x'_{ik}\beta)} = \frac{\sum_{j=1}^{J} \exp(x'_{ij}\beta)}{\sum_{k=1}^{J} \exp(x'_{ik}\beta)} = 1.$$

A distribuição dos erros importa porque ela determina a forma da probabilidade de uma alternativa gerar a maior utilidade. No logit multinomial, a forma fechada acima decorre da hipótese de erros tipo valor extremo i.i.d. Outras distribuições de erro geram outros modelos, como o probit multinomial.

A.3

A hipótese IIA, *Independence of Irrelevant Alternatives*, implica que a razão entre as probabilidades de escolha de duas alternativas quaisquer não depende das demais alternativas disponíveis.

No logit multinomial:

$$\frac{\Pr(y_i = j \mid X_i)}{\Pr(y_i = k \mid X_i)} = \frac{\exp(x'_{ij}\beta)}{\exp(x'_{ik}\beta)} = \exp[(x_{ij} - x_{ik})'\beta].$$

Essa razão não depende de nenhuma alternativa $\ell \neq j, k$.

O exemplo clássico é o ônibus vermelho/azul. Suponha que um indivíduo escolha entre carro e ônibus. Se surge uma nova alternativa, ônibus azul, muito parecida com o ônibus vermelho, seria natural que a maior parte da probabilidade da nova alternativa viesse do ônibus original, não do carro. A hipótese IIA, entretanto, força uma redistribuição proporcional das probabilidades entre todas as alternativas. Isso pode ser implausível quando há alternativas muito próximas entre si.

Parte B Modelos ordinais: ologit e oprobit

B.1

Em modelos ordinais, supõe-se uma variável latente contínua:

$$y_i^* = x'_i\beta + u_i.$$

A variável observada y_i é discreta e ordinal. Ela é determinada por limiares:

$$y_i = j \iff \kappa_{j-1} < y_i^* \leq \kappa_j.$$

Os limiares transformam a variável latente contínua em categorias discretas. Por exemplo, se y_i^* representa uma propensão latente à escolaridade, limiares diferentes separam categorias como fundamental, médio e superior.

B.2

Um modelo ordinal é preferível ao MQO porque y_i é categórica e ordinal, não contínua. O MQO trataria as distâncias entre categorias como iguais. Por exemplo, passaria a supor que a diferença entre fundamental e médio é igual à diferença entre médio e superior, o que pode não ser adequado.

Um modelo ordinal também pode ser preferível ao logit multinomial irrestrito porque aproveita a ordenação natural das categorias. O logit multinomial trataria as categorias como nominais, ignorando que superior está acima de médio e médio está acima de fundamental. O modelo ordinal é mais parcimonioso e usa essa informação de ordenação.

B.3

Se:

$$\hat{\beta}_2 > 0,$$

então aumentos em x_i deslocam a variável latente y_i^* para a direita. Isso tende a reduzir a probabilidade das categorias mais baixas e aumentar a probabilidade das categorias mais altas.

Entretanto, o efeito marginal pode ter sinais diferentes para categorias diferentes. Para a menor categoria, o efeito de x_i tende a ser negativo. Para a maior categoria, tende a ser positivo. Para categorias intermediárias, o sinal pode ser positivo ou negativo, pois aumentar x_i pode deslocar indivíduos tanto para dentro quanto para fora daquela categoria.

B.4

Se y_i representa escolaridade em três categorias – fundamental, médio e superior – há uma ordenação natural. Nesse caso, um modelo ordinal, como ologit ou oprobit, é geralmente adequado porque preserva a informação de que:

$$\text{fundamental} < \text{médio} < \text{superior}.$$

O logit multinomial seria mais apropriado se as categorias não tivessem ordem natural ou se quiséssemos permitir efeitos muito diferentes de x_i sobre cada categoria, sem impor estrutura ordinal. No caso da escolaridade, porém, a estrutura ordinal é substantiva e econometricamente útil.

Portanto, salvo evidência forte contra a hipótese ordinal, a escolha natural para esse exemplo é um modelo ordinal.