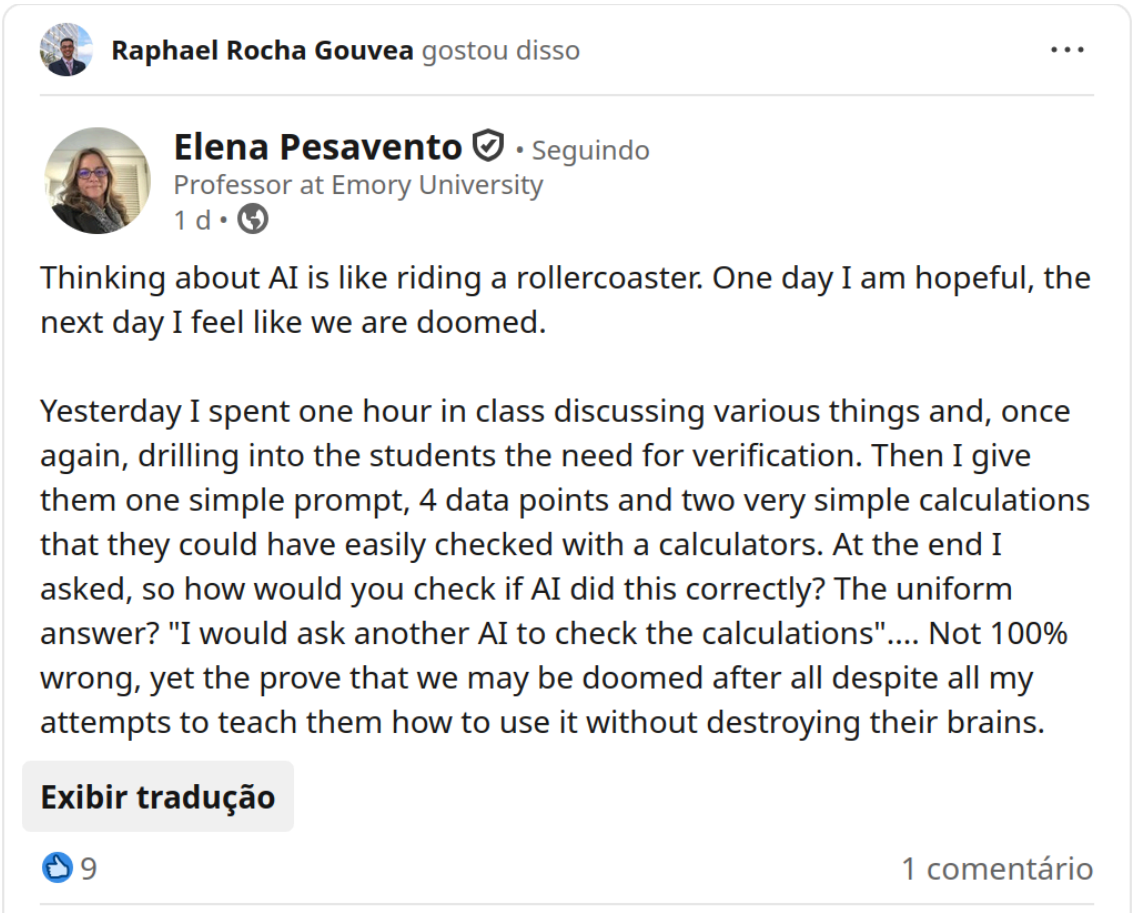
Variável dependente binária: modelos não lineares

IBM0288 - 2026.2

Prof. Raphael Gouvea

1 Para reflexão



2 Aula passada: regressão

- Nos modelos de MQO que estudamos anteriormente, a função de regressão populacional é dada por:

$$E(Y \mid X_1, X_2, ..., X_k)$$

- Quando $Y_i \in \{0, 1\}$, o valor esperado de Y é p, ou seja, a probabilidade de $Y = 1$:

$$E(Y) = 0 \times Pr(Y = 0) + 1 \times Pr(Y = 1) = Pr(Y = 1)$$

- Portanto, para uma variável binária Y, temos:

$$E(Y \mid X_1, X_2, ..., X_k) = Pr(Y = 1 \mid X_1, X_2, ..., X_k)$$

3 Aula passada: LPM

Um ponto de partida natural é o modelo de regressão linear (*Linear Probability Model - LPM*) com um único regressor:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- O que a reta $\beta_0 + \beta_1 X$ significa quando Y é binária?
  - $\Pr(Y = 1|X) = \beta_0 + \beta_1 X_i$
- O que β_1 significa quando Y é binária? $\beta_1 = \frac{\Delta Y}{\Delta X}$?
  - β_1 é a mudança nessa probabilidade para uma variação em X
  - os resultados devem ser interpretados como variação em pontos percentuais (p. p.)
- O que o valor predito $\hat{Y}$ significa quando Y é binária?
  - $\hat{Y}_i = \widehat{\Pr}(Y = 1|X) = \hat{\beta}_0 + \hat{\beta}_1 X_i$

4 Aula passada: vantagens e desvantagens

Vantagens:

- Simples de estimar e interpretar
- A inferência é a mesma da regressão múltipla (é necessário usar erros-padrão robustos à heterocedasticidade)
- Mesmas hipóteses em relação à causalidade e aos efeitos de tratamento

Desvantagens:

- O LPM supõe que a mudança na probabilidade predita é a mesma para todos os valores de X (linearidade!)
- Probabilidades preditas pelo LPM podem ser < 0 ou > 1!

5 O que precisamos?

- O problema com o modelo de probabilidade linear é que ele modela a probabilidade de $Y = 1$ como linear em X:

$$\Pr(Y = 1 \mid X) = \beta_0 + \beta_1 X$$

- Em vez disso, queremos:
  - que $\Pr(Y = 1 \mid X)$ seja crescente em X para $\beta_1 > 0$, e
  - que $0 \leq \Pr(Y = 1 \mid X) \leq 1$ para todo X
- Isso requer o uso de uma forma funcional não linear para a probabilidade.

6 Introdução aos modelos não lineares

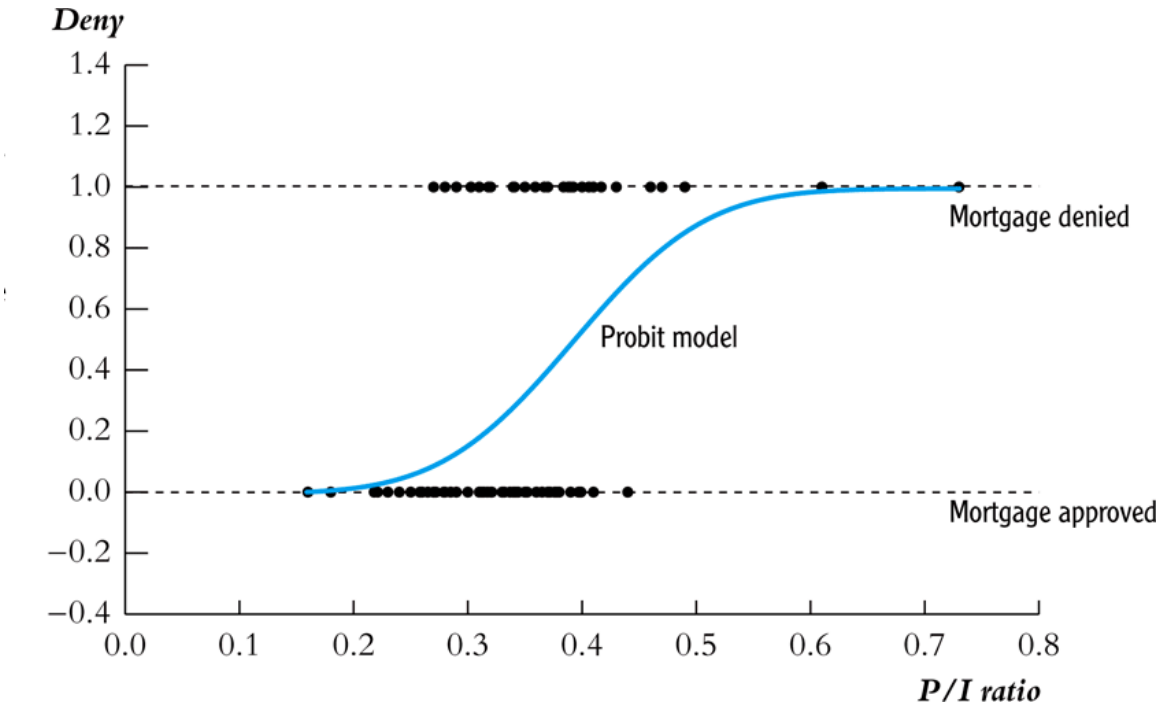
- A ideia é modelar:

$$P(Y_i = 1 \mid X_i) = F(\beta_0 + \beta_1 X_i)$$

  onde $F(\cdot)$ é uma **função de distribuição acumulada**.
- Assim garantimos que $0 \leq P \leq 1$.
- Duas escolhas principais:
  - $F = \Phi$ (CDF normal padrão): **Probit**
  - $F = \Lambda$ (CDF logística): **Logit**

7 Regressão Probit

- O modelo probit satisfaz essas condições:
  - $\Pr(Y = 1 \mid X)$ ser crescente em X para $\beta_1 > 0$, e
  - $0 \leq \Pr(Y = 1 \mid X) \leq 1$ para todo X



8 Probit e a Distribuição Normal

- A regressão probit modela a probabilidade de $Y = 1$ usando a função de distribuição acumulada da normal padrão, $\Phi(z)$, avaliada em $z = \beta_0 + \beta_1 X$:

$$\Pr(Y = 1 \mid X) = \Phi(\beta_0 + \beta_1 X)$$

- $z = \beta_0 + \beta_1 X$ é o "valor-z" ou "índice-z" do modelo probit.
- Exemplo: suponha $\beta_0 = -2$, $\beta_1 = 3$, $X = 0,4$, portanto

$$\Pr(Y = 1 \mid X = 0,4) = \Phi(-2 + 3 \times 0,4) = \Phi(-0,8)$$

- $\Pr(Y = 1 \mid X = 0,4)$ = área sob a curva da densidade normal padrão à esquerda de $z = -0,8$, que é...

9 Tabela da Normal

TABLE 1 *(continued)*

	Second Decimal Value of z									
z	0	1	2	3	4	5	6	7	8	9
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

This table can be used to calculate $\Pr(Z^* z)$ where Z is a standard normal variable. For example, when $z = 1.17$, this probability is 0.8790, which is the table entry for the row labeled 1.1 and the column labeled 7.

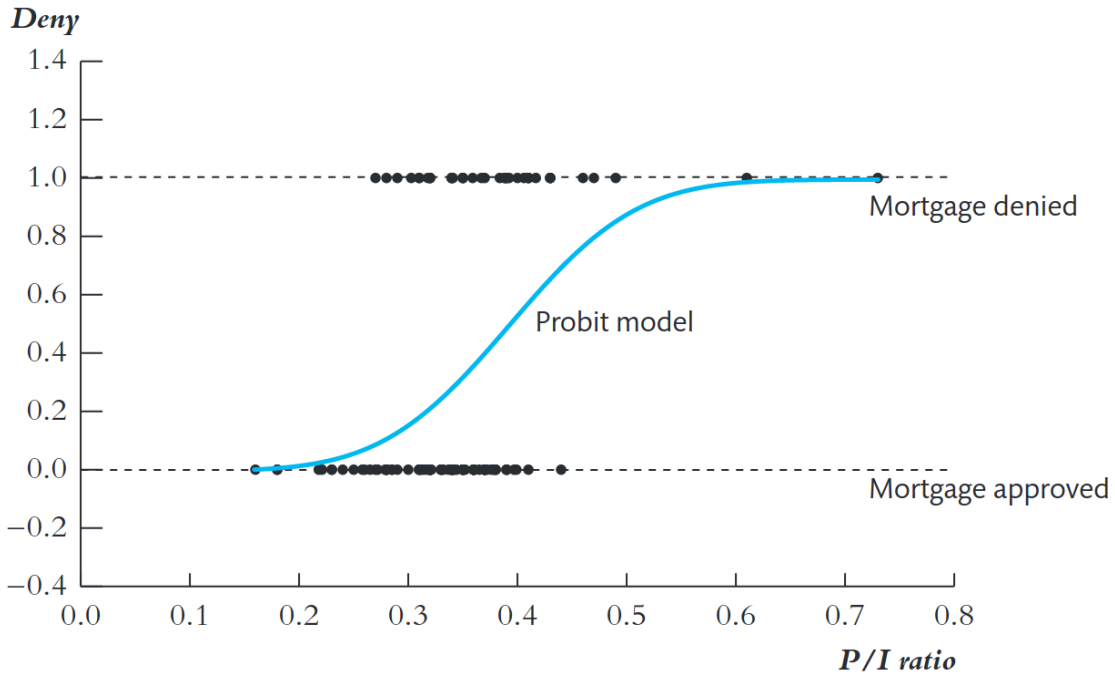
$$\Pr(z \leq -0,8) = 0,2119$$

10 Probit: o caso das hipotecas

$P(Y_i = 1|X_i) = \Phi(\beta_0 + \beta_1 P/I_i)$

$\Phi(\cdot)$: CDF normal padrão.

- Para $P/I = 0,2 \longrightarrow P(deny = 1) \approx 2,1\%$
- Para $P/I = 0,3 \longrightarrow P(deny = 1) \approx 16,1\%$
- Para $P/I = 0,4 \longrightarrow P(deny = 1) \approx 51,9\%$
- Para $P/I = 0,6 \longrightarrow P(deny = 1) \approx 98,3\%$



Interpretação: a probabilidade aumenta **lentamente para valores baixos** de P/I, **cresce rapidamente** para valores intermediários e **satura em 1** para valores elevados.

11 Regressão probit com múltiplos regressores

$$\Pr(Y = 1 \mid X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

- Φ é a função de distribuição acumulada da normal padrão.
- $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ é o "valor-z" ou "índice-z" do modelo probit.
- β_1 é o efeito sobre o valor-z de uma variação unitária em X_1, mantendo X_2 constante (quando uma interpretação causal é justificada).

12 Regressão Probit: resumo

- Por que usar a função de distribuição acumulada da normal padrão?
- A "forma em S" nos dá o que queremos:
  - $\Pr(Y = 1 \mid X)$ é crescente em X para $\beta_1 > 0$
  - $0 \leq \Pr(Y = 1 \mid X) \leq 1$ para todo X
- Fácil de usar: probabilidades tabeladas e facilmente obtidas no R
- A interpretação é relativamente simples:
  - $\beta_0 + \beta_1 X$ = valor-z
  - β_1 é a variação no valor-z para uma variação unitária em X

Ibmec

13 Regressão Logit

A regressão logit modela a probabilidade de $Y = 1$, dado X, como a função de distribuição acumulada da logística padrão, $\Lambda(z)$, avaliada em $z = \beta_0 + \beta_1 X$:

$$\Pr(Y = 1 \mid X) = \Lambda(\beta_0 + \beta_1 X)$$

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

! Probit vs Logit: valores dos coeficientes

Como logit e probit utilizam funções de distribuição diferentes, os coeficientes (β's) são diferentes nos dois modelos.

14 Exemplo

$$\Pr(Y = 1 \mid X) = \Lambda(\beta_0 + \beta_1 X)$$

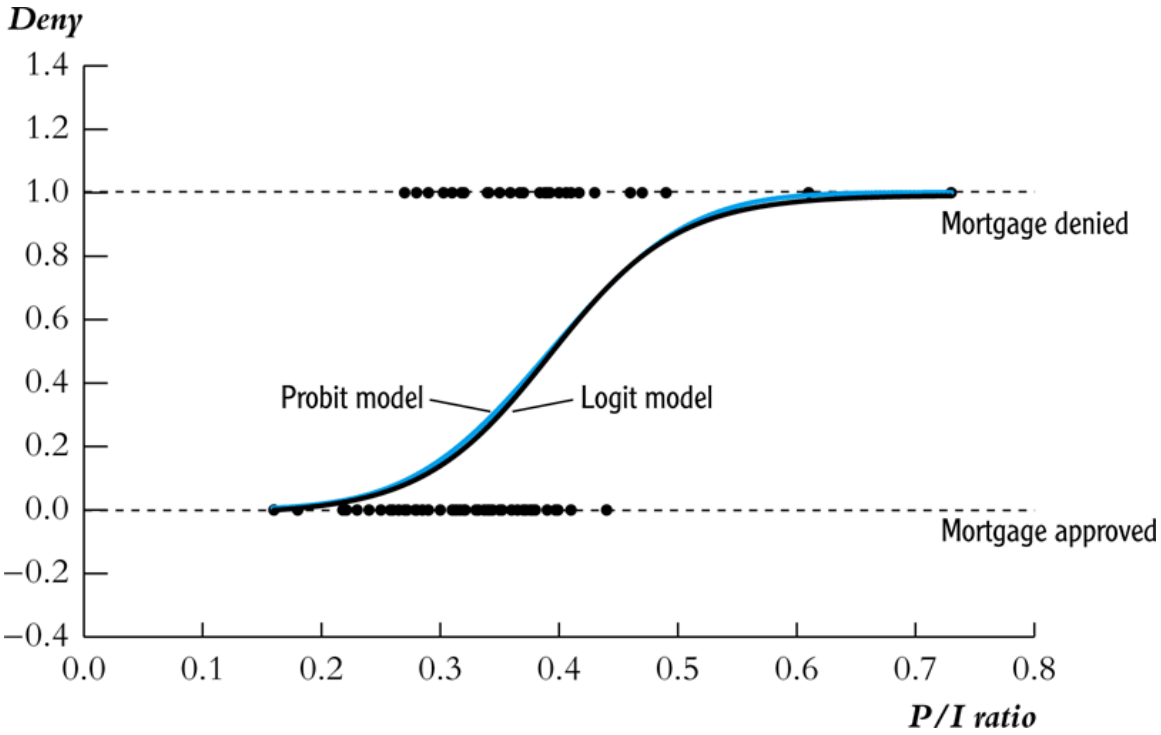
- Exemplo: $\beta_0 = -3$, $\beta_1 = 2$, $X = 0,4$,
  - portanto, $\beta_0 + \beta_1 X = -3 + 2 \times 0,4 = -2,2$, logo
  - $\Pr(Y = 1 \mid X = 0,4) = 1/(1 + e^{-(-2,2)}) = 0,0998$

15 Probit vs Logit

Por que usar logit se temos probit?

- A razão principal é histórica: o logit é computacionalmente mais rápido e simples, mas isso não importa mais hoje em dia.

Na prática, logit e probit são muito semelhantes!



Ibmec

16 Como interpretar os coeficientes?

- O coeficiente β_1 representa a diferença no valor de z associada ao aumento de X_1 em uma unidade, mantendo constantes as demais variáveis (X_2, ..., X_k).
- Como o modelo é **não linear**, o impacto sobre a probabilidade predita não é constante.
- Como calcular o efeito marginal aproximado da mudança de um regressor?
  1. Calcular a probabilidade predita para os valores iniciais;
  2. Calcular a probabilidade predita para o novo valor do regressor;
  3. Calcular a diferença entre as duas probabilidades preditas.
- A interpretação depende do ponto da curva: o mesmo incremento em X_1 pode gerar **efeitos diferentes** conforme o valor inicial.

17 Estimação por Máxima Verossimilhança: intuição

https://www.youtube.com/watch?v=uN7yIhXxHHk

https://www.youtube.com/watch?v=XepXtl9YKwc

18 Estimação por Máxima Verossimilhança: formalização

- Para variáveis binárias, a verossimilhança é:

$$L(\beta) = \prod_{i=1}^{n} [P_i]^{Y_i} [1 - P_i]^{1-Y_i}$$

  com: $P_i = F(\beta_0 + \beta_1 X_i)$
- Maximizamos o logaritmo:

$$\ln L(\beta) = \sum_i [Y_i \ln P_i + (1 - Y_i) \ln(1 - P_i)]$$

- Estimação feita numericamente (ex.: algoritmo Newton-Raphson).

19 Propriedades do estimador MLE

- **Consistente:** converge para o valor verdadeiro conforme $n \to \infty$.
- **Assintoticamente normal:**

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, \sigma^2)$$

- **Eficiente:** menor variância assintótica entre estimadores consistentes.

20 Erros-padrão e inferência

- Erros-padrão baseados na **matriz de informação de Fisher**:$\hat{V}(\hat{\beta}) = (-H)^{-1}$ onde H é a matriz Hessiana da log-verossimilhança.
- Intervalos de confiança: $\hat{\beta}_j \pm 1,96 \cdot SE(\hat{\beta}_j)$
- Testes:
  - Teste z individual.
  - Teste de razão de verossimilhança.

21 Medidas de ajuste

- O R^2 tradicional não é aplicável.
- Usamos o **pseudo-R^2** de McFadden:

$$R^2_{MF} = 1 - \frac{\ln L_{\text{modelo}}}{\ln L_{\text{nulo}}}$$

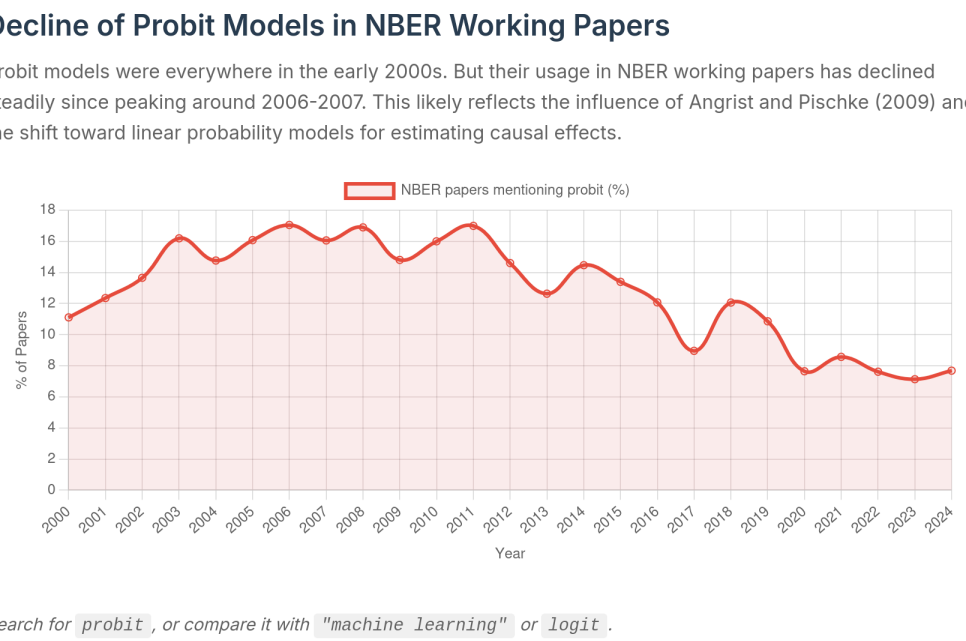
- Fração corretamente prevista

22 Existe discriminação racial no mercado de crédito imobiliário?

23 Angrist & Pischke: visão prática sobre Y binário

- **LPM**
  - Útil e suficiente para **inferência causal**;
  - Fácil de **estimar, interpretar e combinar** com VI, efeitos fixos, dif-in-dif;
  - Use **erros-padrão robustos**;
  - Previsões fora do intervalo $[0, 1]$ não invalidam a **consistência** do efeito médio quando há exogeneidade.
- **Logit/Probit**
  - Impõem probabilidades entre 0 e 1, mas os **efeitos marginais** são geralmente próximos aos do LPM.
  - A escolha da CDF muitas vezes é **secundária**: o foco deve estar no **desenho de pesquisa e na identificação**.

24 Declínio de modelos probit desde Angrist & Pischke



Fonte: Economics Literature Search

25 Guia Prático

Guia Prático

O cerne deve ser a identificação causal. A forma funcional não salva uma identificação ruim!

Prática recomendada: reporte o LPM com erros-padrão robustos e, se quiser, logit/probit para mostrar a robustez dos resultados.