Econometria Aplicada

Modelos de probabilidade (Probit e Logit)

João Ricardo Costa Filho

“The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

“In God we trust. All others must bring data."

William Edwards Deming

Variáveis binárias

Motivação 1 (tudo começa com uma pergunta)

O que influencia a participação feminina no mercado de trabalho?

Motivação 2 (tudo começa com uma pergunta)

Jim O'Neill levanta uma hipótese no seu texto "A Better Year for the Stock Market?": Se a bolsa dos EUA subir nos primeiros cinco pregões do ano, é provável que será um ano de alta. Será que isso é verdade? E será que vale para o Brasil? E para a Argentina?

Motivação 3 (tudo começa com uma pergunta)

Como as características de um cliente afetam a probabilidade de inadimplência de um financiamento de veículo?

Referências

- Como responder as questões que motivaram a nossa análise?

Referências

- Como responder as questões que motivaram a nossa análise?
  - Capítulos 7 e 17 de Wooldridge (2006).
  - Dados: Mroz (1987).

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.
- Mas, no caso da(s) pergunta(s) que motiva(m) esta aula, a variável dependente é **discreta**.

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.
- Mas, no caso da(s) pergunta(s) que motiva(m) esta aula, a variável dependente é **discreta**. E binária.

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.
- Mas, no caso da(s) pergunta(s) que motiva(m) esta aula, a variável dependente é **discreta**. E binária.
- Como podemos lidar com isso?

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.
- Mas, no caso da(s) pergunta(s) que motiva(m) esta aula, a variável dependente é **discreta**. E binária.
- Como podemos lidar com isso? Temos várias opções. Nesta aula, veremos três:

Tipos de variáveis

- Até o momento, trabalhamos com variáveis dependentes (Y) **contínuas**.
- Mas, no caso da(s) pergunta(s) que motiva(m) esta aula, a variável dependente é **discreta**. E binária.
- Como podemos lidar com isso? Temos várias opções. Nesta aula, veremos três:
  - **Modelo de Probabilidade Linear**
  - **Probit**
  - **Logit**

Modelos de Probabilidade

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X?

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X? Ou seja, $P\big(Y = 1|X\big) =?$.

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X? Ou seja, $P\left(Y = 1|X\right) =?$.

- No modelo de regressão linear, seria simplesmente: $P\left(Y = I|X_i\right) = \beta_0 + \beta_1 X + \epsilon$.

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X? Ou seja, $P\left(Y = 1|X\right) =?$.

- No modelo de regressão linear, seria simplesmente: $P\left(Y = I|X_i\right) = \beta_0 + \beta_1 X + \epsilon$.
- Para o caso multivariado: $P\left(Y = I|X_1, X_2, \ldots, X_k\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$.

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X? Ou seja, $P\big(Y = 1|X\big) =?$.

- No modelo de regressão linear, seria simplesmente: $P\big(Y = I|X_i\big) = \beta_0 + \beta_1 X + \epsilon$.
- Para o caso multivariado: $P\big(Y = I|X_1, X_2, \ldots, X_k\big) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$.

Qual é a alternativa?

Modelo de Probabilidade Linear

Qual é a probabilidade de $Y = I$, dados os valores de X? Ou seja, $P(Y = 1|X) = ?$.

- No modelo de regressão linear, seria simplesmente: $P(Y = I|X_i) = \beta_0 + \beta_1 X + \epsilon$.
- Para o caso multivariado: $P(Y = I|X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$.

Qual é a alternativa? Uma função **não-linear** ($G(\cdot)$) de $\beta_0 + \beta_1 X_1$ (ou de $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$)!

O estimador de máxima verossimilhança

O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$.

O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$. Ou seja, três sucessos.

O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$. Ou seja, três sucessos. Qual é a probabilidade de coletarmos **exatamente** essa amostra?

O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$. Ou seja, três sucessos. Qual é a probabilidade de coletarmos **exatamente** essa amostra?

$$L(p) = p \times (1 - p) \times (1 - p) \times p \times p = p^3(1 - p)^2$$

O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$. Ou seja, três sucessos. Qual é a probabilidade de coletarmos **exatamente** essa amostra?

$$L(p) = p \times (1 - p) \times (1 - p) \times p \times p = p^3(1 - p)^2$$

- Se $p = 0,1$, temos: $L(0,1) = 0,1^3(1 - 0,1)^2 \approx 0,00081$.

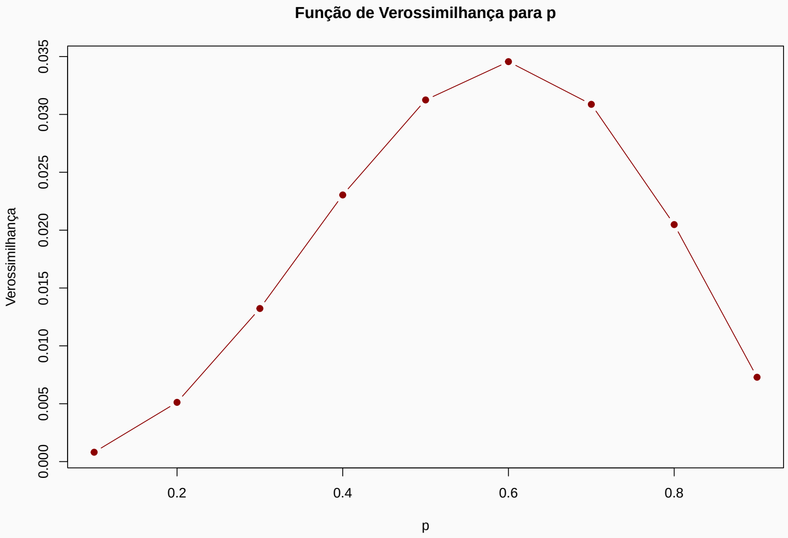
O estimador de máxima verossimilhança

Assuma que coletamos a seguinte amostra: $\{1, 0, 0, 1, 1\}$. Ou seja, três sucessos. Qual é a probabilidade de coletarmos **exatamente** essa amostra?

$$L(p) = p \times (1 - p) \times (1 - p) \times p \times p = p^3(1 - p)^2$$

- Se $p = 0,1$, temos: $L(0,1) = 0,1^3(1 - 0,1)^2 \approx 0,00081$.
- Se $p = 0,8$, temos: $L(0,8) = 0,8^3(1 - 0,8)^2 \approx 0,02048$.

O estimador de máxima verossimilhança



O estimador de máxima verossimilhança

$$L(p) = p^k (1-p)^{n-k},$$

onde k é o número de sucessos e n o tamanho da amostra. A condição de primeira ordem é dada por:

$$\frac{\partial L(p)}{\partial p} = k\, p^{k-1}(1-p)^{n-k} - (n-k)\, p^k (1-p)^{n-k-1} = 0 \iff .$$

$$p^{k-1}(1-p)^{n-k-1}\big[k(1-p) - (n-k)p\big] = 0.$$

O estimador de máxima verossimilhança

Assim,

$$k(1-p)-(n-k)p=0 \quad \Rightarrow \quad k-kp-np+kp=k-np=0.$$

Portanto,

$$\hat{p}_{MLE}=\frac{k}{n}.$$

O estimador de máxima verossimilhança

Alternativamente, podemos maximizar o log da função de verossimlinhaça:

$$\ell(p) = \ln L(p) = k \ln(p) + (n-k) \ln(1-p).$$

A condição de primeira ordem é dada por:

$$\frac{\partial \ell(p)}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0 \implies k(1-p) - (n-k)p = 0 \implies \hat{p}_{MLE} = \frac{k}{n}$$

Probit

- Vamos assumir que a função não-linear, $G\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$ é a função de distribuição acumulada da **Normal-padrão**.

Probit

- Vamos assumir que a função não-linear, $G\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$ é a função de distribuição acumulada da **Normal-padrão**.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $Z \sim N(0, 1)$.

Probit

- Vamos assumir que a função não-linear, $G\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$ é a função de distribuição acumulada da **Normal-padrão**.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $Z \sim N(0, 1)$.
    - Assim, $\phi(Z) = \big(2\pi\big)^{-1/2} exp\big(-Z^2/2\big)$

Probit

- Vamos assumir que a função não-linear, $G\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$ é a função de distribuição acumulada da **Normal-padrão**.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $Z \sim N(0, 1)$.
  - Assim, $\phi(Z) = \big(2\pi\big)^{-1/2} exp\big(-Z^2/2\big)$
- Como o modelo é não-linear, precisamos de outro método para estimar os parâmetros.

Probit

- Vamos assumir que a função não-linear, $G\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$ é a função de distribuição acumulada da **Normal-padrão**.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $Z \sim N(0, 1)$.
  - Assim, $\phi(Z) = \big(2\pi\big)^{-1/2} exp\big(-Z^2/2\big)$
- Como o modelo é não-linear, precisamos de outro método para estimar os parâmetros.
  - Utilizaremos o estimador de máxima verossimilhança.

Logit

- Vamos assumir que a função não-linear, $G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$ é a função de distribuição **logística** acumulada.

Logit

- Vamos assumir que a função não-linear, $G\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$ é a função de distribuição **logística** acumulada.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $G\big(Z\big) = \exp(Z)/\big(1 + \exp(Z)\big)$.

Logit

- Vamos assumir que a função não-linear, $G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$ é a função de distribuição **logística** acumulada.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $G(Z) = \exp(Z)/(1 + \exp(Z))$.
- Como o modelo é não-linear, precisamos de outro método para estimar os parâmetros.

Logit

- Vamos assumir que a função não-linear, $G\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$ é a função de distribuição **logística** acumulada.
- Portanto, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$, tal que $G\left(Z\right) = \exp(Z) / \left(1 + \exp(Z)\right)$.
- Como o modelo é não-linear, precisamos de outro método para estimar os parâmetros.
    - Utilizaremos o estimador de máxima verossimilhança.

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma.

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma. Por exemplo β_1 é o "efeito marginal" do aumento de X_1 em Y.

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma. Por exemplo β_1 é o "efeito marginal" do aumento de X_1 em Y.
- Agora, como no **Probit** e no **Logit** as funções são não-lineares, os parâmetros não têm o mesmo significado.

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma. Por exemplo β_1 é o "efeito marginal" do aumento de X_1 em Y.
- Agora, como no **Probit** e no **Logit** as funções são não-lineares, os parâmetros não têm o mesmo significado.
- A função de "efeito marginal" é dada por:
  - efeito marginal $= \beta_1 \times g\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma. Por exemplo β_1 é o "efeito marginal" do aumento de X_1 em Y.
- Agora, como no **Probit** e no **Logit** as funções são não-lineares, os parâmetros não têm o mesmo significado.
- A função de "efeito marginal" é dada por:
  - efeito marginal $= \beta_1 \times g\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$
  - O que isso significa?

Interpretação

- No modelo de **Modelo de Probabilidade Linear**, a interpretação dos coeficientes segue a mesma. Por exemplo β_1 é o "efeito marginal" do aumento de X_1 em Y.
- Agora, como no **Probit** e no **Logit** as funções são não-lineares, os parâmetros não têm o mesmo significado.
- A função de "efeito marginal" é dada por:
  - efeito marginal $= \beta_1 \times g\big(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\big)$
  - O que isso significa?
  - Que temos que avaliar o efeito marginal "ponto-a-ponto" (e.g. na média das variáveis da amostra).

Efeitos marginais

Efeitos marginais

- Probit: $\frac{\partial \hat{y}}{\partial x_j} = \hat{\beta}_j \cdot \phi(x\hat{\beta})$.

Efeitos marginais

- Probit: $\frac{\partial \hat{y}}{\partial x_j} = \hat{\beta}_j \cdot \phi(x\hat{\beta})$.
- Logit: $\frac{\partial \hat{y}}{\partial x_j} = \hat{\beta}_j \cdot \lambda(x\hat{\beta})$.

Efeitos marginais

- Probit: $\frac{\partial \hat{y}}{\partial x_j} = \hat{\beta}_j \cdot \phi(x\hat{\beta})$.
- Logit: $\frac{\partial \hat{y}}{\partial x_j} = \hat{\beta}_j \cdot \lambda(x\hat{\beta})$.
- Efeitos marginais na média (*Partial effects at the average*): $PEA = \hat{\beta}_j \cdot g(\bar{x}\hat{\beta})$
- Média dos efeitos marginais (*Average partial effects*): $APE = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_j \cdot g(x_i\hat{\beta}) = \hat{\beta}_j \cdot \overline{g(x\hat{\beta})}$

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.
- Mas podemos definir uma estatística análoga: o Pseudo R^2.

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.
- Mas podemos definir uma estatística análoga: o Pseudo R^2.
  - Intuição:

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.
- Mas podemos definir uma estatística análoga: o Pseudo R^2.
  - Intuição: (i) rodo um modelo sem restrições, calculo as estatísticas dele;

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.
- Mas podemos definir uma estatística análoga: o Pseudo R^2.
    - Intuição: (i) rodo um modelo sem restrições, calculo as estatísticas dele; (ii) depois, rodo um modelo com restrições (removo as variáveis e mantenho apenas o intercepto) e calculo as estatísticas dele.

Pseudo R^2

- Não podemos utilizar o R^2 para os modelos Probit e Logit.
- Mas podemos definir uma estatística análoga: o Pseudo R^2.
  - Intuição: (i) rodo um modelo sem restrições, calculo as estatísticas dele; (ii) depois, rodo um modelo com restrições (removo as variáveis e mantenho apenas o intercepto) e calculo as estatísticas dele. (iii) Finalmente, comparo as estatísticas.

Vamos aos dados!

Como poderíamos melhorar os modelos?

Como poderíamos melhorar os modelos?

Nesse caso, como calcularíamos as probabilidades?

Vamos para a atividade em grupo!

Mroz, Thomas A. 1987. "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." *Econometrica* 55 (4): 765–99. https://doi.org/10.2307/1911038.

Wooldridge, Jeffrey M. 2006. *Introdução à Econometria: Uma Abordagem Moderna*. Cengage Learning.