Econometria I

Modelos de Resposta Binária

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024

Sumário I

Modelos Não Lineares

- Em muitas aplicações que nos interessam, a variável dependente y é discreta e assume um número pequeno de valores.
- Quando analisamos decisões de pessoas, domicílios ou firmas, a variável relevante é se aquele agente econômico tomou uma decisão ou não, ou se tomou a decisão A, B ou C.
- Estamos interessados em saber:
  - se uma pessoa decidiu cometer um crime ou não
  - se em uma família a esposa trabalha ou não
  - se uma firma decide oferecer contratos de trabalho com salário fixo ou variável
  - se uma firma decide exportar ou não
  - se um prefeito se reelege ou não.

- Em alguns casos, podemos aproximar $\mathbb{E}[y|x]$ com um modelo linear. Porém, em outros casos, essa aproximação não será adequada.
- Temos uma variável dependente y que assume valores 0 ou 1.
- Podemos escrever a probabilidade condicional de y como:

$$P(y = 1|x) = F(x\beta) \tag{1}$$

- $F(.)$ é uma função de probabilidade acumulada.
    - $P(y = 1|x) \to 1$ quando $x\beta \to \infty$
    - $P(y = 0|x) \to 0$ quando $x\beta \to -\infty$
- Observe que $0 < F(x) < 1$ garante que as probabilidades de resposta estimadas estejam estritamente entre zero e um.

Modelo de Probabilidade Linear

- Uma abordagem simples para estimar o efeito de x na probabilidade de escolha da alternativa A é o modelo de regressão linear.
- A regressão por MQO de y sobre x fornece estimativas consistentes dos efeitos marginais médios amostrais dos regressores x sobre a probabilidade de escolher a alternativa A.
- Como resultado, o modelo linear é muito útil para fins exploratórios. Por exemplo, fornece um bom guia para quais variáveis são estatisticamente significativas e em qual sinal está seu efeito.

- No entanto, o modelo de probabilidade linear tem duas desvantagens importantes para a análise de resultados binários.
    1. A primeira é que as probabilidades previstas $\widehat{p}(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$ não são limitadas entre zero e um.
    2. A segunda desvantagem é que o termo de erro é heterocedástico e (dado $\boldsymbol{x}$) tem suporte discreto. Em particular, $u = -\boldsymbol{x}'\boldsymbol{\beta}$ se $y = 0$ e $u = 1 - \boldsymbol{x}'\boldsymbol{\beta}$ se $y = 1$; e sua variância é $\boldsymbol{x}'\boldsymbol{\beta}(1 - \boldsymbol{x}'\boldsymbol{\beta})$, que depende de x.

Variância Heterocedástica

- Considere o modelo

$$y_i = x_i'\beta + u_i \tag{2}$$

$$\mathbb{E}[u_i] = 0 \tag{3}$$

- Temos que:

$$\mathbb{E}[y_i|x_i] = \{(1) \times P(y_i = 1)\} + \{(0) \times P(y_i = 0)\} = P(y_i = 1) \tag{4}$$

Podemos deduzir que $\mathbb{E}[y_i|x_i] = x_i'\beta$ é quase um modelo de regressão linear padrão e que $\widehat{y}_i = x_i'\widehat{\beta} = \widehat{P}(y_i = 1)$.

- A variável dependente pode assumir dois valores: 1 ou zero. Portanto, o erro de regressão também pode assumir dois valores.

Tabela 1: Tabela para os erros

Valores	u_i	Probabilidade
$y_i = 1$	$1 - x_i'\beta$	$x_i'\beta = P(y_i = 1) = P_i$
$y_i = 0$	$-x_i'\beta$	$1 - x_i'\beta = P(y_i = 0) = 1 - P_i$

- Vamos ver que a variância do termo de erro é heterocedástica.

$$
\begin{aligned}
\text{var}[u_i] &= \mathbb{E}[(u_i - \mathbb{E}(u_i))^2] \\
&= \mathbb{E}[u_i]^2 \\
&= [u_i|y_i = 1]^2(P[u_i|y_i = 1]) + [u_i|y_i = 0]^2(P[u_i|y_i = 0]) \\
&= (1 - x_i'\beta)^2 \times P_i + (-x_i'\beta)^2 \times (1 - P_i) \\
&= (1 - x_i'\beta)^2 \times (x_i'\beta) + (-x_i'\beta)^2 \times (1 - x_i'\beta) \\
&= (1 - x_i'\beta) \times (x_i'\beta) \left[(1 - x_i'\beta) + x_i'\beta\right] \\
&= (1 - x_i'\beta) \times (x_i'\beta) \left[1 - x_i'\beta + x_i'\beta\right] \\
&= (1 - x_i'\beta) \times (x_i'\beta) \times (1) \\
&= (1 - x_i'\beta) \times (x_i'\beta) \qquad (5)
\end{aligned}
$$

Então, a var$[u_i]$ é heterocedástica. Como resolver esse problema? Podemos aplicar Mínimos Quadrados Generalizados ou usar algum dos erros robustos.

Modelo Geral de Escolha Discreta

- A probabilidade condicional de escolher a alternativa A dado $\boldsymbol{x}$ é

$$p(\boldsymbol{x}) \equiv \Pr[y = 1|\boldsymbol{x}] = F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) \tag{6}$$

- $F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)$ é uma função especificada de $\boldsymbol{x}'\boldsymbol{\beta}$. Os modelos dessa classe são chamados de modelos de índice único, pois o argumento da função de probabilidade condicional $F(\cdot)$ é um índice único de regressores, $\boldsymbol{x}'\boldsymbol{\beta}$.
- No modelo de probabilidade linear, a função especificada $F(\cdot)$ é a identidade, ou seja, $F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \boldsymbol{x}'\boldsymbol{\beta}$.
- Nos demais casos, porém, é natural especificar $F(\cdot)$ como uma função de distribuição acumulada para garantir que $0 \leq p \leq 1$.

- É o caso do Logit e do Probit, que assumem, respectivamente, a FDA logística e normal:

$$\text{Logit:} \qquad F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \frac{\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}}{1+\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}} \tag{7}$$

$$\text{Probit:} \qquad F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \Phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \int_{-\infty}^{\boldsymbol{x}'\boldsymbol{\beta}} \phi(z)dz \tag{8}$$

Estimação por MV

- Assumimos uma amostra de N observações independentes de $\{y_i, \boldsymbol{x}_i\}_{i=1}^{N}$.
- Dada a natureza binomial dos dados e a suposição de independência entre as observações, uma característica muito conveniente do modelo binomial é que a distribuição do resultado é conhecida: a distribuição de Bernoulli.
- A escolha de 0 e 1 para os valores da variável de resultado nos permite escrever a função de massa de probabilidade de uma forma muito compacta:

$$g(y|\boldsymbol{x}) = p^y(1-p)^{1-y} = \begin{cases} p, & \text{se } y = 1 \\ 1-p, & \text{se } y = 0 \end{cases} \tag{9}$$

em que $p = F(\boldsymbol{x}'\boldsymbol{\beta})$.

- Portanto, a função de log-verossimilhança é dada por

$$L_N(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left\{ y_i \ln F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) + (1 - y_i) \ln\left(1 - F\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right) \right\} \quad (10)$$

- O estimador de MV é obtido da CPO do problema de otimização acima

$$\begin{aligned} \frac{\partial L_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{N} \left\{ \frac{y_i}{F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right)} f\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right) \boldsymbol{x}_i - \frac{1 - y_i}{1 - F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right)} f\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right) \boldsymbol{x}_i \right\} \\ &= \sum_{i=1}^{N} \frac{y_i - F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right)}{F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right)\left(1 - F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right)\right)} f\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}\right) \boldsymbol{x}_i = 0 \qquad (11) \end{aligned}$$

Propriedades Assintóticas

- Já sabemos que a distribuição de y é uma distribuição Bernoulli. Portanto, para consistência, precisamos adicionalmente que $F\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{\mathbf{0}}\right)$ seja a especificação correta de p.
- O verdadeiro vetor de parâmetros precisa ser o máximo da verossimilhança da população $\mathbb{E}\left[y \ln F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) + (1-y)\ln\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\right]$:

$$\mathbb{E}\left[\frac{y - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\boldsymbol{x}\right] =$$
$$\mathbb{E}\left[\frac{\mathbb{E}[y|\boldsymbol{x}] - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\boldsymbol{x}\right] = 0 \tag{12}$$

em que a primeira igualdade é obtida pela aplicação da lei das expectativas iteradas. Podemos ver facilmente que esta expressão é igual a 0 se $\mathbb{E}[y|\boldsymbol{x}] = F\left(\boldsymbol{x}\boldsymbol{\beta}_0\right)$.

- Do mesmo modo, $\widehat{\boldsymbol{\beta}} \underset{d}{\rightarrow} \mathcal{N}\left(\boldsymbol{\beta}, {}^{\Omega_0}/_{N}\right)$, em que $\Omega_0 = -\mathbb{E}\left[{}^{\partial^2 L_N}/_{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]^{-1}$. Assim,

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial L}{\partial \boldsymbol{\beta}}\frac{\partial L}{\partial \boldsymbol{\beta}'}\right]^{-1} &= \mathbb{E}\left[\left(\frac{y - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)^2 \boldsymbol{x}\boldsymbol{x}'\right]^{-1} \\
&= \mathbb{E}\left[\frac{y^2 + F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 - 2yF\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{\left(F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\right)^2} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 \boldsymbol{x}\boldsymbol{x}'\right]^{-1} \\
&= \mathbb{E}\left[\frac{\mathbb{E}[y|x] + F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 - 2\mathbb{E}[y|x]F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)}{\left(F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\right)^2} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 \boldsymbol{x}\boldsymbol{x}'\right]^{-1} \\
&= \mathbb{E}\left[\frac{F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) + F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 - 2F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2}{\left(F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\right)^2} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 \boldsymbol{x}\boldsymbol{x}'\right]^{-1} \\
&= \mathbb{E}\left[\frac{1}{F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1 - F\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)} f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2 \boldsymbol{x}\boldsymbol{x}'\right]^{-1}
\end{aligned}
$$

15/34

Efeitos Marginais

- Como os parâmetros são, em geral, não interpretáveis neste modelo, nossos objetos de interesse são os efeitos marginais: o efeito de uma mudança marginal no regressor k na probabilidade de escolher a alternativa A:

$$\frac{\partial \Pr[y=1|\boldsymbol{x}]}{\partial x_k} = f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\beta_k \tag{13}$$

- Observe que, ao contrário do modelo de probabilidade linear, onde $f\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = 1$, em qualquer modelo não linear, os efeitos marginais dependem de x.
- Podemos calcular o efeito marginal médio da amostra (ou seja, a média sobre todas as observações), o efeito marginal para o indivíduo médio (ou seja, a avaliação nos valores médios da amostra $\bar{x}$), ou o efeito marginal para qualquer outro indivíduo "representativo" com $\boldsymbol{x} = \boldsymbol{x}^*$.

- Os coeficientes estimados ainda carregam informações relevantes. Dado que $f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)$ é uma função densidade de probabilidade, ela é positiva para todos os valores de $\boldsymbol{x}$.
- Portanto, os sinais dos efeitos marginais são determinados pelo sinal dos coeficientes estimados.
- Além disso, a proporção de efeitos marginais para dois regressores diferentes é constante entre os indivíduos e igual à proporção dos dois coeficientes:

$$\frac{\partial \Pr[y=1|\boldsymbol{x}] / \partial x_k}{\partial \Pr[y=1|\boldsymbol{x}] / \partial x_\ell} = \frac{f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\beta_k}{f\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\beta_\ell} = \frac{\beta_k}{\beta_\ell} \tag{14}$$

- O efeito marginal para regressores discretos é calculado como a diferença nas probabilidades previstas.
- Em particular, no caso de um regressor dicotômico, o efeito marginal seria calculado como:

$$\begin{aligned} &\Pr\left[y = 1 | \boldsymbol{x}_{-k}, x_k = 1\right] - \Pr\left[y = 1 | \boldsymbol{x}_{-k}, x_k = 0\right] = \\ &F\left(\boldsymbol{x}'_{-k}\boldsymbol{\beta}_{-k} + \beta_k\right) - F\left(\boldsymbol{x}'_{-k}\boldsymbol{\beta}_{-k}\right) \end{aligned} \tag{15}$$

em que $\boldsymbol{x}'_{-k}$ denota um vetor com todos os regressores em x, exceto x_k.

Modelo Logit

- A FDA é

$$F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \frac{\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}}{1+\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}} \tag{16}$$

- A expressão para a condição de primeira ordem do estimador de MV é

$$\sum_{i=1}^{N}\left(y_i - \Lambda\left(\boldsymbol{x}_i'\widehat{\beta}\right)\right)\boldsymbol{x}_i = 0 \tag{17}$$

- Observe que esta expressão implica que o termo de erro é ortogonal a $\boldsymbol{x}$, como no caso do estimador de MQO.

- A expressão para a variância assintótica é

$$\Omega_0 = \mathbb{E}\left[\Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1-\Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\boldsymbol{x}\boldsymbol{x}'\right]^{-1} \tag{18}$$

- Os efeitos marginais são dados por

$$\frac{\partial \Pr[y=1|\boldsymbol{x}]}{\partial x_k} = \Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1-\Lambda\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)\beta_k \tag{19}$$

- Uma característica adicional interessante do modelo logit é que a odds ratio é linear nos regressores:

$$p = \frac{\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}}{1+\mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}}} \Longleftrightarrow \ln\frac{p}{1-p} = \boldsymbol{x}'\boldsymbol{\beta} \tag{20}$$

- Esta expressão é interessante porque nos permite interpretar $\boldsymbol{\beta}$ como semielasticidade.
- Além disso, permite estimar $\boldsymbol{\beta}$ com dados agregados sob certas hipóteses.

Modelo Probit

- A FDA é

$$F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \Phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \int_{-\infty}^{\boldsymbol{x}'\boldsymbol{\beta}} \phi(z)dz \tag{21}$$

- A expressão para a condição de primeira ordem do estimador de MV é

$$\sum_{i=1}^{N} \frac{y_i - \Phi\left(\boldsymbol{x}_i'\widehat{\beta}\right)}{\Phi\left(\boldsymbol{x}_i'\widehat{\beta}\right)\left(1 - \Phi\left(\boldsymbol{x}_i'\widehat{\beta}\right)\right)} \phi\left(\boldsymbol{x}_i'\widehat{\beta}\right) \boldsymbol{x}_i = 0 \tag{22}$$

- A expressão para a variância assintótica é

$$\Omega_0 = \mathbb{E}\left[\frac{\phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right)^2}{\Phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\left(1-\Phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right)}\boldsymbol{x}\boldsymbol{x}'\right]^{-1} \tag{23}$$

- Os efeitos marginais são dados por

$$\frac{\partial \Pr[y=1|\boldsymbol{x}]}{\partial x_k} = \phi\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\beta_k \tag{24}$$

Variável Latente

- Uma forma de dar uma interpretação mais estrutural ao modelo é concebê-lo em termos de uma medida latente de utilidade.
- Uma variável latente é uma variável que não observamos completamente.
- Por exemplo, podemos observar a decisão final de um indivíduo, mas não a utilidade intrínseca experimentada por ele.
- As decisões individuais podem ser modeladas com base em sua função de utilidade, cujos parâmetros podem ser estimados graças à preferência revelada.

- Variáveis latentes são tipicamente incluídas em um modelo econométrico/estatístico (modelo de variável latente) com diferentes objetivos:
    1. representar o efeito de covariáveis/fatores não observáveis e, em seguida, contabilizando a heterogeneidade não observada entre os sujeitos (variáveis latentes são usadas para representar o efeito desses fatores não observáveis)
    2. contabilizando erros de medição (as variáveis latentes representam os resultados "verdadeiros" e as variáveis manifestas representam suas versões "perturbadas")
    3. resumindo diferentes medidas das mesmas características não observáveis (por exemplo, qualidade de vida), de modo que as unidades amostrais possam ser facilmente ordenadas/classificadas com base nessas características (representadas pelas variáveis latentes)

Modelos de Fatores Latentes

- Modelo de análise fatorial: ferramenta fundamental em estatística multivariada para resumir várias medidas (contínuas) através de um pequeno número de traços latentes (contínuos). Nenhuma covariável está incluída.
- Modelos de Teoria da Resposta ao Item: modelos para itens (respostas categóricas) que medem um traço latente comum assumido como contínuo (ou menos frequentemente discreto) e tipicamente representando uma habilidade ou uma atitude psicológica. Normalmente nenhuma covariável é incluída.
- Modelos lineares generalizados mistos (modelos de efeitos aleatórios): extensão da classe de modelos lineares generalizados (GLM) para respostas contínuas ou categóricas que respondem por heterogeneidade não observada, além do efeito de covariáveis observáveis.

- Existem duas formas alternativas de modelar o modelo de resultado binário em termos de uma variável latente.
    1. O primeiro é chamado de modelo de função índice, no qual um limite na variável latente determina se o indivíduo escolhe uma alternativa ou outra (por exemplo, eu compro um produto se minha utilidade de comprá-lo for positiva).
    2. O segundo é chamado de modelo de utilidade aleatória (aditivo), no qual o indivíduo compara a utilidade latente associada às duas alternativas e escolhe aquela que fornece a maior utilidade.
- No caso binário, essas interpretações são um tanto equivalentes, mas a distinção entre elas torna-se relevante no caso multinomial.

- Seja y^* uma variável latente de interesse, tal que

$$y^* = \boldsymbol{x}'\boldsymbol{\beta}^* + u^* \tag{25}$$

em que $\boldsymbol{x}$ é um vetor de regressores e u é um componente de erro não observado com FDA $F(\cdot)$.

- Observa-se somente qual alternativa é escolhida pelo indivíduo, isto é,

$$y = \begin{cases} 1, & \text{se } y^* > 0 \\ 0, & \text{se } y^* \leq 0 \end{cases} \tag{26}$$

em que 0 é uma normalização de um threshold c, não identificado se o modelo inclui um termo de intercepto.

- A probabilidade de observar $y = 1$ é obtida como segue:

$$
\begin{aligned}
\Pr[y = 1|\boldsymbol{x}] &= \Pr[y^* > 0|x] \\
&= \Pr[\boldsymbol{x}'\boldsymbol{\beta}^* + u^* > 0] \\
&= \Pr[u^* > -\boldsymbol{x}'\boldsymbol{\beta}^*] \\
&= \Pr[u^* < \boldsymbol{x}'\boldsymbol{\beta}^*] \\
&= \Pr\left[\frac{u^*}{\sigma_u} < \frac{\boldsymbol{x}'\boldsymbol{\beta}^*}{\sigma_u}\right] \\
&= \Pr[u < \boldsymbol{x}'\boldsymbol{\beta}] \\
&= F\left(\boldsymbol{x}'\boldsymbol{\beta}\right) \qquad (27)
\end{aligned}
$$

em que a última igualdade decorre do fato de que a função densidade de probabilidade de u é simétrica em torno de zero (por exemplo, a logística e as distribuições normal padrão).

Figura 1: Distribuição Logística

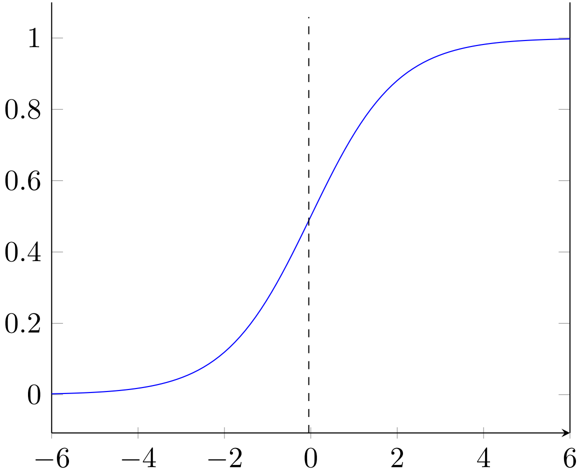
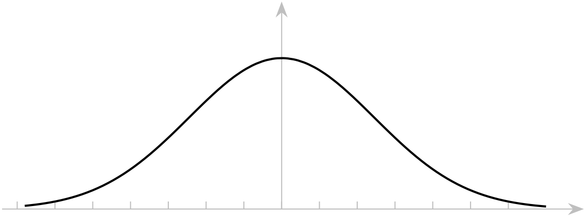


Figura 2: Distribuição Normal



- Na expressão anterior, verifica-se que um limiar c não seria identificado se o modelo incluísse um intercepto porque poderíamos aumentar o intercepto e reduzir o limiar na mesma quantidade a e a probabilidade não mudaria.
- Da mesma forma, os parâmetros são identificados apenas até uma escala:

$$\Pr\left[u > -\boldsymbol{x}'\boldsymbol{\beta} >\right] = \Pr\left[ua > -\boldsymbol{x}'\boldsymbol{\beta}a >\right] \tag{28}$$

- Portanto, temos que impor uma restrição na variância do termo de erro para que os parâmetros possam ser identificados de forma única.
- Essa restrição está implícita na distribuição logística (impomos que a variância seja ${k^2\pi^2}/{3}$ fazendo $k = 1$), e no caso de erros normais, normalmente impomos que a variância seja 1 para que trabalhemos com a distribuição normal padrão e o problema se reduza a o modelo probit.
- Essas restrições devem ser levadas em consideração ao interpretar os coeficientes.
- Na verdade, esta é a razão pela qual os próprios coeficientes são difíceis de serem interpretados diretamente e nos concentramos nos efeitos marginais de cada regressor (ou na razão entre dois coeficientes, que fornece a razão dos efeitos marginais).

- Para a maioria dos propósitos, essas normalizações são inócuas.
- Infelizmente, no contexto da análise de decomposição para modelos probit e logit, essas normalizações não são tão inócuas.
- Ao decompor as diferenças médias nas probabilidades de resultado para duas populações, o objetivo natural seria estimar o quanto as diferenças de grupo nos parâmetros β^* do modelo de variável latente contribuem para as diferenças médias nas probabilidades de resultado.
- Decomposições baseadas apenas nos parâmetros β estimados mensuram os efeitos das diferenças de grupo em $\dfrac{\beta^*}{\sigma_u}$ ao invés de β^*.
- Consequentemente, diferenças de grupo nos parâmetros β^* são confundidas com diferenças de grupo nos parâmetros σ_u.

Econometria I

Modelos de Resposta Binária

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024