二項選択モデルII

ロジットモデル・プロビットモデル

モデルの紹介
係数の解釈

プロビットモデル・ロジットモデルへ

- 線形確率モデルでは、説明変数が取る値によって、応答確率がゼロより小さいとか1より大きい、なんてことが起きました。

- この問題の解決策として、応答確率を次のようにモデル化します（式1）。

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = G(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})$$

ここで$G(\cdot)$は何らかの関数です。

- ちなみに線形確率モデルも、この形で表すことができます。

$$G(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$$

- 関数$G(\cdot)$を「インプットをアウトプットとしてそのまま返す」ものに設定すると、（式1）は線形確率モデルになるということです。

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = G(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})$$

- ここで関数$G(\cdot)$に次のようなものを選びましょう：

「インプットが何であれ、アウトプットは必ず0より大きく1より小さくなる」

- つまり

$$0 < G(z) < 1 \; for \; any \; z$$

- そのように関数$G(\cdot)$を設定すれば、
  - $(\beta_0, \beta_1, \dots, \beta_k)$がどんな値を取ったとしても、
  - $(X_{i1}, \dots, X_{ik})$がどんな値を取ったとしても、

応答確率は、0より小さくならないし、1より大きくなることもない。

- こうすれば、線形確率モデルに起こっていた問題を回避できます。

- そういう特性を持つ関数はいくつもあるかもしれませんが、次の関数がよく使われます。

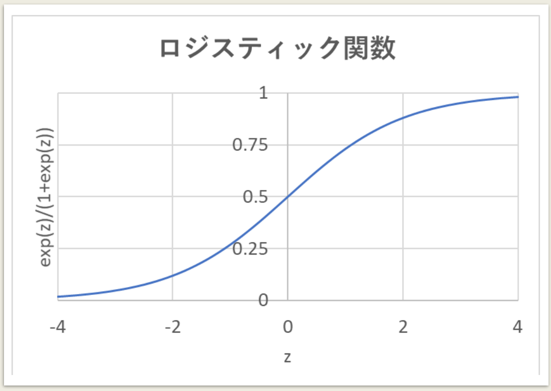
（1）ロジスティック関数（以下では$\Lambda(z)$と表すことがあります）

$$\boxed{G(z) = \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}}$$

- この関数はインプット(z)がどんな値であれ、アウトプット$G(z)$は0より小さくならないし、1より大きくもなりません。

$$z \to -\infty \Rightarrow \exp(z) \to 0 \Rightarrow \Lambda(z) \to 0$$
$$z \to +\infty \Rightarrow \exp(z) \to +\infty \Rightarrow \Lambda(z) \to 1$$

- この関数がどのようなものかプロットしてみます。



（二項）ロジットモデル

- 関数Gにロジスティック関数Λを使ったもの、すなわち応答関数が

$$\boxed{Pr(Y_i=1|X_{i1},\dots,X_{ik})=\frac{\exp(\beta_0+\beta_1X_{i1}+\cdots+\beta_kX_{ik})}{1+\exp(\beta_0+\beta_1X_{i1}+\cdots+\beta_kX_{ik})}}$$

で表されるモデルを

（二項）ロジットモデル

と言います。

- ロジスティック関数以外にも関数$G(\cdot)$としてよく使われるものがあります。

(2) 標準正規変数の累積分布関数（以下では$\Phi(z)$と表すことがあります）

  - 標準正規確率変数Zを考えます。
  - Zがある値zより小さい値をとる確率は$Pr(Z\le z)$です。
  - この確率を$\Phi(z)$と表します$\Phi(z)\equiv Pr(Z\le z)$です。
  - 確率・統計の教科書の後ろに載っているテーブルが$\Phi(z)$です。
  - 式を使って表すと

$$\boxed{\Phi(z)=\int_z^{\infty}\phi(z)}$$

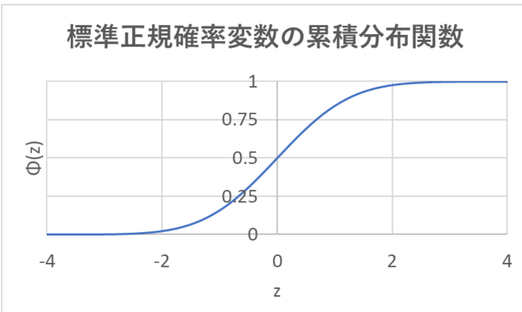
ここで$\phi(z)$は標準正規確率変数の確率密度関数です。

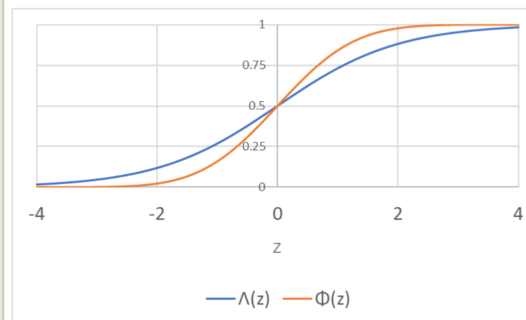
- 関数$\Phi(z)$は以下の特性を持っています。

$$\mathrm{z}\to-\infty\Rightarrow\Phi(z)\to0$$
$$\mathrm{z}\to+\infty\Rightarrow\Phi(z)\to1$$

- $\Phi(z)\equiv Pr(Z\le z)$から上の特性は明らかですね。
  - $Pr(Z\le-\infty)$はゼロ、って直観的に分かりますよね（表記の厳密さは無視）。
  - $Pr(Z\le+\infty)$は1、って直観的に分かりますよね（表記の厳密さは無視）。

- この関数がどのようなものかプロットしてみます。Φ





（二項）プロビットモデル

- 関数Gに標準正規確率変数の累積分布関数Φを使ったもの、すなわち応答関数が

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \Phi(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

で表されるモデルを

（二項）プロビットモデル

と言います。

二項選択モデル：これまでのまとめ

- 線形確率モデル

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

- 二項ロジットモデル

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})}$$

- 二項プロビットモデル

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \Phi(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

ただしΦは標準正規確率変数の累積分布関数。

潜在的変数を使ったプロビットモデルの表し方

- ロジットモデルとプロビットモデルは、線形確率モデルの弱点を克服するために、関数Gに適当な関数を選ぶことで導出されました。
- 「テキトー」に関数を選んで、無理やり、応答確率が0から1の範囲以外の値をとらないようにしただけなんじゃないの、、、と思われたかもしれません。
- ここでは、個人の選択行動に関して無理のない仮定を置き、応答確率を導出します。
- そしてその応答確率がプロビット・モデルのものになることを示します。
- プロビットは、「テキトー」なものではないということです。

13

潜在的変数を使ったプロビットモデルの表し方

- 個人iがする・しないを選択するとします。
- ここで連続的な潜在的変数（latent variable）Y_i^*を考えます。
- この変数は、個人iが「どの程度したいか」を表すもの。
  - Y_i^*がとても大きい→とてもしたい
  - Y_i^*がとても小さい→とてもしたくない
- ただし分析者はこの変数自体は観測しません（だから「潜在的」です）。
  - 個人iが「どの程度したいか」は分かりません。
  - 観測できるのは、個人iがするか・しないかですね。

14

- この潜在的変数Y_i^*は次のようにモデル化できるとします：

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i, \quad U_i \sim N(0,1)$$

- 説明変数$X_{ij}(j = 1, \ldots, k)$と誤差項U_iに相関は無いとします。
- この式は、回帰分析初級で学習した線形モデルと同じです、2点を除いて。

（1） 初級で学習した線形モデル: 従属変数が観測可能
この式: Y_i^*は観測されないもの

（2） 初級で学習した線形モデル: 誤差項に特定の分布の仮定はしない
この式: 誤差項に正規分布（標準正規分布）の仮定をする

- 誤差項の分散が1という仮定については後で説明します。
- ここで、分析者は個人iの潜在的変数Y_i^*は観測できませんが、個人iがするかしないか（Y_i）は観測できるとします。
- この二つの変数に以下の関係を仮定します。

15

$$Y_i = 1 \; if \; and \; only \; if \; Y_i^* \geq 0$$
$$Y_i = 0 \; if \; and \; only \; if \; Y_i^* < 0$$

- 言葉を使えば、この仮定は

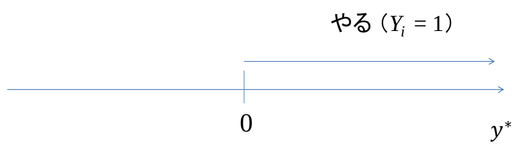
したい程度が0以上なら、「やる」を選択

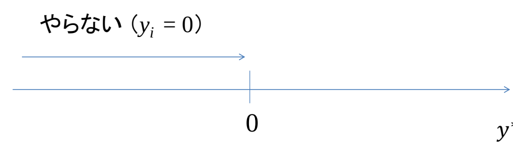
したい程度が0未満なら、「やらない」を選択

ということ（正確には逆も、です。）

- これが個人の選択行動に置く仮定です
  - それなりに合理的な個人であればこんな感じで選択してますかね。
  - したい程度が小さいなら「やる」、大きいなら「やらない」人は変わってますよね。
- なぜ0が閾値になるかについては後で説明します。
  - 閾値は「いきち」または「しきいち」と読みます。
  - 英語ではthresholdとかcutoff pointといいます。

16





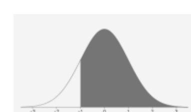
これまでのまとめ

- Y_i^* : 潜在的変数（どの程度したいかを表す）　観測しない
- Y_i : 実際の選択 （$Y_i = 1$ する; $Y_i = 0$ しない） 観測する

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i, \quad U_i \sim N(0,1)$$
$$Y_i = 1 \; if \; and \; only \; if \; Y_i^* \geq 0$$
$$Y_i = 0 \; if \; and \; only \; if \; Y_i^* < 0$$

- 説明変数$X_{ij}(j = 1, \dots, k)$と誤差項U_iに相関は無し。
- 応答確率を求めてみましょう。
- $Pr(Y_i = 1 | X_{i1}, \dots, X_{ik})$ 、、、、

$Pr(Y_i = 1 | X_{i1}, \dots, X_{ik})$
$= Pr(Y_i^* \geq 0 | X_{i1}, \dots, X_{ik})$
$= Pr(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i \geq 0 | X_{i1}, \dots, X_{ik})$
$= Pr(U_i \geq -(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}) | X_{i1}, \dots, X_{ik})$



$U_i \sim N(0,1)$なので標準正規分布の対称性より

$= Pr(U_i < \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} | X_{i1}, \dots, X_{ik})$
$= \Phi(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$

ということでプロビットモデルの応答確率になりました。

- よって、プロビットモデルは（合理的な）個人の選択行動と整合的であると言えます。
- U_iにロジスティック分布を仮定することで、ロジットモデルも同様に導出することができます。
- よって、ロジットモデルも（合理的な）個人の選択行動と整合的であると言えます。

潜在的変数を使ったプロビットモデルの表し方 なぜ誤差項の分散を1と設定するの？

- 潜在的変数を使ったプロビットモデルの表し方では、誤差項の分散を1と設定しました。
- なんかテキトーな感じがする、、、、
- ここではその疑問に答えます。結論からいうと、これはテキトーではありません。
- これを見るために、誤差項の分散をσ_*^2とします。

$$Y_i^* = \beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik} + V_i, \quad V_i \sim N(0, \sigma_*^2)$$

$$y_i = 1 \; if \; and \; only \; if \; y_i^* \geq 0$$
$$y_i = 0 \; if \; and \; only \; if \; y_i^* < 0$$

- 応答確率を求めてみましょう
  - 以後$X_{i1}, \dots, X_{ik}$の条件は適宜省略します。$X_{i1}, \dots, X_{ik}$を条件とした確率を考えていますから、。$X_{i1}, \dots, X_{ik}$は定数扱いです。

$Pr(Y_i = 1)$

$= Pr(Y_i^* \geq 0)$

$= Pr(\beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik} + V_i \geq 0)$

$= Pr(V_i \geq -(\beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik})) = Pr\left(\frac{V_i}{\sigma_*} \geq -\frac{\beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik}}{\sigma_*}\right)$

$V_i / \sigma_* \sim N(0,1)$なので標準正規分布の対称性より

$= Pr\left(\frac{V_i}{\sigma_*} < \frac{\beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik}}{\sigma_*}\right)$

$= \Phi\left(\frac{\beta_0^*}{\sigma_*} + \frac{\beta_1^*}{\sigma_*} X_{i1} + \cdots \frac{\beta_k^*}{\sigma_*} X_{ik}\right)$

- ここで話を簡単にするために、

$$Pr(Y_i = 1) = \Phi\left(\frac{\beta_0^*}{\sigma_*} + \frac{\beta_1^*}{\sigma_*} X_{i1}\right)$$

を考えましょう。

- このモデルにおいて、以下のパラメータの組み合わせはすべて同じ応答確率を与えます。

$(\beta_0^*, \beta_1^*, \sigma_*) = (3,2,1)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (6,4,2)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (9,6,3)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (12,8,4)$

- 実際、この応答確率と同じになるパラメータの組み合わせは無限にあります。
- 別の例も考えられます。例えば、

$(\beta_0^*, \beta_1^*, \sigma_*) = (3,8,1)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (6,16,2)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (9,24,3)$
$(\beta_0^*, \beta_1^*, \sigma_*) = (12,32,4)$

- これらのパラメータの組み合わせはすべて同じ応答確率を与えます（先ほどとは別ですが）。

- このことは、

「$(\beta_0^*, \beta_1^*, \sigma_*)$を個別に識別することはできない」

ことを示しています。

- ある確率に対して、それと整合的なパラメータの組み合わせ$(\beta_0^*, \beta_1^*, \sigma_*)$が無限にある、、、、
- そうであるなら、どれか一つには決められません（識別できない）。

- ここでもう一度モデルを見てみましょう:

$$Pr(Y_i = 1) = \Phi\left(\frac{\beta_0^*}{\sigma_*} + \frac{\beta_1^*}{\sigma_*}X_{i1}\right)$$

- 応答確率は、「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」によって決まっていることが分かります。

- よって、$(\beta_0^*, \beta_1^*, \sigma_*)$を個別に識別することはできませんが、

「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」

は識別できるということになります。

- $(\beta_0^*, \beta_1^*, \sigma_*)$を個別に識別することはできない、ただし「$\beta_0^*$と$\sigma_*$の比」そして「$\beta_1^*$と$\sigma_*$の比」は分かるので、一般的に$\sigma_* = 1$とセットします。

25

従って、

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + U_i, \quad U_i \sim N(0,1)$$
$$y_i = 1\ if\ and\ only\ if\ y_i^* \geq 0$$
$$y_i = 0\ if\ and\ only\ if\ y_i^* < 0$$

このモデルにおける(β_0, β_1)は、

$$Y_i^* = \beta_0^* + \beta_1^* X_{i1} + V_i, \quad V_i \sim N(0, \sigma_*^2)$$
$$Y_i = 1\ if\ and\ only\ if\ Y_i^* \geq 0$$
$$Y_i = 0\ if\ and\ only\ if\ Y_i^* < 0$$

このモデルの「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」ということになります。

26

$\sigma_*^2 = 1$でなければだめ？

- 計量のソフトウェアでは、一般的に$\sigma_* = 1$とセットされています。
  - よって σ_*は推定するパラメーターではありません。

- ただし自分でプログラムを書くなら、$\sigma_* = 1$以外の値を自分で設定することもできます。

- 例えば、$\sigma_* = 2$と設定した場合、定数項やx_{i1}の係数は、$\sigma_* = 1$と設定した場合に得られるものの2倍となります。

- ただし「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」はいずれの場合も同じです。

27

$(\beta_0^*, \beta_1^*, \sigma_*)$を個別に識別することはできない、問題ない？

- 応答確率が、「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」によって決まっている以上、$(\beta_0^*, \beta_1^*, \sigma_*)$を個別に識別できなくても分析上問題ありません。

- 「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」が分析に必要な情報のすべてです。

- そして、「β_0^*とσ_*の比」そして「β_1^*とσ_*の比」は分析から分かります。

- それは

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + U_i, \quad U_i \sim N(0,1)$$

$$Y_i = 1\ if\ and\ only\ if\ Y_i^* \geq 0$$
$$Y_i = 0\ if\ and\ only\ if\ Y_i^* < 0$$

このモデルのパラメータ(β_0, β_1)です。

- すなわちX_{i1}を説明変数とするプロビットモデルです。

28

なぜσ_*^2の値は分からない？直感的説明

$$Y_i^* = \beta_0^* + \beta_1^* X_{i1} + \cdots + \beta_k^* X_{ik} + V_i, \quad V_i \sim N(0, \sigma_*^2)$$

- ここからスタートしてみましょう。
- この式において、Y_i^*が観測可能なら、これは線形回帰モデルです。
- Y_i^*が観測可能なら、Y_i^*の「スケール」の情報があります。
- そこからσ_*^2が分かります。
- しかし、プロビットモデルでは Y_i^*は潜在的変数、つまり観測不可能です。
- 分かるのは「する」「しない」。よって Y_i^*のスケールの情報はありません。
- そのためσ_*^2は分からないということになります、分からなくても分析上困りませんが。

29

潜在的変数を使ったプロビットモデルの表し方 なぜ閾値を0と設定するの？

$$Y_i = 1 \; if \; and \; only \; if \; Y_i^* \geq 0$$
$$Y_i = 0 \; if \; and \; only \; if \; Y_i^* < 0$$

- すなわち、やりたい程度が0以上なら、「やる」を選択、やりたい程度が0未満なら、「やらない」を選択、と仮定しました。
- ここでは、なぜ閾値を0と設定してよいのか、について説明します。
- これを見るために、閾値をμとします。

$$Y_i^* = \beta_0^* + \beta_1 X_{i1} + U_i, \quad U_i \sim N(0,1)$$
$$Y_i = 1 \; if \; and \; only \; if \; Y_i^* \geq \mu$$
$$Y_i = 0 \; if \; and \; only \; if \; Y_i^* < \mu$$

30

- 応答確率を求めてみましょう。

$Pr(Y_i = 1)$

$= Pr(Y_i^* \geq 0)$

$= Pr(\beta_0^* + \beta_1 X_{i1} + U_i \geq \mu)$

$= Pr(U_i \geq -((\beta_0^* - \mu) + \beta_1 X_{i1}))$

$U_i \sim N(0,1)$なので標準正規分布の対称性より

$= Pr((U_i < (\beta_0^* - \mu) + \beta_1 X_{i1})$

$= \Phi((\beta_0^* - \mu) + \beta_1 X_{i1})$

31

- 所与の値β_1に対して、例えば、

$$(\beta_0^*, \mu) = (6,1)$$
$$(\beta_0^*, \mu) = (7,2)$$
$$(\beta_0^*, \mu) = (8,3)$$
$$(\beta_0^*, \mu) = (9,4)$$

……………………

はすべて同じ応答確率を与えます。

これは、「(β_0^*, μ)を個別に識別することはできない」

ことを示しています。

- 同じ応答確率を与えるパラメーターの組み合わせがいっぱいある、、、
- そのときに、一つのパラメーターの組み合わせを選ぶことはできませんから。

32

- ここで応答確率

$$Pr(Y_i = 1) = \Phi\big((\beta_0^* - \mu) + \beta_1 X_{i1}\big)$$

をよく見てみると、応答確率は$(\beta_0^* - \mu)$に依存していることが分かります。

- 従って、(β_0^*, μ)を個別に識別することはできないけれど、$(\beta_0^* - \mu)$を識別することはできるということになります。
- (β_0^*, μ)は個別には分からないけれど、
- $(\beta_0^* - \mu)$は分かる、、、、
- そのため、普通は$\mu = 0$と設定します。

33

従って、

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + U_i, \quad U_i \sim N(0,1)$$
$$Y_i = 1\ if\ and\ only\ if\ Y_i^* \geq 0$$
$$Y_i = 0\ if\ and\ only\ if\ Y_i^* < 0$$

このモデルにおけるβ_0は、

$$Y_i^* = \beta_0^* + \beta_1 X_{i1} + U_i, \quad u_i \sim N(0,1)$$
$$Y_i = 1\ if\ and\ only\ if\ Y_i^* \geq \mu$$
$$Y_i = 0\ if\ and\ only\ if\ Y_i^* < \mu$$

このモデルのβ_0^*とμの差、ということになります。

34

閾値は0でなければだめ？

- 計量のソフトウェアでは、一般的に$\mu = 0$とセットされています。
  - よって μは推定するパラメーターではありません。
- ただし自分でプログラムを書くなら、$(\beta_0^* - \mu)$は分かるので、$\mu = 0$ではなく$\beta_0^* = 0$と設定することもできます。
- その場合は、モデルには定数項がなく、閾値が推定するパラメーターの一つになります。
- 見た目は違いますが、応答確率に関するすべての情報は同一です。

35

なぜμの値は分からない？直感的説明

- Y_i^*が観測可能なら、Y_i^*の「水準」の情報があります。
- でもY_i^*は潜在的変数、つまり観測不可能です。
- 分かるのは「する」「しない」。よって Y_i^*の「水準」の情報はありません。
- 従って、Y_i^*の「水準」の情報が無い以上、定数項とμの水準を決めることができません、分からなくても分析上困りませんが。

36

潜在的変数を使ったプロビットモデルの表し方：まとめ

- Y_i^*：潜在的変数（どの程度したいかを表す） 観測しない
- Y_i：実際の選択 （$Y_i = 1$ する; $Y_i = 0$ しない） 観測する

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i, \quad U_i \sim N(0,1)$$

$$Y_i = 1 \; if \; and \; only \; if \; Y_i^* \geq 0$$
$$Y_i = 0 \; if \; and \; only \; if \; Y_i^* < 0$$

- このように書くと、「これはプロビットモデルです」と言っているのと同じこと。
- なぜなら応答確率が

$$Pr(Y_i = 1|X_{i1}, \ldots, X_{ik}) = \Phi(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

になるから。

37

推定結果の解釈

- 潜在的変数y_i^*（どの程度したいか）を使ったプロビットモデルの表し方を考えましょう。

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

- この式の係数（$\beta_1, \ldots, \beta_k$）は何を意味するでしょうか？
- 線形回帰モデルと同様に解釈すると、、、

$$\frac{\partial E(Y_i^*|X_{i1}, \cdots X_{ik})}{\partial X_{ij}} = \beta_j$$

ですから、「X_{ij}（$j = 1, \ldots, k$）が1単位増えると、したい程度がβ_j（$j = 1, \ldots, k$）だけ変化する」になりますよね。

38

- 「X_{ij}が1単位増えると、したい程度がβ_jだけ変化する」、、、、
- 「したい程度」というのは、仮想的に構築された概念です。
- そのため、「したい程度」には、明確に定義された単位はありません。
- したい程度が20減ったとか、50増えたとか、、、、明確に定義された単位が無い以上、あまり有用な情報ではありません。
- 従って、

プロビットモデルの係数のサイズを解釈するのはあまり意味がない

ということになります。

- ロジットモデルの係数にも同じことが言えます。

39

推定結果の解釈：限界効果

- そのためロジットやプロビットの結果を解釈する際には、**限界効果**（**marginal effects**）と呼ばれるものが使われます。
- 限界効果とは、より正確に言うと、

説明変数の応答確率への限界効果

です。

- 式で表すと、説明変数X_j（$j = 1, \ldots, k$）の限界効果は

$$\frac{\partial Pr(Y = 1|X_1, \cdots X_k)}{\partial X_j}$$

で、

説明変数X_jが一単位増加したときに、応答確率がどれだけ変化するか

を意味します。

40

推定結果の解釈：限界効果

- ここではiの添字を落とします。
- 線形確率モデルを思い出しましょう。

$$Pr(Y=1|X_1,\dots,X_k)=\beta_0+\beta_1X_1+\cdots+\beta_kX_k$$

- このモデルでは、X_j ($j=1,\dots,k$)の限界効果は

$$\frac{\partial Pr(Y=1|X_1,\cdots X_k)}{\partial X_j}=\beta_j$$

です。

- すなわち、説明変数の係数は、その説明変数の限界効果になります。
- 従って、線形確率モデルの結果の解釈は容易です。
- ロジットやプロビットは、残念ながら、これほど単純ではありません。

41

限界効果：ロジットモデル

- まずロジットモデルから考えます。

$$Pr(Y=1|X_1,\dots,X_k)=\frac{\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}{1+\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}$$

- 求めたいのは、

$$\frac{\partial Pr(Y=1|X_1,\cdots X_k)}{\partial X_j}$$

- ということは、

$$\frac{\partial Pr\left[\frac{\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}{1+\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}\right]}{\partial X_j}$$

を計算しなければならないということです。

42

- これは合成関数の微分のルールを使って計算でき、以下のようになります。

ロジットモデルにおける説明変数X_j ($j=1,\dots,k$)の限界効果は

$$\boxed{\frac{\partial Pr(Y_i=1|X_1,\cdots X_k)}{\partial X_j}=\frac{\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}{[1+\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)\ ]^2}\cdot\beta_j}$$

- この式の導出は試験の範囲外、この式自体も暗記する必要はないです。
- ただし、次のことは知っている必要があります。

説明変数X_jの限界効果は、その係数、その説明変数の値、他の説明変数の係数、そして他の説明変数の値に依存する。

43

- 従って、一般的に言って、

説明変数X_j ($j=1,\dots,k$)の限界効果は、個人個人で異なる

ことになります、個人個人で説明変数の値は異なりますから。

- この点において、ロジットモデルは線形確率モデルと大きく異なります。

44

限界効果：プロビットモデル

- 次にプロビットモデルを考えます。

$$Pr(Y=1|X_1,\dots,X_k)=\Phi(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)$$

- 求めたいのは、

$$\frac{\partial Pr(Y=1|X_1,\cdots X_k)}{\partial X_j}$$

- ということは、

$$\frac{\partial Pr[\Phi(\beta_0+\beta_1X_{i1}+\cdots+\beta_kX_{ik})]}{\partial X_j}$$

を計算しなければならないということです。

45

- これは、「累積分布関数の微分は確率密度関数」という事実と、合成関数の微分のルールを使って計算でき、以下のようになります。

- プロビットモデルにおける説明変数$X_j\,(j=1,\dots,k)$の限界効果は

$$\frac{\partial Pr(Y_i=1|X_1,\cdots X_{ik})}{\partial X_j}=\emptyset(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)\cdot\beta_j$$

ただし$\emptyset(\cdot)$は正規確率変数の確率密度関数。

- こちらの式は簡単なので覚えてください。

- また次のことは知っている必要があります。

説明変数X_jの限界効果は、その係数、その説明変数の値、他の説明変数の係数、そして他の説明変数の値に依存する。

46

- 従って、一般的に言って、

説明変数$X_j\,(j=1,\dots,k)$の限界効果は、個人個人で異なる

ことになります、個人個人で説明変数の値は異なりますから。

- この点において、プロビットモデルは線形確率モデルと大きく異なります。

47

これまでのまとめ：
ロジットモデル・プロビットモデルにおける限界効果

- 説明変数$X_j\,(j=1,\dots,k)$の応答確率への限界効果は

$$\frac{\partial Pr(Y=1|X_1,\cdots X_k)}{\partial X_j}$$

- ロジットモデルでは

$$\frac{\exp(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}{[1+\exp(\beta_0+\beta_1X_{i1}+\cdots+\beta_kX_k)]^2}\cdot\beta_j$$

- プロビットモデルでは

$\emptyset(\beta_0+\beta_1X_{i1}+\cdots+\beta_kX_{ik})\cdot\beta_j$ （ただし$\emptyset(\cdot)$は正規確率変数の確率密度関数）

- ロジット、プロビットともに、説明変数x_{ij}の限界効果は、その係数、その説明変数の値、他の説明変数の係数、そして他の説明変数の値に依存する。

48

- 説明変数$X_j\ (j=1,\dots,k)$の限界効果が個人個人で異なるのであれば、何を論文でリポートすればよいのでしょうか？

- 一般的に論文でリポートされるのは、

（1）説明変数の標本平均における限界効果
（2）平均限界効果

のいずれかです。

- これらについては次の講義ノート（演習）で詳しく解説します。

49

限界効果：ダミー変数の場合

- これまでは、説明変数が連続型という前提で、限界効果を説明してきました。

- 説明変数が連続型という前提 → 微分を使っていた

- 説明変数がダミー変数の場合、0と1しか値を取らないので、微分を取るのはちょっと、、、ですね。

- 説明変数X_jがダミー変数の場合、限界効果は以下のものになります。

$$\boxed{Pr\left(Y=1\middle|X_1,\dots,X_j=1,\dots,X_k\right)-Pr\left(Y_i=1\middle|X_1,\dots,X_j=0,\dots,X_k\right)}$$

「$X_j=1$のときの応答確率と$X_j=0$のときの応答確率の差」

50

ロジット・プロビット 説明変数の係数から直接言えること

- 「ロジット、プロビットの係数のサイズを解釈することにあまり意味は無い」ことを説明しました。

- そして結果の解釈では主に限界効果を使うことも説明しました。

- それでは、ロジット、プロビットの係数から直接言えることは無いのでしょうか？

- あります、限界効果の符号です。

説明変数$X_j\ (j=1,\dots,k)$の限界効果の符号は、説明変数X_jの係数β_jの符号と同じになる。

- なぜでしょうか？

51

- 説明変数$X_j\ (j=1,\dots,k)$の限界効果は、ロジットモデルでは

$$\boxed{\frac{\exp(\beta_0+\beta_1X_1+\dots+\beta_kX_k)}{[1+\exp(\beta_0+\beta_1X_1+\dots+\beta_kX_k)\,]^2}}\cdot\beta_j$$

- ここは常に正。従って、限界効果の符号とβ_jの符号は一致する。

- プロビットモデルでは

$$\boxed{\emptyset(\beta_0+\beta_1X_{i1}+\dots+\beta_kX_{ik})}\cdot\beta_j$$

- ここは常に正。従って、限界効果の符号とβ_jの符号は一致する。

- 従って、ロジット、プロビットともに、係数の符号を見ればすぐに限界効果の符号が分かります。

- しかし限界効果の大きさは、限界効果を計算しなければ分かりません。

52

まとめ：ロジットモデル・プロビットモデル

- ロジットモデル

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})}$$

- プロビットモデル

$$Pr(Y_i = 1|X_{i1}, \dots, X_{ik}) = \Phi(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

ただしΦは標準正規確率変数の累積分布関数。

- 潜在的変数を使ったプロビットモデルの表し方：

Y_i^*を潜在的変数（どの程度したいかを表す）とすると、

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i, \quad U_i \sim N(0,1)$$

$$Y_i = 1 \ if \ and \ only \ if \ Y_i^* \geq 0$$
$$Y_i = 0 \ if \ and \ only \ if \ Y_i^* < 0$$

53

- ロジット、プロビットにおいて、係数のサイズを解釈することにあまり意味は無い。
- これは、係数が潜在的変数に対応していて、「説明変数が1単位増えたときに、<u>したい程度</u>がどれだけ変化する」かを捉えているため。
- よって、ロジット、プロビットの結果の解釈には、説明変数の応答確率への「限界効果」を使う。
- これは、「説明変数が一単位増加したときに、応答確率がどれだけ変化するか」
- 説明変数X_j $(j = 1, \dots, k)$の限界効果は

ロジット：

$$\frac{\partial Pr(Y = 1|X_1, \cdots X_k)}{\partial X_j} = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}{[1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)]^2} \cdot \beta_j$$

プロビット：

$$\frac{\partial Pr(Y = 1|X_1, \cdots, X_k)}{\partial X_j} = \emptyset(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k) \cdot \beta_j$$

54

- 説明変数X_j $(j = 1, \dots, k)$がダミー変数の場合、限界効果は「$X_j = 1$のときの応答確率と$X_j = 0$のときの応答確率の差」

$$Pr(Y = 1|X_1, \dots, X_j = 1, \dots, X_{ik}) - Pr(Y = 1|X_1, \dots, X_j = 0, \dots, X_k)$$

- ロジット、プロビットにおいて、、、、

説明変数X_j $(j = 1, \dots, k)$の限界効果は、その係数、その説明変数の値、他の説明変数の係数、他の説明変数の値、これらすべてに依存する。

- 従って、個人個人で、説明変数X_j $(j = 1, \dots, k)$の限界効果の大きさは異なる。
- ロジット、プロビットともに、係数の符号を見ればすぐに限界効果の符号が分かる。
- しかし分析の目的は、限界効果の符号に加えて、限界効果の大きさを知ることである。
- 限界効果の大きさは、限界効果を計算しなければ分からない。
- よって、ロジット、プロビットを推定した場合、必ず限界効果を計算する必要がある。

55