

福山平成大学経営学部紀要
第 16 号 (2020), 85-104 頁

社会システム分析のための統合化プログラム 3 5 — 回帰分析の拡張・トービット回帰分析 —

福井 正康^{*1}・兔内 祥子^{*}

^{*1} 福山平成大学経営学部経営学科

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回はこのプログラムに計量経済分野で使われる統計分析手法を追加する。この論文では、中心極限定理を基にして拡張された回帰分析、床効果や天井効果を持つデータに対するトービット回帰分析について、その理論とプログラムの利用法を解説する。

キーワード：計量経済、トービット回帰分析、College Analysis、統計

1. はじめに

これまで分析プログラム College Analysis の統計分析は、データ数があまり多くない場合でも使えることを考えて作られてきた。そのため、比較の検定などにも積極的にノンパラメトリック検定を使ってきた。しかし、このような進め方の裏には問題もある。たとえば回帰分析の誤差項には、あまり深く考えることなく等分散の仮定をおいている。計量経済の分野ではこの等分散性を問題をにする。しかし、計量経済分野の統計分析がより厳密なのかというと、そういうわけでもない。計量経済の分野ではデータが多く集められるため、データの平均値やパラメータの推定値などで中心極限定理が成り立ち、その分布は基本的に正規分布になる。つまり、データ数があまり多く取れない、すなわち中心極限定理が保証されない分野では仮定を余分に付けて計算を可能にし、データ数が多く中心極限定理が成り立つ分野では、仮定を減らして一般的な議論を行う。これは分析実行者の置かれた状況で使い分けられる。

College Analysis は元々保健分野の統計学から始まっている。そのため上で述べた後者の視点が抜け落ちていた。今回著者らは後者の立場に立ってこれまでの分析を見直すことにした。これまでの分析の中にはそのまま利用できるものもあるが、機能追加を必要とするものもある。この論文では、2群間の比較検定について注釈を加え、(重) 回帰分析については計量経済でよく使われる機能を追加することにする。

また、計量経済では多くの場面で回帰分析を使うことから、回帰分析に関する手法が特に発達した。著者らはこれらの手法についても順を追って追加して行く。ここではまず、トービット回帰分析を取り上げる。例えば、収入と車への支出の関係について考えると、車に対する支出は、車を持っていなければ収入が多くても 0 である、また 5 点評価の成

績ではいくら勉強しても最高は 5 点である。このような下限や上限を持つことを床効果や天井効果と呼ぶ。車に対する支出や勉強に対する評価など、通常の回帰分析ではこの効果に引きずられ、分析者の求める傾きが得られない。このようなときに利用されるのがトービット回帰分析である。これについては章を設けて紹介する。最後に、この論文で利用される、回帰分析に関する行列と確率分布の公式について補遺にまとめておく。本文の中からは【公式1】などのように引用する。

2. 回帰分析の拡張

計量経済分析には回帰分析が頻繁に利用される。その目的は、因果関係の導出やデータの予測などであるが、データ数が多い対象を扱うので、パラメータの推定に対して、中心極限定理が成り立つと仮定することが多く、基本的に正規分布を元にした検定が行われる。これらの検定への理解を深めるために、多少教科書的になるが、今回の拡張を含めた重回帰分析の理論について復習しておく^[1]。

2.1 回帰分析の理論

均一分散誤差の場合の理論

目的変数を k 個の説明変数と定数項で回帰する重回帰式を以下のように仮定する。

$$\mathbf{y} = \mathbf{Z}\mathbf{d} + \mathbf{u}$$

ここに、

$$\mathbf{y}(N \times 1), \mathbf{Z} = (\mathbf{1} \quad \mathbf{X}(N \times k)), \mathbf{d} = (b_0 \quad \mathbf{b}'(1 \times k)),$$

$$\mathbf{u}(N \times 1) \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$$

最小 2 乗法で以下の量の最小化を考える。

$$L = (\mathbf{y} - \mathbf{Z}\mathbf{d})'(\mathbf{y} - \mathbf{Z}\mathbf{d})$$

これを回帰係数 \mathbf{d} で微分して、回帰係数の推定値 $\hat{\mathbf{d}}$ を求めると以下となる。

$$\frac{\partial L}{\partial \mathbf{d}} = -2\mathbf{Z}'(\mathbf{y} - \mathbf{Z}\mathbf{d}) = \mathbf{0} \quad \text{より、} \hat{\mathbf{d}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

以後、推定値は真の値 \mathbf{a} に対して $\hat{\mathbf{a}}$ で表す。この $\hat{\mathbf{d}}$ を書き換えると、

$$\hat{\mathbf{d}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\mathbf{d} + \mathbf{u}) = \mathbf{d} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}$$

となり、これを用いると $\hat{\mathbf{d}}$ の平均と分散は以下となる。

$$E[\hat{\mathbf{d}}] = \mathbf{d} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E[\mathbf{u}] = \mathbf{d}$$

$$\begin{aligned} \text{Cov}[\hat{\mathbf{d}}, \hat{\mathbf{d}}] &= E[(\hat{\mathbf{d}} - \mathbf{d})(\hat{\mathbf{d}} - \mathbf{d})'] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E[\mathbf{u}\mathbf{u}']\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \\ &= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} \equiv \Sigma_{\hat{\mathbf{d}}} \end{aligned}$$

これより、【公式 3】を用いると

$$\hat{\mathbf{d}} - \mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u} \sim N(\mathbf{0}, (\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2)$$

よって、

$$(\hat{\mathbf{d}} - \mathbf{d})_i / \sqrt{[(\mathbf{Z}'\mathbf{Z})^{-1}]_{ii}} \sigma \sim N(0, 1) \quad (1)$$

回帰係数の検定

推定値を使って $\mathbf{y} = \mathbf{Z}\hat{\mathbf{d}} + \hat{\mathbf{u}}$ と定義すると、

$$\begin{aligned} \hat{\mathbf{u}} &= -\mathbf{Z}\hat{\mathbf{d}} + \mathbf{y} = -\mathbf{Z}(\hat{\mathbf{d}} - \mathbf{d}) + \mathbf{u} \\ &= -\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u} + \mathbf{u} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u} \end{aligned}$$

表式の簡単化のために、以下の定義をする。

$$\mathbf{P}_Z \equiv \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}', \quad \mathbf{M}_Z \equiv \mathbf{I} - \mathbf{P}_Z$$

対象行列 \mathbf{P}_Z と \mathbf{M}_Z はべき等行列であり、 $\mathbf{P}_Z\mathbf{M}_Z = \mathbf{M}_Z\mathbf{P}_Z = \mathbf{0}$ も成り立つ。

また、べき等行列の固有値は 0 か 1 であることから、以下も分かる。

$$\text{rank}(\mathbf{P}_Z) = k + 1, \quad \text{rank}(\mathbf{M}_Z) = N - k - 1$$

これらを使うと、

$$\hat{\mathbf{u}} = (\mathbf{I} - \mathbf{P}_Z)\mathbf{u} = \mathbf{M}_Z\mathbf{u}$$

ここで、 u_λ の分散 σ^2 が不明であるので、以下の推定値 $\hat{\sigma}^2$ で置き換える。

$$\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} / (N - k - 1) = EV / (N - k - 1)$$

最初にこの置き換えの妥当性をみるために、この式が σ^2 の不偏推定量であることを示す。それには以下の関係を使う。

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}_Z\mathbf{M}_Z\mathbf{u} = \mathbf{u}'\mathbf{M}_Z\mathbf{u}$$

これを用いると、

$$\begin{aligned} E[\hat{\mathbf{u}}'\hat{\mathbf{u}}] &= E[\mathbf{u}'\mathbf{M}_Z\mathbf{u}] = \sigma^2 \sum_{\lambda=1}^N (\mathbf{M}_Z)_{\lambda\lambda} = \sigma^2 \text{tr}(\mathbf{M}_Z) = (N - k - 1)\sigma^2 \\ E[\hat{\sigma}^2] &= E[\hat{\mathbf{u}}'\hat{\mathbf{u}} / (N - k - 1)] = \sigma^2 \end{aligned}$$

次に、 $\mathbf{u}/\sigma \sim N(\mathbf{0}, \mathbf{I}_N)$ であるから、【公式 4】を用いて以下となる。

$$EV / \sigma^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} / \sigma^2 = (\mathbf{u}'/\sigma)\mathbf{M}_Z(\mathbf{u}/\sigma) \sim \chi_{N-k-1}^2 \quad (2)$$

最後に $\hat{\mathbf{d}} - \mathbf{d}$ と EV が独立であることを示す。 \mathbf{M}_Z がべき等行列であり、

$$EV = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}_Z\mathbf{u} = \mathbf{u}'\mathbf{M}_Z\mathbf{M}_Z\mathbf{u}$$

の関係から、 $\hat{\mathbf{d}} - \mathbf{d}$ と $\mathbf{M}_Z\mathbf{u}$ の独立性を示せばよい。

$$\hat{\mathbf{d}} - \mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}, \quad \mathbf{M}_Z\mathbf{u} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u}$$

であるから、

$$\begin{aligned} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\Sigma_u(\mathbf{I}-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' &= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{I}(\mathbf{I}-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' \\ &= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{0} \end{aligned}$$

よって、【公式 2】より、 $\hat{\mathbf{d}} - \mathbf{d}$ と EV は独立である。

(1)式と(2)式、および $\hat{\mathbf{d}} - \mathbf{d}$ と EV の独立性より、

$$t_i \equiv \frac{(\hat{\mathbf{d}} - \mathbf{d})_i}{\sqrt{[(\mathbf{Z}'\mathbf{Z})^{-1}]_{ii} EV / (N - k - 1)}} \sim t_{N-k-1}, \quad EV = \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

結合仮説の検定

ここでは q 個の制約が付いた検定 $\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d}) = \mathbf{0} (q \times 1)$ を考える。これが同時に成り立つことを検定する場合、【公式 3】を利用する。

$$[\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})]'\Sigma_{\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})}^{-1}\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d}) \sim \chi_q^2$$

ここに、 $\text{rank}(\mathbf{R}) = q$

$$\Sigma_{\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})} = \Sigma_{\mathbf{R}\hat{\mathbf{d}}} = \mathbf{R}\Sigma_{\hat{\mathbf{d}}}\mathbf{R}' = \sigma^2\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}'$$

より、 $\mathbf{R}\mathbf{d} = \mathbf{r}$ とすると

$$\begin{aligned} &[\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})]'\Sigma_{\mathbf{R}\hat{\mathbf{d}}}^{-1}\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d}) \\ &= \frac{1}{\sigma^2}(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})'[\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r}) \sim \chi_q^2 \end{aligned}$$

また、(2)式より、

$$EV/\sigma^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2 = \mathbf{u}'\mathbf{M}_x\mathbf{u}/\sigma^2 \sim \chi_{N-k-1}^2$$

さらに、以前の議論から $\hat{\mathbf{d}} - \mathbf{d}$ と EV は独立であるので、

$$F = \frac{(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})'[\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})/q}{EV/(N - k - 1)} \sim F_{q, N-k-1}$$

例として、単独の検定 $b_i = 0$ について考える。その場合は以下である。

$$\mathbf{R}(1 \times (k+1)) = [0 \quad \mathbf{0}(1 \times (i-1)) \quad 1 \quad \mathbf{0}(1 \times (k-i))], \quad r = 0$$

また、有効性の検定 $b_1 = 0, \dots, b_k = 0$ の場合は以下となる。

$$\mathbf{R}(k \times (k+1)) = [\mathbf{0}(k \times 1) \quad \mathbf{I}(k \times k)], \quad \mathbf{r}(k \times 1) = \mathbf{0}$$

また、結合仮説検定 $b_1 = 0, \dots, b_q = 0$ の場合は以下である。

$$\mathbf{R}(q \times (k+1)) = [\mathbf{0}(q \times 1) \quad \mathbf{I}(q \times q) \quad \mathbf{0}(q \times (k-q))], \quad \mathbf{r}(q \times 1) = \mathbf{0}$$

定数項を分離した形式との比較

これらの検定を、定数項を分離する形式と比較してみる。回帰係数の検定では、

$$t_i = \frac{(\hat{\mathbf{d}} - \mathbf{d})_i}{\sqrt{[(\mathbf{Z}'\mathbf{Z})^{-1}]_{ii} EV / (N - k - 1)}} \sim t_{N-k-1}, \quad \mathbf{Z}'\mathbf{Z} = \begin{pmatrix} N & N\bar{\mathbf{x}}' \\ N\bar{\mathbf{x}} & \mathbf{X}'\mathbf{X} \end{pmatrix}$$

$(\mathbf{Z}'\mathbf{Z})^{-1}$ はどのような形をしているのであろうか。そのために、以下の公式を利用する。

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \text{ならば、} \mathbf{A}^{-1} = \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \text{とすると、}$$

$$\mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}, \quad \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}^{-1}, \quad \mathbf{B}_{22} = \mathbf{C}^{-1}$$

$$\text{ここに、} \mathbf{C} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

この公式を用いると、

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \text{として、}$$

$$\mathbf{B}_{11} = 1/N + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}}, \quad \mathbf{B}_{12} = -\bar{\mathbf{x}}'\mathbf{C}^{-1}, \quad \mathbf{B}_{21} = -\mathbf{C}^{-1}\bar{\mathbf{x}}, \quad \mathbf{B}_{22} = \mathbf{C}^{-1}$$

ここに、

$$\mathbf{C} = \mathbf{X}'\mathbf{X} - N\bar{\mathbf{x}}\bar{\mathbf{x}}' = \mathbf{X}'\mathbf{X} - \bar{\mathbf{X}}'\bar{\mathbf{X}} = (\mathbf{X} - \bar{\mathbf{X}})'(\mathbf{X} - \bar{\mathbf{X}})$$

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_k \end{pmatrix}, \quad \bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{x}}' \\ \vdots \\ \bar{\mathbf{x}}' \end{pmatrix} = \begin{pmatrix} \bar{x}_1 & \cdots & \bar{x}_k \\ \vdots & \ddots & \vdots \\ \bar{x}_1 & \cdots & \bar{x}_k \end{pmatrix}$$

以上、まとめると

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{pmatrix} 1/N + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\bar{\mathbf{x}} & \mathbf{C}^{-1} \end{pmatrix}, \quad \mathbf{C} = (\mathbf{X} - \bar{\mathbf{X}})'(\mathbf{X} - \bar{\mathbf{X}})$$

この関係を使って、もう一つ式を作っておく。

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad \bar{\mathbf{y}} = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}$$

$$\begin{aligned}
 (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} &= \begin{pmatrix} 1/N + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\bar{\mathbf{x}} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}' \\ \mathbf{X}' \end{pmatrix} \mathbf{y} \\
 &= \begin{pmatrix} 1/N + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\bar{\mathbf{x}} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} N\bar{y} \\ \mathbf{X}'\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} + N\bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}}\bar{y} - \bar{\mathbf{x}}'\mathbf{C}^{-1}\mathbf{X}'\mathbf{y} \\ -\mathbf{C}^{-1}\bar{\mathbf{x}}N\bar{y} + \mathbf{C}^{-1}\mathbf{X}'\mathbf{y} \end{pmatrix} \\
 &= \begin{pmatrix} \bar{y} + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{X}}'\bar{y} - \bar{\mathbf{x}}'\mathbf{C}^{-1}\mathbf{X}'\mathbf{y} \\ -\mathbf{C}^{-1}\bar{\mathbf{X}}'\bar{y} + \mathbf{C}^{-1}\mathbf{X}'\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{\mathbf{x}}'\mathbf{C}^{-1}(\mathbf{X} - \bar{\mathbf{X}})'(\mathbf{y} - \bar{y}) \\ \mathbf{C}^{-1}(\mathbf{X} - \bar{\mathbf{X}})'(\mathbf{y} - \bar{y}) \end{pmatrix}
 \end{aligned}$$

以上より

$$\begin{aligned}
 t_0 &= \frac{\hat{b}_0 - b_0}{\sqrt{[(\mathbf{Z}'\mathbf{Z})^{-1}]_{00} EV / (N - k - 1)}} \\
 &= \frac{\hat{b}_0 - b_0}{\sqrt{(1/N + \bar{\mathbf{x}}'\mathbf{C}^{-1}\bar{\mathbf{x}}) EV / (N - k - 1)}} \sim t_{N-k-1} \\
 t_i &= \frac{\hat{b}_i - b_i}{\sqrt{[(\mathbf{Z}'\mathbf{Z})^{-1}]_{ii} EV / (N - k - 1)}} = \frac{\hat{b}_i - b_i}{\sqrt{[\mathbf{C}^{-1}]_{ii} EV / (N - k - 1)}} \sim t_{N-k-1}
 \end{aligned}$$

また、有効性の検定は、

$$F = \frac{(\mathbf{R}\hat{\mathbf{b}} - \mathbf{r})'[\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\mathbf{b}} - \mathbf{r})/q}{EV/(N - k - 1)} \sim F_{q, N-k-1}$$

より、 $\mathbf{R}(k \times (k+1)) = [\mathbf{0}(k \times 1) \quad \mathbf{I}(k \times k)]$, $\mathbf{r}(k \times 1) = \mathbf{0}$ として、

$$F = \frac{\hat{\mathbf{b}}'\mathbf{C}\hat{\mathbf{b}}/k}{EV/(N - k - 1)} \sim F_{k, N-k-1}$$

不均一分散誤差の場合の理論

ここでは不均一分散誤差の場合の重回帰分析について説明するが、均一分散の場合にそって一から始めることにする。

目的変数を k 個の説明変数と定数項で回帰する重回帰式を以下のように仮定する。

$$\mathbf{y} = \mathbf{Z}\mathbf{d} + \mathbf{u}$$

ここに、 $\mathbf{y}(N \times 1)$, $\mathbf{Z} = (\mathbf{1} \quad \mathbf{X}(N \times k))$, $\mathbf{d}' = (b_0 \quad \mathbf{b}'(1 \times k))$, $\mathbf{u}(N \times 1)$

$$\mathbf{u}'(1 \times N) = (u_1 \quad \cdots \quad u_N), \quad u_\lambda \sim N(\mathbf{0}, \sigma_\lambda^2), \quad Cov(u_\lambda, u_{\lambda'}) = \sigma_\lambda^2 \delta_{\lambda\lambda'}$$

最小 2 乗法で以下の量の最小化を考える。

$$L = (\mathbf{y} - \mathbf{Z}\mathbf{d})'(\mathbf{y} - \mathbf{Z}\mathbf{d})$$

回帰係数 \mathbf{d} で微分して、回帰係数の推定値 $\hat{\mathbf{d}}$ を求めると以下となる。

$$\frac{\partial L}{\partial \mathbf{d}} = -2\mathbf{Z}'(\mathbf{y} - \mathbf{Zd}) = \mathbf{0} \quad \text{より、} \quad \hat{\mathbf{d}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

この $\hat{\mathbf{d}}$ を書き換えると、

$$\hat{\mathbf{d}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Zd} + \mathbf{u}) = \mathbf{d} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}$$

となり、これを用いると $\hat{\mathbf{d}}$ の平均と分散は、

$$E[\hat{\mathbf{d}}] = \mathbf{d} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E[\mathbf{u}] = \mathbf{d}$$

$$\begin{aligned} \text{Cov}[\hat{\mathbf{d}}, \hat{\mathbf{d}}] &= E[(\hat{\mathbf{d}} - \mathbf{d})(\hat{\mathbf{d}} - \mathbf{d})'] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E[\mathbf{u}\mathbf{u}']\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \end{aligned}$$

ここに、 $(\mathbf{U})_{\lambda\lambda'} = E[u_{\lambda}u_{\lambda'}] = \sigma_{\lambda}^2\delta_{\lambda\lambda'}$

以上より、回帰係数は以下の分布となる。

$$\hat{\mathbf{d}} - \mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u} \sim N(\mathbf{0}, \Sigma_{\hat{\mathbf{d}}}), \quad \Sigma_{\hat{\mathbf{d}}} \equiv (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

分散の推定値としては \mathbf{U} の代わりに自由度を考慮して以下と置く。

$$(\hat{\mathbf{U}})_{\lambda\lambda'} = \frac{N\hat{u}_{\lambda}^2}{N-k-1}\delta_{\lambda\lambda'}$$

計算には上の推定値を用いて、

$$\hat{\mathbf{d}} - \mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u} \sim N(\mathbf{0}, \hat{\Sigma}_{\hat{\mathbf{d}}}), \quad \hat{\Sigma}_{\hat{\mathbf{d}}} \equiv (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{U}}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

回帰係数の検定

上で述べた関係式より、回帰係数の検定には以下の関係を用いる。

$$(\hat{b}_i - b_i) / \sqrt{(\hat{\Sigma}_{\hat{\mathbf{d}}})_{ii}} \sim N(0, 1)$$

結合仮説の検定

ここでは制約が付いた q 個の検定 $\mathbf{R}(\hat{\mathbf{b}} - \mathbf{b}) = \mathbf{0} (q \times 1)$ を考える。これが同時に成り立つことを検定する場合、【公式3】を利用する。

$$[\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})]' \Sigma_{\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})}^{-1} \mathbf{R}(\hat{\mathbf{d}} - \mathbf{d}) \sim \chi_q^2$$

ここに、 $\text{rank}(\mathbf{R}) = q$ とする。また、

$$\Sigma_{\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})} = \Sigma_{\mathbf{R}\hat{\mathbf{d}}} = \mathbf{R}\Sigma_{\hat{\mathbf{d}}}\mathbf{R}' \equiv \mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}'$$

より、 $\mathbf{R}\hat{\mathbf{d}} = \mathbf{r}$ とすると

$$\begin{aligned} &[\mathbf{R}(\hat{\mathbf{d}} - \mathbf{d})]' \Sigma_{\mathbf{R}\hat{\mathbf{d}}}^{-1} \mathbf{R}(\hat{\mathbf{d}} - \mathbf{d}) \\ &= (\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})' [\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\hat{\mathbf{d}} - \mathbf{r}) \sim \chi_q^2 \end{aligned}$$

検定のために推定値を使って以下の関係を利用する。

$$(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})' [\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{U}}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\hat{\mathbf{d}} - \mathbf{r}) \sim \chi_q^2$$

また、F 検定に書式を合わせると以下となる。

$$F = (\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})'[\mathbf{R}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{U}}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{R}]^{-1}(\mathbf{R}\hat{\mathbf{d}} - \mathbf{r})/q \sim F_{q,\infty}$$

2.2 プログラムの利用法

ここでは重回帰分析の中に新しく導入した機能について解説する。1つは重回帰分析の誤差項の不均一分散性の問題である。基本的な重回帰分析では、回帰の誤差は説明変数の値によらず一定と考える。しかし、経済分野では基本的に回帰の誤差は説明変数に依存すると考えることが多い。そのため誤差項の扱いや検定に利用する分布が異なってくる。我々の重回帰分析のプログラムでは図1の上側の枠の「不均一分散回帰分析」チェックボックスにチェックを入れることで、その処理を行う。

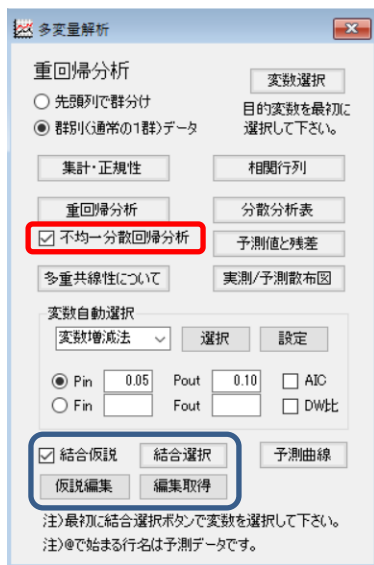


図1 重回帰分析実行画面

卒業試験	入試点数	内申点数	勉強時間	出席率
84	67	3.0	7.4	100
90	71	3.7	8.0	100
80	57	3.9	6.5	78
39	43	2.8	1.8	38
81	63	3.6	6.1	100
47	51	3.7	2.7	55
92	72	4.1	7.9	100
75	62	3.8	4.6	90
77	69	3.6	5.8	89
64	59	3.6	4.2	60
76	65	3.7	5.9	97
97	60	4.3	6.2	100
65	69	3.3	5.5	60
72	67	3.0	5.3	89
49	53	3.4	2.0	46
90	62	3.6	9.0	100
71	62	3.3	4.4	75
54	75	3.7	2.6	57

図2 サンプルデータ

図2のデータで、不均一分散にチェックを入れない場合と入れた場合を分けて、図3a、図3bと図4a、図4bに示す。

卒業試験	偏回帰係数	標準化係数	t検定値	自由度	確率値	相関係数	偏相関係数
▶ 入試点数	0.1490	0.0749	1.3885	45	0.1718	0.386	0.203
内申点数	2.2283	0.0527	1.0327	45	0.3073	0.121	0.152
勉強時間	2.7614	0.3522	3.5671	45	0.0009	0.887	0.469
出席率	0.4314	0.5888	5.9546	45	0.0000	0.920	0.664
切片	6.8654	0.0000	0.7806	45	0.4391		
R	0.943	R ²	0.889	調整済R	0.938	調整済R ²	0.879

図3a 均一分散誤差の結果

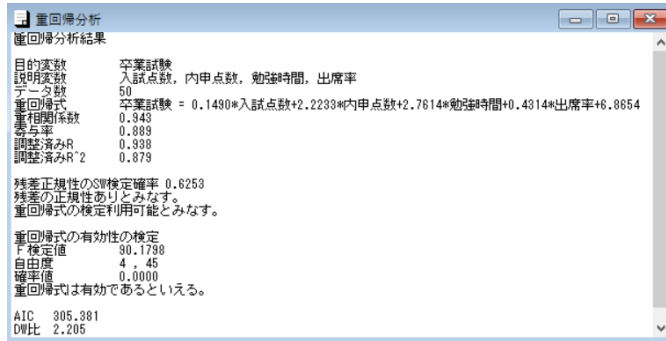


図 3b 均一分散誤差の有効性の検定

卒業試験	偏回帰係数	標準化係数	標準誤差	z統計量	確率値	95%下限	95%上限	相関係数	偏相関係数
▶ 入試点数	0.1490	0.0749	0.1092	1.3647	0.1724	-0.0650	0.3629	0.386	0.203
内申点数	2.2233	0.0527	2.7280	0.8150	0.4151	-3.1295	7.5702	0.121	0.152
勉強時間	2.7614	0.3522	0.7032	3.9271	0.0001	1.3832	4.1396	0.887	0.469
出席率	0.4314	0.5888	0.0702	6.1439	0.0000	0.2938	0.5690	0.920	0.664
切片	6.8654	0.0000	8.1304	0.8444	0.3984	-9.0699	22.8007		
R	0.943	R ²	0.889	調整済R	0.938	調整済R ²	0.879		

図 4a 不均一分散誤差の結果 1

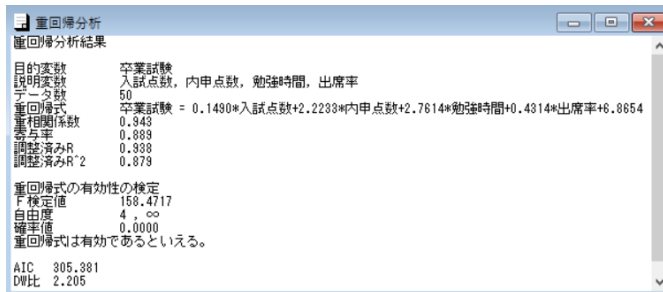


図 4b 不均一分散誤差の有効性の検定

均一分散誤差では、回帰係数の検定に t 検定が、有効性の検定に自由度が有限な F 検定が用いられている。しかし、不均一分散誤差では、回帰係数の検定に正規分布による検定が使われ、標準誤差や 95%の区間推定結果が表示されている。また、有効性の検定では、自由度 4,∞の F 検定（自由度 4 の χ^2 検定に相当）が使われている。

さらに、経済分野での重回帰分析では、結合仮説の検定が使われることもある。重回帰式の有効性の検定は、すべての説明変数の係数が 0 であるという帰無仮説を用いた例であるが、特定の係数の線形結合がある値を取るという検定も使われることがある。そのため我々は、図 1 の下側の枠で囲まれた部分に結合仮説検定の機能を追加した。

「結合仮説」チェックボックスボタンをチェックし、「結合選択」ボタンをクリックすると、図 5 のような設定画面が表示される。

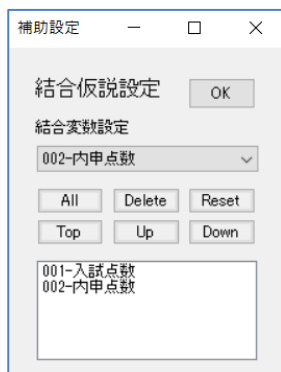


図5 結合仮説検定の変数設定

ここでは2つの変数を選んでいる。この状態で「OK」ボタンを押すと2つの説明変数の係数が共に0であるという設定が出来上がる。そのまま「重回帰分析」ボタンをクリックするとテキストエディタに図6のような結果が追加出力される。

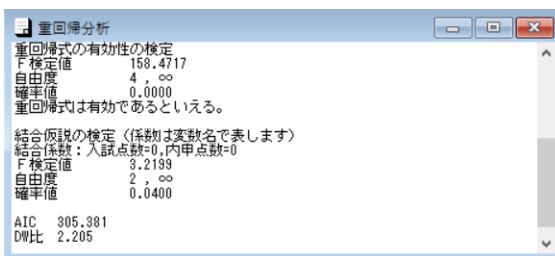


図6 結合仮説検定結果

この中に、「入試点数=0,内申点数=0」というのが見えるが、これが結合仮説の検定で、2つの変数の係数が共に0であることを検定している。ここで、「」内の変数名は変数に付いている係数のことを表している。

次に、例えば、「10*入試点数+内申点数=4」という特殊な検定を行ってみよう。結合仮説設定はそのまま、「仮説編集」ボタンをクリックすると図7のような編集画面が表示される。これが図6の結果を与えた変数選択である。

	入試点数	内申点数	勉強時間	出席率	切片		
▶ 制約1		1				=	0
▶ 制約2						=	0

図7 結合仮説検定編集画面

これを編集して図8のような形にする。

	入試点数	内申点数	勉強時間	出席率	切片		
▶ 制約1	10					=	4
▶ 制約2		1				=	0

図8 結合仮説検定編集後画面

これが「 $10 \times \text{入試点数} + \text{内心点数} = 4$ 」を検定するデータである。この画面を表示したままで「編集取得」ボタンをクリックすると、この設定を読み込み、「重回帰分析」ボタンをクリックすると、テキストエディタに図 9 のような結合仮説検定結果が追加出力される。

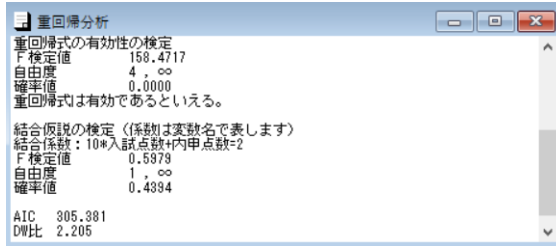


図 9 結合仮説検定結果

ここでは、不均一分散の場合を扱ったが、均一分散の場合でも同様のことを実行することができる。但し、この検定は先頭列で群分けの場合には対応していない。

また現在、結合仮説検定は 2 値ロジスティック分析にも含まれている。2 値ロジスティック分析の分析実行メニューを図 10 に、結合仮説検定を与えた場合に追加されるテキスト出力画面を図 11 に示す。

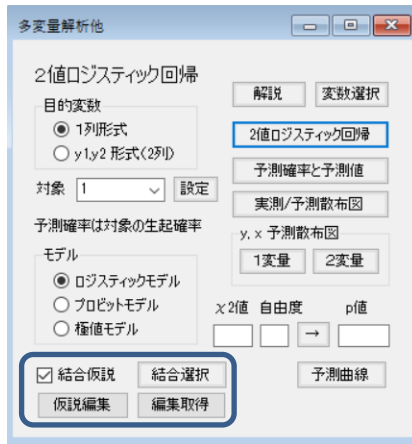


図 10 分析実行画面

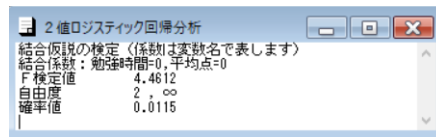


図 11 結合仮説検定結果

3. トービット回帰分析

トービット回帰分析は目的変数のデータに切断のある場合の回帰分析である^[2]。例えば、車への支出と所得との関係を考えて、車を持っている人、または購入する人は車にお金をかけるが、車を持っていない人の車への支出は 0 である。このデータのようにあるところに下限や上限があるデータを切断されたデータと呼ぶ。データの切断はデータの床効果や天井効果とも呼ばれる。データの切断では、切断された値以下または以上のデータはすべて切断された値で置き換えられ、それ以外のデータの情報は不明なものとして扱われる。

例えば、図 1 は 0 で切断のあるデータである。左のグラフの直線は通常の回帰分析で、切断された値はそのままの値として最小 2 乗法で計算されている。それに対して右のグラフでは同じデータにトービット回帰分析を適用した場合の回帰直線が引かれている。

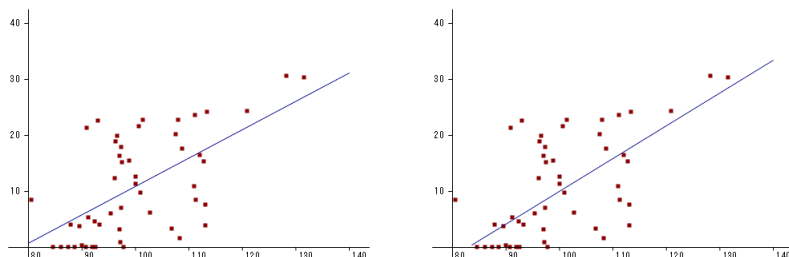


図 1 通常の回帰直線とトービット回帰分析の回帰直線

トービット回帰分析の回帰直線は、通常の回帰直線に比べて、切断データがあたかもマイナスの位置にあるかのように、傾きが少し急になっている。これは、下からの切断の場合、同じ切断の点でも、横軸の負の方向に進むほどウェイトが小さく与えられるというトービット回帰分析の特徴である。

トービット回帰分析は、誤差分布が正規分布に従うという最尤法を使ってパラメータの値を求める。切断データについては、密度関数の値を使わず、切断値を取る確率だけが分かるものとして尤度関数に取り入れられる。これは生存時間分析の打ち切りデータの扱いと同様である。具体的には下からの切断の場合、尤度関数 L_λ は正規分布の密度関数 $f(x)$ と分布関数 $F(x)$ を用いて以下で与えられる。

$$L_\lambda = \prod_{\lambda=1}^N f(y_\lambda - Y_\lambda)^{\delta_\lambda} F(a - Y_\lambda)^{1-\delta_\lambda}, \quad Y_\lambda = \sum_{i=1}^p b_i x_{i\lambda} + b_0$$

ここで、 δ_λ は λ 番目の個体が切断値の場合に $\delta_\lambda = 0$ 、切断値でない場合 $\delta_\lambda = 1$ をとる。また、 y_λ は目的関数の実測値、 Y_λ は目的関数の予測値、 a は切断値である。

下からの切断と同様に上からの切断の場合も考えられる。計算の詳細は次節で述べる。

3.1 トービット回帰分析の理論

トービット回帰分析のモデルは、回帰式の誤差項が正規分布することが元になっている。まず、以下の回帰式を考える。

$$y_\lambda = \sum_{j=0}^p b_j x_{j\lambda} + u_\lambda \quad (x_{0\lambda} = 1; \lambda = 1, \dots, N)$$

誤差項を $u_\lambda \sim N(0, \sigma^2)$ で個体について独立とすると、密度関数は以下になる。

$$f(u_\lambda) = \frac{1}{\sqrt{2\pi\sigma}} \exp[-u_\lambda^2/2\sigma^2] = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2} \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda}\right)^2\right]$$

ここで、標準正規分布の密度関数を $\phi(x)$ とすると、以下のようにも表される。

$$f(u_\lambda) = \frac{1}{\sigma} \phi(u_\lambda/\sigma) = \frac{1}{\sigma} \phi \left[\frac{1}{\sigma} \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right) \right]$$

また、分布関数 $F(x)$ は標準正規分布の分布関数 $\Phi(x)$ を用いて以下のようにも表される。

$$\begin{aligned} F(u_\lambda) &= \int_{-\infty}^{u_\lambda} f(t) dt = \int_{-\infty}^{u_\lambda/\sigma} \phi(x) dx \\ &= \Phi(u_\lambda/\sigma) = \Phi \left[\frac{1}{\sigma} \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right) \right] \end{aligned}$$

ここで、各データに対して、切斷データの場合 $\delta_\lambda = \mathbf{0}$ 、そうでないデータの場合 $\delta_\lambda = 1$ とすると、例えば下からの $y_\lambda = a$ の切斷（床効果）の場合、尤度関数 L は以下となる。

$$L_l = \prod_{\lambda=1}^N f(u_\lambda)^{\delta_\lambda} F(u_\lambda)^{1-\delta_\lambda}$$

また、上からの $y_\lambda = a$ の切斷（天井効果）の場合、尤度関数 L は以下となる。

$$L_u = \prod_{\lambda=1}^N f(u_\lambda)^{\delta_\lambda} [1 - F(u_\lambda)]^{1-\delta_\lambda}$$

$y_\lambda = a$ のとき、 u_λ の値は以下である。

$$u_\lambda = a - \sum_{j=0}^p b_j x_{j\lambda}$$

具体的には、下からの切斷の場合、尤度関数 L_l は以下で与えられる。

$$L_l = \prod_{\lambda=1}^N \left[\frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right)^2 \right\} \right]^{\delta_\lambda} \Phi \left[\frac{1}{\sigma} \left(a - \sum_{j=0}^p b_j x_{j\lambda} \right) \right]^{1-\delta_\lambda}$$

同様に上からの切斷の場合、尤度関数 L_u は以下となる。

$$L_u = \prod_{\lambda=1}^N \left[\frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right)^2 \right\} \right]^{\delta_\lambda} \left\{ 1 - \Phi \left[\frac{1}{\sigma} \left(a - \sum_{j=0}^p b_j x_{j\lambda} \right) \right] \right\}^{1-\delta_\lambda}$$

対数尤度関数はそれぞれ以下となる。

$$\begin{aligned} \log L_l &= -\log \sqrt{2\pi} \sum_{\lambda=1}^N \delta_\lambda - \log \sigma \sum_{\lambda=1}^N \delta_\lambda - \frac{1}{2\sigma^2} \sum_{\lambda=1}^N \delta_\lambda \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right)^2 \\ &\quad + \sum_{\lambda=1}^N (1 - \delta_\lambda) \log \Phi \left[\frac{1}{\sigma} \left(a - \sum_{j=0}^p b_j x_{j\lambda} \right) \right] \end{aligned}$$

$$\begin{aligned} \log L_u = & -\log \sqrt{2\pi} \sum_{\lambda=1}^N \delta_\lambda - \log \sigma \sum_{\lambda=1}^N \delta_\lambda - \frac{1}{2\sigma^2} \sum_{\lambda=1}^N \delta_\lambda \left(y_\lambda - \sum_{j=0}^p b_j x_{j\lambda} \right)^2 \\ & + \sum_{\lambda=1}^N (1-\delta_\lambda) \log \left\{ 1 - \Phi \left[\frac{1}{\sigma} \left(a - \sum_{j=0}^p b_j x_{j\lambda} \right) \right] \right\} \end{aligned}$$

この後、ニュートン・ラフソン法を用いてパラメータの推定を行うが、その際、パラメータを以下のように設定する。

$$\tilde{\mathbf{b}} = \begin{pmatrix} \mathbf{b} \\ \sigma \end{pmatrix}, \quad (\mathbf{b})_i = b_i \quad (i=0, \dots, p)$$

対数尤度をパラメータで微分してスコアベクトル \mathbf{U} と情報行列 \mathfrak{I} を求めると以下となる。

$$\mathbf{U} = \begin{pmatrix} \partial \log L / \partial \mathbf{b} \\ \partial \log L / \partial \sigma \end{pmatrix}, \quad \mathfrak{I} = - \begin{pmatrix} \partial^2 \log L / \partial \mathbf{b} \partial \mathbf{b}' & \partial^2 \log L / \partial \mathbf{b} \partial \sigma \\ \partial^2 \log L / \partial \mathbf{b} \partial \sigma & \partial^2 \log L / \partial \sigma^2 \end{pmatrix}$$

これらを用いてニュートン・ラフソン法でパラメータの推定を行う。

$$\tilde{\mathbf{b}}^{(m+1)} = \tilde{\mathbf{b}}^{(m)} + (\mathfrak{I}^{(m)})^{-1} \mathbf{U}^{(m)}$$

ここで右肩の (m) や $(m+1)$ は、ニュートン・ラフソン法の計算のステップを表す。この情報行列の逆行列の対角成分はパラメータの分散を与える。

ニュートン・ラフソン法のために、定義と計算式を書いておく。

下からの切断の場合

$$\begin{aligned} \sum_{\lambda=1}^N \delta_\lambda = m, \quad Y_\lambda = \sum_{j=1}^p b_j x_{j\lambda} \\ \frac{\partial}{\partial b_i} \log L_l = \frac{1}{\sigma^2} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} (y_\lambda - Y_\lambda) \\ \quad - \frac{1}{\sigma} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \\ \frac{\partial}{\partial \sigma} \log L_l = -\frac{m}{\sigma} + \frac{1}{\sigma^3} \sum_{\lambda=1}^N \delta_\lambda (y_\lambda - Y_\lambda)^2 \\ \quad - \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2}{\partial b_i \partial b_k} \log L_t &= -\frac{1}{\sigma^2} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} x_{k\lambda} \\
 &\quad - \frac{1}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} x_{k\lambda} (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \\
 &\quad - \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} x_{k\lambda} \phi[(a-Y_\lambda)/\sigma]^2 / \Phi[(a-Y_\lambda)/\sigma]^2 \\
 \frac{\partial^2}{\partial b_i \partial \sigma} \log L_t &= -\frac{2}{\sigma^3} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} (y_\lambda - Y_\lambda) \\
 &\quad + \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \\
 &\quad - \frac{1}{\sigma^4} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^2 x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \\
 &\quad - \frac{1}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma]^2 / \Phi[(a-Y_\lambda)/\sigma]^2 \\
 \frac{\partial^2}{\partial \sigma^2} \log L_t &= \frac{m}{\sigma^2} - \frac{3}{\sigma^4} \sum_{\lambda=1}^N \delta_\lambda (y_\lambda - Y_\lambda)^2 \\
 &\quad + \frac{2}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \Phi[(a-Y_\lambda)/\sigma] \\
 &\quad - \frac{1}{\sigma^5} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^3 \phi[-Y_\lambda/\sigma] / \Phi[-Y_\lambda/\sigma] \\
 &\quad - \frac{1}{\sigma^4} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^2 \phi[(a-Y_\lambda)/\sigma]^2 / \Phi[(a-Y_\lambda)/\sigma]^2
 \end{aligned}$$

上からの切断の場合

$$\begin{aligned}
 \frac{\partial}{\partial b_i} \log L_t &= \frac{1}{\sigma^2} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} (y_\lambda - Y_\lambda) \\
 &\quad + \frac{1}{\sigma} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\
 \frac{\partial}{\partial \sigma} \log L_t &= -\frac{m}{\sigma} + \frac{1}{\sigma^3} \sum_{\lambda=1}^N \delta_\lambda (y_\lambda - Y_\lambda)^2 \\
 &\quad + \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\}
 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial b_l \partial b_k} \log L_l &= -\frac{1}{\sigma^2} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} x_{k\lambda} \\ &+ \frac{1}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} x_{k\lambda} (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\ &- \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} x_{k\lambda} \phi[(a-Y_\lambda)/\sigma]^2 / \{1-\Phi[(a-Y_\lambda)/\sigma]\}^2 \\ \frac{\partial^2}{\partial b_l \partial \sigma} \log L_l &= -\frac{2}{\sigma^3} \sum_{\lambda=1}^N \delta_\lambda x_{i\lambda} (y_\lambda - Y_\lambda) \\ &- \frac{1}{\sigma^2} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\ &+ \frac{1}{\sigma^4} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^2 x_{i\lambda} \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\ &- \frac{1}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) x_{i\lambda} (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma]^2 / \{1-\Phi[(a-Y_\lambda)/\sigma]\}^2 \\ \frac{\partial^2}{\partial \sigma^2} \log L_l &= \frac{m}{\sigma^2} - \frac{3}{\sigma^4} \sum_{\lambda=1}^N \delta_\lambda (y_\lambda - Y_\lambda)^2 \\ &- \frac{2}{\sigma^3} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda) \phi[(a-Y_\lambda)/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\ &+ \frac{1}{\sigma^5} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^3 \phi[-Y_\lambda/\sigma] / \{1-\Phi[(a-Y_\lambda)/\sigma]\} \\ &- \frac{1}{\sigma^4} \sum_{\lambda=1}^N (1-\delta_\lambda) (a-Y_\lambda)^2 \phi[(a-Y_\lambda)/\sigma]^2 / \{1-\Phi[(a-Y_\lambda)/\sigma]\}^2 \end{aligned}$$

3.2 プログラムの利用法

メニュー [分析-多変量解析他-予測手法-トービット回帰分析] を選択すると、図 1 で与えられる分析実行画面が表示される。

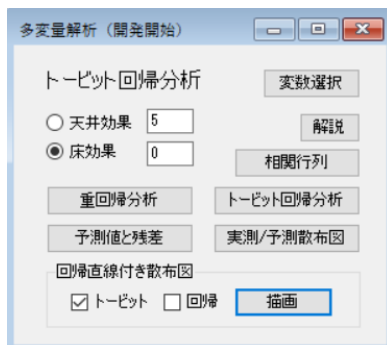


図 1 トービット回帰分析実行画面

データは通常の重回帰分析と同様のデータ形式で、例えば図 2 のように与えられる。

	目的変数	説明変数
▶ 1	12.3	96.1
2	4.1	87.9
3	9.8	100.9
4	22.7	93.0
5	24.4	120.9
6	0	84.6
7	3.4	106.8
8	15.3	112.8
9	16.5	112.0
10	7.7	113.1

図 2 トービット回帰分析データ（トービット回帰 1.txt）

ここでは説明変数が 1 だけの場合であるが、複数の場合も重回帰分析と同じである。「重回帰分析」ボタンをクリックすると比較のための通常の重回帰分析の結果が表示される。但し、トービット回帰の計算に合わせて、誤差項については均一分散として計算されている。結果を図 3 に示す。

目的変数	偏回帰係数	標準化係数	検定値	自由度	確率値	相関係数	偏相関係数
▶ 説明変数	0.5066	0.6045	5.2569	48	0.0000	0.604	0.604
切片	-39.7512	0.0000	-4.0868	48	1.9998		
R	0.604	R ²	0.365	調整済R	0.593	調整済R ²	0.352
F値	27.6347	自由度	1, 48	検定確率	0.0000		

図 3 重回帰分析結果

トービット回帰分析の計算は、まず目的変数に天井効果があるか床効果があるかを選択し、切断値を右のテキストボックスに入力する（両方の場合は必要に応じて加える）。その後、変数を選んで、「トービット回帰」ボタンをクリックすると図 4 で与えられる結果が表示される。

目的変数	偏回帰係数	標準化係数	標準誤差	z統計量	確率値	95%下限	95%上限
▶ 説明変数	0.5855	0.6986	0.1112	5.2671	0.0000	0.3676	0.8033
切片	-48.5193	0.0000	11.2870	-4.2987	0.0000	-70.6414	-26.3973
実測予測R	0.608	R ²	0.369				

図 4 トービット回帰分析結果

ここで、通常の回帰分析の場合、回帰分散と目的変数の分散との比が重相関係数の 2 乗（寄与率）に一致するが、トービット回帰分析の場合は一般に一致しない。どのように表示すべきか考えたが、意味がよく分かる方が良く考え、目的変数値と予測値との相関を採用した。また、予測値は「予測値と残差」ボタンで図 5 のように表示されるが、例えば下からの切断の場合、予測値が切断値より小さい場合は、切断値に置き換えている。

	実測値	予測値	残差
▶ 1	12.3	7.744	4.556
2	4.1	2.944	1.156
3	9.8	10.555	-0.755
4	22.7	5.929	16.771
5	24.4	22.264	2.136
6	0	1.011	-1.011
7	3.4	14.009	-10.609
8	15.3	17.522	-2.222
9	16.5	17.053	-0.553

図5 予測値と残差

「実測/予測散布図」は、この実測値と予測値を使って、図6のように表される。

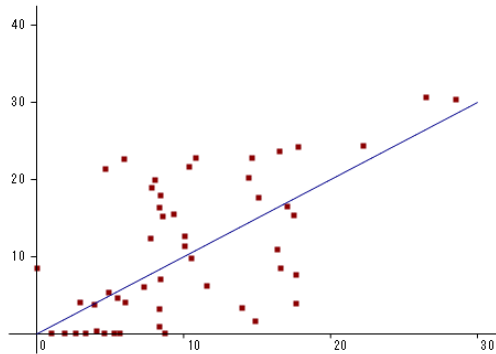


図6 実測/予測散布図

「回帰直線付き散布図」グループボックス内の「描画」ボタンでは、図7のような散布図が表示される。

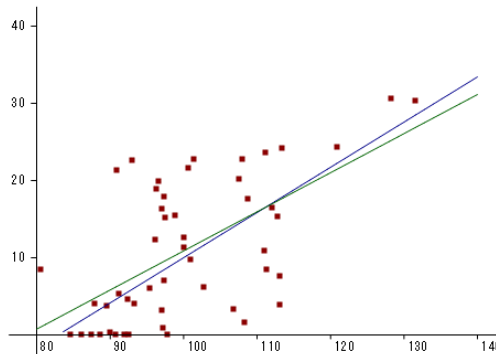


図7 回帰直線付き散布図

ここでは、「トービット」および「回帰」チェックボックスにチェックを入れ、両方の回帰直線を表示している。傾きが大きい方がトービット回帰分析の結果である。

4. おわりに

経済分野の統計処理の大きな特徴の一つは、データ数の多さから、平均値やパラメータの推定値に中心極限定理が適用でき、正規分布を用いた検定が利用できる点である。二つ目は因果関係の証明に回帰分析を使うことが多い点である。そのため、回帰分析の手法が発達し、欠落変数のバイアスを除去するような、特別な分析手法も作られた。

我々はまず、重回帰分析のプログラムの中で、これまで一定と仮定していた回帰分析の誤差分散について、説明変数に依存する場合を取り扱えるようにした。また、回帰係数の特別な組み合わせがある定数となる可能性を調べる、結合仮説の検定を簡単に実行できるようにした。結合仮説の検定はロジスティック回帰分析でも可能なので、2値ロジスティック回帰分析のプログラムにも組み込んだ。

新しい分析としては、経済の分野で作られたトービット回帰分析を加えた。これは床効果や天井効果がある場合の回帰分析である。この手法は他の分野でも利用可能である。今回別の報告で、パネルデータ分析と操作変数回帰分析について紹介したが、残る経済特有の分析では、誤差の時間相関が重要になる時系列を扱う分析がある。今後このような分析についてもプログラム化して行きたい。

参考文献

- [1] Stock, J.H., Watson, M.W., 宮尾龍蔵訳, 「入門計量経済学」, 共立出版, (2016)
- [2] 浅野哲, 中村二郎, 「計量経済学」第2版, 有斐閣, (2009)

補遺 行列と確率分布の公式

ここではこの論文で使う行列と分布の公式をまとめておく。 \mathbf{u} は確率変数である。

【公式1】 $\text{Cov}(\mathbf{A}\mathbf{u}, \mathbf{B}\mathbf{u}) = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{B}'$

【公式2】 $\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{B}' = \mathbf{0}$ ならば、 $\mathbf{A}\mathbf{u}$ と $\mathbf{B}\mathbf{u}$ は独立した分布である。

【公式3】 $\mathbf{u}(m \times 1) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}})$ のとき、

$$\mathbf{d} + \mathbf{A}\mathbf{u} \sim N(\mathbf{d}, \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{A}'), \quad \mathbf{u}'\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{u} \sim \chi_m^2$$

【公式4】 $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_m)$ で $\mathbf{C}(m \times m)$ がべき等行列 ($\mathbf{C}\mathbf{C} = \mathbf{C}$) のとき、

$$\mathbf{u}'\mathbf{C}\mathbf{u} \sim \chi_r^2 \quad \text{但し、} \text{rank}(\mathbf{C}) = r$$

Multi-purpose Program for Social System Analysis 35 - Extension of Regression Analysis, Tobit Regression Analysis -

Masayasu FUKUI^{*1} and Shoko TONAI^{*1}

**1 Department of Business Administration, Faculty of Business Administration,
Fukuyama Heisei University*

Abstract: We have been constructing a unified program on the social system analysis for purpose of education. This time, we made programs for statistical analysis used in the econometrics. In this paper, we introduce the programs for the regression analysis based on the central limit theorem and Tobit regression analysis for the data with floor and ceiling effects.

Key Words: econometrics, Torbit regression analysis, College Analysis, statistics