第 5 章　重回帰モデルによる分析

1　重回帰モデルの古典的仮定

重回帰モデル
$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i \tag{1}$$

古典的仮定：

(i)　k（< n）個の説明変数 $x_{1i},\cdots,x_{ki}$ $(i=1,\cdots,n)$ は，被説明変数 y が実現する以前からわかっており，確率変数ではない．さらに，これらの説明変数および定数項の間には一次従属の関係がない．

(ii) 誤差項 $\varepsilon_1,\cdots,\varepsilon_n$ は互いに独立に同一分布 $N(0,\sigma^2)$ に従う．

(i),(ii) から，$y_1,\cdots,y_n$ は互いに独立であり，各 y_i は
$N(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}, \sigma^2)$ に従う．

2　重回帰モデルの推定

重回帰モデル (1) には，定数項 β_0，回帰係数 $\beta_1,\cdots,\beta_k$ および誤差項の分散 σ^2 の計 $k+2$ 個のパラメータが含まれている．

定数項と回帰係数の推定

まず，$\beta_0,\beta_1,\cdots,\beta_k$ の LSE $\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_k$ は，第 3 章で説明した $k+1$ 本からなる正規方程式の確率変数バージョン

$$\hat\beta_0\sum_{i=1}^n 1 + \hat\beta_1\sum_{i=1}^n x_{1i} + \cdots + \hat\beta_k\sum_{i=1}^n x_{ki} = \sum_{i=1}^n y_i$$
$$\hat\beta_0\sum_{i=1}^n x_{1i} + \hat\beta_1\sum_{i=1}^n x_{1i}^2 + \cdots + \hat\beta_k\sum_{i=1}^n x_{1i}x_{ki} = \sum_{i=1}^n x_{1i}y_i$$
$$\cdots\cdots\cdots\cdots$$
$$\hat\beta_0\sum_{i=1}^n x_{ki} + \hat\beta_1\sum_{i=1}^n x_{ki}x_{1i} + \cdots + \hat\beta_k\sum_{i=1}^n x_{ki}^2 = \sum_{i=1}^n x_{ki}y_i$$

をみたす．

解を表現するためには，行列やベクトルの知識を必要とするのでここでは示さない．しかし，次のことが成り立つことがわかる．

(a) 各 $\hat\beta_h$ は $y_1,\cdots,y_n$ の一次結合で表される

(b) $\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_k$ は，全体として $k+1$ 次元の正規分布に従う

(c) 上の正規方程式の両辺の期待値をとることにより $E(\hat\beta_h)=\beta_h$，すなわち各 $\hat\beta_h$ は不偏であることがわかる

以上より，各 $\hat\beta_h$ の分布は
$$\hat\beta_h = \sum_{i=1}^n c_{hi}y_i \quad \sim \quad N(\beta_h, \sigma^2\sum_{i=1}^n c_{hi}^2) \tag{2}$$
のように表現できる．ただし，c_{hi} は説明変数のみに依存する非確率的な値である．

古典的仮定のもとでの LSE $\hat{\beta}_h$（$h=0,1,\cdots,k$）の性質

(a) 最尤推定量と一致する.

(b) 最小分散不偏推定量である.

(c) 正規性を仮定しない場合は，各 $\hat{\beta}_h$ は，$y_1,\cdots,y_n$ の一次結合で表されるすべての不偏推定量の中で分散が最小となる．この特性を，重回帰モデルの推定量に関する**ガウス・マルコフの定理**という．このとき，LSE は BLUE と呼ばれる.

分散の推定

誤差項の分散 σ^2 の推定量として

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2 \tag{3}$$

を考えると，$\hat{\sigma}^2$ は σ^2 の不偏推定量となることが示される.

信頼区間

(3) の $\hat{\sigma}^2$ は 各 $\hat{\beta}_h$ と独立であり，$(n-k-1)\hat{\sigma}^2/\sigma^2$ が自由度 $n-k-1$ の χ^2 分布に従うことが知られている．このことから,

$$\frac{\hat{\beta}_h - \beta_h}{\sqrt{\hat{V}(\hat{\beta}_h)}} \sim t(n-k-1) \tag{4}$$

ただし，$\hat{V}(\hat{\beta}_h) = \hat{\sigma}^2 \sum_{i=1}^{n} c_{hi}^2$ である．このことから，信頼係数 $1-\alpha$ の β_h に関する信頼区間は，不等式

$$\hat{\beta}_h - t_{\alpha/2}^{(n-k-1)} \times \sqrt{\hat{V}(\hat{\beta}_h)} \le \beta_h \le \hat{\beta}_h + t_{\alpha/2}^{(n-k-1)} \times \sqrt{\hat{V}(\hat{\beta}_h)} \tag{5}$$

に基づいて構成することができる.

〔**例題 5.1**〕**付録** 4 の年次マクロ・データにおいて，輸入額を GDE と交易条件指数（＝輸出デフレーター／輸入デフレーター）で説明する重回帰式を求め，信頼度 90% で輸入の GDE 弾力性の信頼区間を計算せよ．ただし，変数についてはすべて対数変換せよ.

（解）輸入額を y，GDE を x，交易条件指数を z とすると，重回帰式は

$$\log y = -2.768 + 1.119\log x + 0.195\log z \qquad R^2 = 0.915\,,\ \overline{R}^2 = 0.907\,,\ \hat{\sigma} = 0.093$$

(-5.37) (12.69)　　(1.62)

となる．したがって，GDE 弾力性の推定値は 1.119 であり，標準誤差は 1.119/12.69=0.088 となる．また，$n=24, k=2, \alpha=0.1$ より，$t_{\alpha/2}^{(n-k-1)} = 1.721$ であるから，求める信頼区間の下限と上限は

1.119±1.721×0.088 = 0.97, 1.27

となる．GDE 弾力性は信頼度 90% でほぼ弾力的であるといえる.

3　線形制約仮説の F 検定

重回帰モデルの個々のパラメータに関する検定は，t 検定で実行できる.

例えば，検定問題

$$H_0\!:\beta_h=\beta_{h0} \qquad \text{vs} \qquad H_1\!:\beta_h\neq\beta_{h0}$$

は，

$$\frac{|\hat{\beta}_h-\beta_{h0}|}{\sqrt{\hat{V}(\hat{\beta}_h)}}\leq t_{\alpha/2}^{(n-k-1)}$$

のときに帰無仮説を受容し，さもなければ棄却する．

特に，$\beta_{h0}=0$ のときの統計量 $\hat{\beta}_h/\sqrt{\hat{V}(\hat{\beta}_h)}$ は，β_h を係数にもつ説明変数 x_h の有意性を検定する t 値を与えるものである．

重回帰モデルにおいては，単回帰モデルとは異なり，複数の回帰係数の線形制約に関する仮説，例えば

$$H_0\!:\beta_1+\beta_2=1,\beta_3=\beta_{30},\beta_4=\beta_5,\cdots$$

のような帰無仮説を検定する必要が生じる．この検定は t 検定では不可能である．そのために使われる検定が F **検定**である．

複数の回帰係数の線形制約に関する検定

$RSS\,(H_0)=$ 帰無仮説のもとでの残差 2 乗和

$RSS\,(H_1)=$ 対立仮説のもとでの残差 2 乗和

(i) F 統計量

$$F=\frac{\left(RSS(H_0)-RSS(H_1)\right)/p}{RSS(H_1)/(n-k-1)} \tag{6}$$

を計算する．ここで，p は帰無仮説に課せられる独立な制約条件の数である．

(ii) 有意水準を α とするとき，

$$F \text{ 値 } \leq F_\alpha(p,n-k-1)$$

ならば H_0 を受容し，さもなければ棄却する．ただし，$F_\alpha(p,n-k-1)$ は自由度 $(p,n-k-1)$ の F 分布の上側 100α% 点である．

次の例題は制約条件が 1 個の F 検定の例である．その場合の F 検定は，両側 t 検定と同等であることを注意しておく．

〔**例題 5.2**〕例題 3.3 で取り上げた生産関数（データは**付録 2**）の規模に関する収穫不変性を検定せよ．

（解）付加価値額を y，労働投入額を x_1，資本設備額を x_2 とする．

モデル： $\log y=\beta_0+\beta_1\log x_1+\beta_2\log x_2+\varepsilon$

検定問題： $H_0\!:\beta_1+\beta_2=1 \qquad \text{vs} \qquad H_1\!:\beta_1+\beta_2\neq 1$

帰無仮説のもとでの重回帰式：

$$\log y=1.069+0.637\log x_1+0.363\log x_2 \qquad R^2=0.943, \overline{R}^2=0.941\,, \hat{\sigma}=0.185$$

(8.12)　(8.45)　　(4.81)　　残差 2 乗和＝0.85574

制約なしの重回帰式：

$$\log y = 1.171 + 0.603\log x_1 + 0.376\log x_2 \qquad R^2 = 0.944, \overline{R}^2 = 0.939, \hat{\sigma} = 0.188$$

(3.58)（4.79） (4.40) 残差 2 乗和＝0.85163

$$F \text{ 値 } = \frac{(0.85574 - 0.85163)/1}{0.85163/(27-3)} = 0.116$$

自由度（1,24）の F 分布の上側 70% 点は 0.152，80% 点は 0.066 であることから，帰無仮説は有意水準 70% でも受容されることがわかる．したがって，与えられたデータから規模に関する収穫不変性を反証することはできない．

〔**例題** 5.3〕例題 3.5 で取り上げた貯蓄関数（データは**付録** 3）において，季節ダミーの有意性を検定せよ．

（解）

モデル： $y = \beta_0 + \beta_1 x + \gamma_1 q_1 + \gamma_2 q_2 + \gamma_3 q_3 + \varepsilon$ 　q_1, q_2, q_3：季節ダミー

帰無仮説： $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$ のもとでの回帰式

$$y = -11.14 + 0.321x \qquad R^2 = 0.468, \overline{R}^2 = 0.459, \hat{\sigma} = 5.718$$

(-3.96)　(7.38)　残差 2 乗和＝2026.90

制約なしの重回帰式,

$$y = 10.23 + 0.102x - 16.466q_1 - 3.721q_2 - 10.385q_3$$

(5.84)　(4.70)（-16.47）（-4.37）　(-11.68)

$R^2 = 0.918, \overline{R}^2 = 0.913, \hat{\sigma} = 2.298$　残差 2 乗和＝311.69

$$F \text{ 値 } = \frac{(2026.90 - 311.69)/3}{311.69/(64-5)} = 108.22$$

自由度（3,59）の F 分布の上側 0.01% 点は 8.378 であるから，帰無仮説は有意水準 0.0001 でも棄却され，季節ダミーの有意性が示唆される．

重回帰モデルにおけるもっとも基本的な F 検定：

$H_0 : \beta_1 = 0, \cdots, \beta_k = 0$ $\Leftrightarrow$ すべての説明変数が説明力をまったくもたない

単に F 値というときは，この値を意味する．

4　グループ間の差の検定

図 5-1： 企業の設備投資の伸び率（前年同期比：%）と設備稼働率の伸び率（前年同期比：%）に関する四半期データの散布図

白丸・・ 2 期前の稼働率水準が 90% 以上の四半期,

黒丸・・ 2 期前の稼働率水準が 90% 未満の四半期

実際のデータは**付録 6**（平成 6 年版「経済白書」からの抜粋）

分析の目的：設備投資の伸びが稼働率のレベルに依存しているかどうかということ.
図の実線：稼働率が 90% 以上の時期における回帰直線
図の点線：稼働率が 90% 未満の時期における回帰直線

統計的に有意かどうかは次のように考える.

〔**例題 5.4**〕**付録 6** の設備投資のデータにおいて，製造業における設備投資の伸び率と稼働率の伸び率の関係は，2 つの稼働率のレベルに依存して異なるかどうかを検定せよ.

（解）設備投資の伸び率を y，稼動率の伸び率を x，稼働率水準が 90% 以上のとき 1, 90% 未満のとき 0 をとるダミー変数を d とする.

モデル：　$y=\beta_0+\beta_1 x+\gamma_1 d+\gamma_2(d\times x)+\varepsilon$

帰無仮説：$H_0:\gamma_1=\gamma_2=0$ のもとでの重回帰式

$$y=5.423+2.022x \qquad R^2=0.541, \bar{R}^2=0.534, \hat{\sigma}=8.941$$

(4.99)（8.81)　　残差 2 乗和＝5276.689

制約なしの重回帰式

$$y=-4.493+0.711x+12.329d+1.456(d\times x) \qquad R^2=0.714, \bar{R}^2=0.701, \hat{\sigma}=7.159$$

残差 2 乗和＝3279.892

,

$$F\text{ 値 }=\frac{(5276.689-3279.892)/2}{3279.892/(68-4)}=19.482$$

自由度（2,64）の F 分布の上側 1% 点は 4.953 であることから，帰無仮説は有意水準 1% でも容易に棄却され，2 つの回帰の間には有意な差があると結論される.

5　構造変化の検定

チャウ検定： 時系列データの回帰分析においては，ある時点を境にして構造変化が起きているかどうかが問題になる場合が多い．この場合に使われる F 検定のこと.

図 5-2： 家計の平均消費性向（＝家計最終消費支出／家計可処分所得）の年次データ（単位：%）である.

このデータから，1975 年頃を境にして平均消費性向がそれまでの減少傾向から増加傾向に転じていることがわかる．以下の例題では，平均消費性向の変動を所得，金融資産，物価の 3 つの要因で説明する重回帰モデルを作り，構造変化を検定してみよう.

〔**例題 5.5**〕**付録 7** のデータにおいて，平均消費性向（y）を，家計可処分所得の伸び率（x_1），金融資産の純増が可処分所得に占める割合（x_2），物価上昇率（x_3）で説明する重回帰モデルを考えて，1975 年までと以後とではモデルが異なるかどうかを検定せよ.

（解）1975 年までは 0，それ以後は 1 をとるダミー変数を d とすると，モデルは

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \gamma_0 d + \gamma_1 (d \times x_1) + \gamma_2 (d \times x_2) + \gamma_3 (d \times x_3) + \varepsilon$$

帰無仮説 $H_0 : \gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 0$ のもとでの重回帰式：

$$y = 86.73 + 0.0682 x_1 - 0.0649 x_2 - 0.476 x_3$$

(29.1) (0.58) (-0.42) (-5.46)

$R^2 = 0.526, \bar{R}^2 = 0.479$, $\hat{\sigma} = 2.204$ 残差 2 乗和＝145.736

制約なしの重回帰式：

$$y = 90.93 + 0.398 x_1 - 0.499 x_2 - 0.182 x_3$$

(26.4) (1.69) (-3.17) (-1.31)

$$-10.88d - 1.598(d \times x_1) + 1.084(d \times x_2) - 0.948(d \times x_3)$$

(-2.57) (-4.52) (5.02) (-5.02)

$R^2 = 0.843, \bar{R}^2 = 0.800$ 残差 2 乗和＝48.395

$$F \text{ 値 } = \frac{(145.736 - 48.395)/4}{48.395/(34-8)} = 13.07$$

自由度 (4,26) の F 分布の上側 1% 点は 4.140 であることから，帰無仮説は有意水準 1% でも容易に棄却されることがわかる．