回帰分析（重回帰）

麻生良文

重回帰分析の厳密な議論のためには線形代数の知識が必要になります。線形代数の知識がないと，以下の議論の理解は難しいと思います。その場合でも，ここでの議論に目を通し，最小二乗推定量の統計的性質や仮説検定の考え方の概略を把握してください。入門的な計量経済学の教科書ではここで展開されているような議論はほとんど省略されています（正確な理解のためには大学院レベルの教科書を読む必要があります)。

1.前提

重回帰モデル (multiple regression model) とは説明変数が複数個ある回帰モデルのことである。k 個の説明変数からなるモデルを考える。

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

ここで，$\alpha, \beta_1, \beta_2, \ldots, \beta_k$ は定数で，y, x_i, u は次のようによばれる（単回帰の場合と全く同様)。

- y：被説明変数(explained variable)，従属変数(dependent variable), regressand
- x_i :(i=1,2,..,k) 説明変数(explanatory variable)，独立変数(independent variable), regressor
- u: 誤差項(error term)，撹乱項(disturbance term)

モデルの特徴

このモデルは次のような特徴を持つ。

- 線型モデル
- x 以外の効果は誤差項に集約されていると考える。

誤差項には，他の変数，モデルで想定していない変数，観察不可能な変数の影響や y の測定誤差が反映されていると考えられる。

回帰分析の前提

観察された(x,y)からパラメータ$\alpha, \beta_1, \beta_2, \ldots, \beta_k$ を推定するためには，誤差項の性質についていくつかの前提を置く必要がある。通常，次のような仮定がおかれる（単回帰の場合と同じで仮定 7 が付け加わるだけ)。

（仮定 1）線型性

真のモデルが次の式で表される（α，βに関して線形関数という点が重要)。

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$$

α, β_1,β_2, …, β_k （および誤差項の分散）が推定すべき未知パラメータである。また，(x_{1i}, x_{2i}, ,...,x_{ki}, y_i)は i 番目の観測値。u_i は誤差の実現値を表す。

（仮定 2）誤差項の期待値は 0（すべての i について)。

$$\mathrm{E}\ (u_i) = 0$$

単回帰の場合と同様に，定数項αが存在するため，この仮定は何ら制約的ではない。

（仮定 3）誤差項の分散はすべての i について等しい（分散均一性 homoskedasticity)。

$$\mathrm{var}\ (u_i) = \sigma^2$$

（仮定 4）誤差項に系列相関は存在しない。

誤差項 u_i と u_j の共分散は 0 である(ただし，$i \neq j$)。

$$\mathrm{cov}\ (u_i, u_j) = 0 \quad \text{for all } i \neq j$$

（仮定 5) 説明変数と誤差項の独立性。

説明変数 x（k 個の説明変数からなるベクトル）と誤差項 u はすべての x の成分と独立である。古典的回帰モデルでは，x は非確率変数であると仮定される。その場合には自動的にこの仮定は満たされる。

なお，現在の教科書のほとんどは，x を非確率変数とせず,x が与えられた場合の誤差項の条件付分布について(仮定 2)以下が成り立つという前提で議論を進めている。

（仮定 6）正規分布の仮定

誤差項の確率分布は正規分布に従う。(仮定 2)，(仮定 3)，(仮定 4）とこの仮定をあわせると，誤差項は互いに独立で同一の正規分布 $N(0,\sigma^2)$に従う。

$$u_i \sim N(0, \sigma^2) \quad i.i.d.$$

（*i.i.d.*は independently identically distributed の略）

（仮定 7）説明変数 $x_1,x_2,\ldots,x_k$ の間に多重共線性の問題は無い（詳細は後述）。

2.最小二乗法

単回帰の場合と全く同様な考え方である。最小二乗法においては，残差平方和を最小にするようにα,β₁,β₂,...,β$_k$ の推定値を決定する。$\alpha,\beta_1,\beta_2,\ldots,\beta_k$ の推定値を $a,b_1,b_2,\ldots,b_k$ で表すと，残差は次の式で与えられる[1]。

$$e_i = y_i - \hat{y}_i = y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki} \tag{1}$$

したがって，残差平方和は次の式で与えられる。

$$S = \sum_{i=1}^{n}(y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki})^2 \tag{2}$$

(2)式は，$a,b_1,b_2,\ldots,b_k$ の 2 次関数である。(2)式の最小化のための必要条件を求めると次の通りになる（$a,b_1,b_2,\ldots,b_k$ で微分して 0）。

$$-2\sum_i(y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki}) = 0$$

$$-2\sum_i(y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki})\,x_{1i} = 0$$

$$\cdots$$

$$-2\sum_i(y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki})\,x_{ji} = 0$$

$$\cdots$$

$$-2\sum_i(y_i - a - b_1x_{1i} - b_2x_{2i} - \cdots - b_kx_{ki})\,x_{ki} = 0$$

これらの式の両辺を−2 で割って整理すると次の式を得る。これを正規方程式 normal equation と呼ぶ。

---

[1] x_{ji} は， j 番目の説明変数の i 番目のオブザベーションであることに注意。

$$\sum_i y_i = \sum_i \left(a + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_j x_{ji} + \cdots + b_k x_{ki}\right)$$

$$\sum_i x_{1i} y_i = \sum_i \left(a x_{1i} + b_1 {x_{1i}}^2 + b_2 x_{1i} x_{2i} + \cdots + b_j x_{1i} x_{ji} + \cdots + b_k x_{1i} x_{ki}\right)$$

$$\cdots$$

$$\sum_i x_{ji} y_i = \sum_i \left(a x_{ji} + b_1 x_{ji} x_{1i} + b_2 x_{ji} x_{2i} + \cdots + b_j {x_{ji}}^2 + \cdots + b_k x_{ji} x_{ki}\right)$$

$$\cdots$$

$$\sum_i x_{ki} y_i = \sum_i \left(a x_{ki} + b_1 x_{ki} x_{1i} + b_2 x_{ki} x_{2i} + \cdots + b_j x_{ki} x_{ji} + \cdots + b_k {x_{ki}}^2\right)$$

これらは，k+1 個の未知パラメータ $a,b_1,b_2,\ldots,b_k$ に関する k+1 本の連立 1 次方程式である。この方程式を $a,b_1,b_2,\ldots,b_k$ について解くことで，最小二乗法の推計値が求まる。なお，k=2 の場合については，(4)の最初の方程式を残りの 2 本の方程式に代入することで，2 本の連立方程式に帰着させることができ，解析的に解を求めることができるが，k が 3 以上になると，行列を用いないと解析的な解を求めるのは困難になる。

さて，正規方程式の最初の式を n でわると次の式を得る。

$$\bar{y} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + \cdots b_k \bar{x}_k$$

ここで，$\bar{y} = (1/n)\sum_i y_i\,, \bar{x}_j = (1/n)\sum_i x_{ji}\,, (j = 1,2,\ldots,k)$ で被説明変数および説明変数の標本平均を表す。この式は，回帰直線は必ず点$(\bar{y}, \bar{x}_1 \cdots, \bar{x}_k)$を通ることを意味しており，単回帰の場合にも成立していた性質である。また，正規方程式から，

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} x_{ji} e_i = 0 \quad \text{for } j = 1,2,\ldots,k$$

が導かれる。これは，残差が説明変数と必ず直交する（無相関である）ことを意味する[2]。

次に，正規方程式を行列で表現する。まず，n 個のオブザベーション y_i (i=1,2,..,n)からなるベクトルを y で表す。また，定数項の 1 を第 1 列として，j 番目の説明変数 x_j の n 個のオブザベーションを第 j+1 列目の要素として持つ行列 X を考える。y と X は次の式で与えられる。

---

[2] $\sum_i e_i = 0$は定数項と残差が直交するという条件であることに注意せよ。

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}$$

y は n 次元ベクトル（$n\times 1$ 行列），Xは $n\times(k+1)$行列である。
　さて，正規方程式の左辺は

$$\begin{bmatrix} y_1 + y_2 + \cdots + y_n \\ x_{11}y_1 + x_{12}y_2 + \cdots + x_{1n}y_n \\ \cdots \\ x_{k1}y_1 + x_{k2}y_2 + \cdots + x_{kn}y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}' \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{X}'\boldsymbol{y}$$

と表される。ここで，$\boldsymbol{X'}$は $\boldsymbol{X}$ の転置行列 (transposed matrix)で，$(k+1)\times n$ 行列で，$\boldsymbol{X'y}$ は $k+1$ 次元ベクトルになる。また，右辺は次のように書き直すことができる。

$$\begin{bmatrix} n & \sum_i x_{1i} & \cdots & \sum_i x_{ki} \\ \sum_i x_{1i} & \sum_i {x_{1i}}^2 & \cdots & \sum_i x_{1i}x_{ki} \\ \vdots & \vdots & & \vdots \\ \sum_i x_{ki} & \sum_i x_{ki}x_{1i} & \cdots & \sum_i {x_{ki}}^2 \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}' \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = X'Xb$$

ここで, 最後の b は係数の推定値からなら $k+1$ 次元の列ベクトルで $\boldsymbol{b}=(a,b_1,b_2,...,b_k)'$である。結局，行列を用いると正規方程式は次のように表現できる。

$$\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \tag{3}$$

(3)式において，$(k+1)\times(k+1)$行列である $\boldsymbol{X'X}$ の逆行列が存在すれば（$\boldsymbol{X'X}$ の階数(rank)が $k+1$ なら，逆行列は存在する），

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{4}$$

が得られる。(4)式は，最小二乗推定量 b が y の線型関数であることを示している。

以上の結果を，最初から行列を用いて示しておこう。まず，y_i（被説明変数 y の i 番目のオブザベーション）が次のように書けたとする。

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$$

説明変数のベクトルの i 番目のオブザベーションの x_i で表し（定数項 1 を含んだ k+1 次元の行ベクトル），係数ベクトルをβで表すことにしよう（βは k+1 次元の列ベクトルであるとする）。すなわち，

$$\boldsymbol{x_i} = [1 \quad x_{1i} \quad x_{2i} \quad \ldots \quad x_{ki}]$$

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

とする。このとき，y_i は

$$y_i = \boldsymbol{x_i}\boldsymbol{\beta} + u_i$$

とコンパクトに表すことができる。さらに， y_i，u_i の n 個のオベザベーションを縦に並べてできる n 次元の列ベクトルを y，u で表す。同様に，x_i の n 個のオブザベーションを縦に並べてできる n×(k+1)行列を X で表す。(6)式を i=1,2,..,n について縦に並べて表示すると

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & \vdots & \cdots & \vdots \\ 1 & x_{1i} & \cdots & x_{ki} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_i \\ \vdots \\ u_n \end{bmatrix}$$

となる。つまり，モデルは

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{5}$$

と表せる。なお，行列 $\boldsymbol{X}$ の階数は k+1 であるとする(列数に等しい)。これが回帰分析の前提で述べた仮定 7（多重共線性は存在しない）の正確な表現である（行列 $\boldsymbol{X}$ の階数が k+1 のとき，$\boldsymbol{X'X}$ の逆行列は存在する）。また，誤差項に関する仮定 2 から仮定 6 は

$$\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \tag{6}$$

と表すことができる。$\boldsymbol{0}$ は全ての要素が 0 であるような n 次元の列ベクトル（ゼロ・ベクトル），$\boldsymbol{I}$ は n 次元の単位行列を表す。

さて，真のモデルが(5)式で表されるとき，βの最小二乗推定量を $\boldsymbol{b}$ として，残差 $\boldsymbol{e}$ は次の式で与えられる（$\boldsymbol{e}$ は第 i 番目の要素が第 i 番目のオブザベーションの残差 e_i であるような n 次元の列ベクトルである）。

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{Xb} \tag{7}$$

そして，残差平方和 S は次のとおりになる。

$$S = \boldsymbol{e}'\boldsymbol{e} = [\boldsymbol{y} - \boldsymbol{Xb}]'[\boldsymbol{y} - \boldsymbol{Xb}] = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{Xb} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{Xy} = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{Xb}$$

ここで $\boldsymbol{y'}, \boldsymbol{e'}, \boldsymbol{X'}$は，$\boldsymbol{y}, \boldsymbol{e}, \boldsymbol{X}$ の転置行列である。S 最小化の 1 階の条件は

$$\frac{\partial S}{\partial \boldsymbol{b}} \equiv \begin{bmatrix} \partial S/\partial a \\ \partial S/\partial b_1 \\ \vdots \\ \partial S/\partial b_k \end{bmatrix} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{Xb} = \boldsymbol{0} \tag{8}$$

となり，b に関する方程式が得られる（正規方程式）。これを $\boldsymbol{b}$ について解くと，

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{9}$$

が得られる。(9)式を(7)式に代入すると，残差は

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{Xb} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$$

となるが，この式の両辺に左から $\boldsymbol{X'}$をかけると

$$\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$$

という関係が導かれる。これは，残差と説明変数（定数項も含む）が直交していることを意味する。

行列の微分

$\boldsymbol{a}$ と $\boldsymbol{x}$ が次のような列ベクトルであるとする。

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad , \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$y = \boldsymbol{a}'\boldsymbol{x} = \sum_{i=1}^{n} a_i x_i$として，$y$ の x に関する微分を考える。

$$\frac{\partial y}{\partial x} = \frac{\partial}{\partial x}(\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial x}(\boldsymbol{x}'\boldsymbol{a}) = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \vdots \\ \partial y/\partial x_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{a}$$

が成り立つ。

次に 2 次形式の微分を求めよう。$\boldsymbol{A}$ は $n \times n$ の対称行列で，(i,j)成分が a_{ij} で与えられているものとする。$\boldsymbol{A}$ は対称行列なので $a_{ij}=a_{ji}$ が成立する。このとき $y=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ の x に関する微分を考える。まず，

$$y = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_i \sum_j a_{ij} x_i x_j$$

である。このとき，

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \vdots \\ \partial y/\partial x_n \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 + \cdots + 2a_{1n}x_n \\ 2a_{21}x_1 + 2a_{22}x_2 + \cdots + 2a_{2n}x_n \\ \vdots \\ 2a_{n1}x_1 + 2a_{n2}x_2 + \cdots + 2a_{nn}x_n \end{bmatrix} = 2\boldsymbol{A}\boldsymbol{x}$$

が成立する。

3.最小二乗推定量の確率分布

最初に述べた回帰モデルの前提（仮定 1 から仮定 7）をもう一度述べておこう。まず，真のモデルが次のように表せるとする（線形モデルの仮定）。

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$$

また，誤差項に関する仮定 2 から仮定 6 が成り立つとき，誤差項は期待値が 0，分散が$\sigma^2 I$であるような多変量正規分布に従う。すなわち，

$$\mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}$$
$$\mathrm{var}(\boldsymbol{u}) = \sigma^2 \boldsymbol{I}$$

が成立するものとする。なお，上式の $\boldsymbol{0}$ は n 個の 0 からなるベクトル，$\boldsymbol{I}$ は n 次元の単位行列である。b の確率分布を求めるために，(9)式に $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$を代入すると

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta} + \boldsymbol{u}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \qquad (10)$$

が得られる。したがって, 誤差項 $\boldsymbol{u}$ が正規分布に従えば, $\boldsymbol{b}$ も正規分布に従うことがわかる。なお，$\boldsymbol{b}$ の期待値は，(10)式より，

$$\mathrm{E}(\boldsymbol{b}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}(\boldsymbol{u}) = \boldsymbol{\beta} \qquad (11)$$

となり， $\boldsymbol{b}$ は不偏性（期待値が真の値に一致する性質）を持つことがわかる。次に，b の分散（分散共分散行列）を求める。(10)式と(11)式より，$\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}) = \boldsymbol{b} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$であるから，

$$(\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}))(\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}))' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

となる。したがって，

$$\begin{aligned}\mathrm{var}(\boldsymbol{b}) &= \mathrm{E}[(\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}))(\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}))'] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}(\boldsymbol{u}\boldsymbol{u}')\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

が得られる。特に，j 番目の説明変数の係数 b_j の期待値と分散は次の通りになる。

$$\mathrm{E}(b_j) = \beta_j$$

$$\mathrm{var}(b_j) = \sigma^2 a_{jj} = \frac{\sigma^2}{S_{xx}^j}$$

ここで，a_{jj} は$(X'X)^{-1}$の$(j+1,j+1)$要素[3]，S_{xx}^j は j 番目の説明変数「固有」の平方和である（ここで「固有」とは，j 番目の説明変数を定数項＋他の k-1 個の説明変数で回帰し，それらの説明変数で説明できる部分を除いた x_j 独自の要因に起因する平方和という意味である[4]）。そして，これらの結果は，単回帰の場合に成立していた性質が重回帰の場合にも成立することを表している。以上の結果をまとめると，最小二乗推定量は，期待値が$\boldsymbol{\beta}$，分散（分散・共分散行列）が$\sigma^2(\boldsymbol{X'X})^{-1}$の（$k$+1 次元の）多変量正規分布に従う。あるいは，

$$\boldsymbol{b} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X'X})^{-1})$$

が成立する。σ^2 は誤差項の分散で未知のパラメータであることに注意。

残差の確率分布

次に最小二乗残差の性質を述べる。まず，残差は

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{Xb} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}]\boldsymbol{y}$$

で与えられる。これから，残差 $\boldsymbol{e}$ は $\boldsymbol{y}$ の線型関数になっていることがわかる。今，

$$\boldsymbol{M} \equiv \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}$$

とおくと $\boldsymbol{M}$ は $n \times n$ 行列の対称行列で

$$\boldsymbol{e} = \boldsymbol{My} \tag{12}$$

と表すことができる。一方， y の推定値を$\hat{\boldsymbol{y}}$で表すと

$$\hat{\boldsymbol{y}} = \boldsymbol{Xb} = \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$$

---

[3] 行列 X において j 番目の説明変数のベクトル x_j は第 j+1 列にある。第 1 列は定数項の 1 からなるベクトルである。

[4] この性質の導出は煩雑なので省略する。

である。ここで

$$\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

とおけば（P もまた $n \times n$ の対称行列である），

$$\hat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y} \tag{13}$$

が成立する。$\boldsymbol{M}$ と $\boldsymbol{P}$ の間には $\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{P}$ という関係が成り立ち，さらに，次のような性質が成立する。

$$\begin{gathered}
\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X} \\
\boldsymbol{M}\boldsymbol{X} = \boldsymbol{O} \\
\boldsymbol{P}^2 = \boldsymbol{P} \\
\boldsymbol{M}^2 = \boldsymbol{M} \\
\boldsymbol{M}\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{P} = \boldsymbol{P} - \boldsymbol{P}^2 = \boldsymbol{P} - \boldsymbol{P} = \boldsymbol{O}
\end{gathered}$$

ただし，$\boldsymbol{O}$ は全ての要素が 0 からなる $n \times n$ 行列である（ゼロ行列）。なお，$\boldsymbol{P}^2 = \boldsymbol{P}$ のような性質（行列の平方がもとの行列に等しい）が成り立つ行列を**べき等行列**(idempotent matrix)という。計量経済学の応用では，対称なべき等行列の固有値は 0 または 1 であるという性質が重要である。さらに，P は X の張る空間への射影行列という意味合いがある（M はその空間に直交する空間への射影である）。詳しくは，Greene 等の大学院レベルの教科書を参照のこと。

さて，(12)式に$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$を代入し，$\boldsymbol{M}\boldsymbol{X} = \boldsymbol{O}$を用いると

$$\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) = \boldsymbol{M}\boldsymbol{u}$$

が得られる。これから，

$$\begin{gathered}
\mathrm{E}(\boldsymbol{e}) = \boldsymbol{0} \\
\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{X}'\boldsymbol{M}\boldsymbol{u} = (\boldsymbol{M}\boldsymbol{X})'\boldsymbol{u} = \boldsymbol{0}
\end{gathered}$$

という性質が成り立つこともわかる（後者の式は残差と説明変数が直交するという意味）。

また，残差平方和は

$$\boldsymbol{e'e} = \boldsymbol{u'M'Mu} = \boldsymbol{u'Mu}$$

で与えられる。残差平方和の期待値を求めると

$$\mathrm{E}(\boldsymbol{e}'\boldsymbol{e}) = \mathrm{E}(\boldsymbol{u}'\boldsymbol{Mu}) = \mathrm{E}(\mathrm{tr}(\,\boldsymbol{u}'\boldsymbol{Mu})) = \mathrm{E}(\mathrm{tr}(\,\boldsymbol{Muu}'))$$

$$= \mathrm{tr}\left(\boldsymbol{M}\mathrm{E}(\boldsymbol{uu}')\right) = \mathrm{tr}(\boldsymbol{M}(\sigma^2\boldsymbol{I})) = \sigma^2\mathrm{tr}(\boldsymbol{M}) = \sigma^2(n-(k+1))$$

となる。tr( )はトレース（行例の対角成分の和）を表す。$\boldsymbol{M}$ のトレースが $n-(k+1)$になることについては Greene 等の教科書を参照せよ。なお，n はサンプル数，k+1 は定数項を含んだ説明変数の個数である。

以上から，誤差項の分散の不変推定量は次の式で与えられることがわかる（単回帰と同様の結果）。

$$s^2 = \frac{SSR}{n-(k+1)}$$

単回帰の場合と同様に，s^2 の平方根を，回帰の**標準誤差**(standard error of regression)と呼ぶ。

また，<u>残差平方和を真の分散で割った変数は，自由度 n-(k+1)のカイ二乗分布に従う</u>ことを示すことができる（n 個が独立に動けるわけではない）。

$$\frac{SSR}{\sigma^2} = \frac{\boldsymbol{e'e}}{\sigma^2} = \frac{\boldsymbol{u'Mu}}{\sigma^2} = \frac{(\boldsymbol{Mu})'(\boldsymbol{Mu})}{\sigma^2} \sim X^2(n-(k+1))$$

決定係数

単回帰と同様に次の式が成立する。

$$TSS = ESS + RSS$$

TSS は全平方和（平均値の回り），ESS は回帰式で説明できる平方和，RSS は残差平方和である。この関係は, $\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e} = \boldsymbol{Py} + \boldsymbol{My}$ で $\boldsymbol{M}$ と $\boldsymbol{P}$ が直交することから導くことができる。決定係数は次の式で定義される。

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

なお，R^2 は説明変数の個数が増加すれば単調に増加する(ESS が単調に増加するから)。そこで，説明変数の個数の増加をペナルティーとするように修正した決定係数を考えることができる。**自由度修正済み決定係数**(adjusted R^2)は次の式で定義される。

$$\bar{R}^2 = 1 - \frac{RSS/(n-(k+1))}{TSS/(n-1)}$$

重回帰分析では，通常は，この自由度修正済み決定係数を用いて，当てはまりの良さを評価する。

仮説検定

以下では，個々の係数に関する仮説検定と，複数の係数制約についての仮説検定を説明する。前者は t 検定を用い，後者は F 検定を用いる。なお，個々の係数についての検定では，片側検定と両側検定がある。

個々の係数に関する仮説検定（両側検定）

次の仮説（帰無仮説）を考える。

$$H_0: \quad \beta_j=\beta_j^0$$

対立仮説は　$H_1: \beta_j \neq \beta_j^0$ である。H_0 が真の時，最小二乗推定量 $\boldsymbol{b}$ は次のような分布をする。

$$\boldsymbol{b} \sim N(\boldsymbol{\beta}^0, \sigma^2(\boldsymbol{X'X})^{-1})$$

ここで，σ^2 は誤差項の分散で，未知のパラメータである。特に，$\boldsymbol{b}$ の要素 b_j の分布は次の通りであった。

$$b_j \sim N(\beta_j{}^0, \sigma^2 a_{jj})$$

ここで，a_{jj} は$(\boldsymbol{X'X})^{-1}$ の$(j+1,j+1)$要素であり（定数項があるため，x_j の分散に対応する成分は$(j+1,j+1)$要素になる），　$a_{jj}= 1 / S_{xx}{}^j$ である（$S_{xx}{}^j$ は説明変数 x_j を他の説明変数および定数項に回帰した残差の平方和：x_j から他の変数の影響を除去した x_j 固有の平方和)。これから，次の関係が導かれる。

$$\frac{b_j - {\beta_j}^0}{\sigma\sqrt{a_{jj}}} \sim N(0,1)$$

ただし，この式の分母には未知のパラメータである誤差項の標準偏差σが含まれているので，このままでは仮説検定に用いることができない。そこで，σの代わりにその最小二乗推定量を用いる。

誤差項の分散に関する最小二乗推定量を s^2 で表すと，s^2 は次の式で与えられた。

$$s^2 = \frac{\boldsymbol{e'e}}{n-(k+1)} = \frac{RSS}{n-(k+1)}$$

さらに，残差平方和を誤差項の真の分散で割った値は，自由度 n-(k+1)のカイ二乗分布にしたがう（ここの証明も行列の知識が必要になるので省略)。すなわち，次の関係が成り立つ。

$$\frac{\boldsymbol{e'e}}{\sigma^2} = \frac{RSS}{\sigma^2} = \frac{(n-(k+1))s^2}{\sigma^2} \sim X^2(n-(k+1))$$

さらに，最小二乗推定量 b_j の標準誤差は次の式で与えられる。

$$s.e.(b_j) = s\sqrt{a_{jj}} = s/\sqrt{{S_{xx}}^j}$$

さて，z が標準正規分布に従う確率変数で，x が自由度 m のカイ二乗分布に従う確率変数で，互いに独立であるとき， $z/\sqrt{x/m}$は自由度 m の t 分布に従う（講義資料の「確率・統計の基礎を参照せよ」。ここで，

$$z = \frac{b_j - \beta_j^0}{\sigma\sqrt{a_{jj}}}$$

$$\sqrt{\frac{x}{m}} = \sqrt{\frac{RSS/\sigma^2}{n-(k+1)}} = \sqrt{\frac{RSS/(n-(k+1))}{\sigma^2}} = \sqrt{\frac{s^2}{\sigma^2}} = \frac{s}{\sigma}$$

とすると（m=$n-$(k+1)としている），z と x は互いに独立であることを示すことができる（残差と b が直交することを用いる：証明は行列の知識が必要なので省略)。z と$\sqrt{x/m}$の比を計算すると

$$\frac{z}{\sqrt{x/m}}=\frac{(b_j-\beta_j{}^0)/\left(\sigma\sqrt{a_{jj}}\right)}{s/\sigma}=\frac{b_j-\beta_j{}^0}{s\sqrt{a_{jj}}}=\frac{b_j-\beta_j{}^0}{s.e.(b_j)}$$

となる。z と$\sqrt{x/m}$の比は，自由度 m ($=n-(k+1)$) の t 分布に従うので（講義資料「確率・統計の基礎」を参照)，結局

$$\frac{b_j-\beta_j{}^0}{s.e.(b_j)}\sim t(n-(k+1))$$

が成り立つ。これが，個々の係数に関する仮説検定の基礎になる。

帰無仮説が $H_0: \beta_j=\beta_j^0$ であり，対立仮説が $H_1: \beta_j\neq\beta_j^0$ であるとき，推定された b_j がβ_j^0 とあまり離れていなければ，仮説 H_0 を受け入れ，そうでなければ仮説 H_0 を棄却する。今，推定された係数が$\tilde{b}_j$という特定の値をとったとしよう。仮説 H_0 を棄却するのは，

$$\Pr\left(\left|b_j-\beta_j^0\right|>\left|\tilde{b}_j-\beta_j^0\right|\right)<\alpha$$

を満たす場合とする。通常は，通常は$\alpha=0.05$ として検定を行う。つまり，推計された推定値と仮説 H_0 で想定した係数の距離が，想定した確率分布から考えて，十分にありそうも無いとき仮説 H_0 を棄却するのである。

さて，

$$\Pr\left(\left|b_j-\beta_j^0\right|>\left|b_j^*-\beta_j^0\right|\right)=\alpha$$

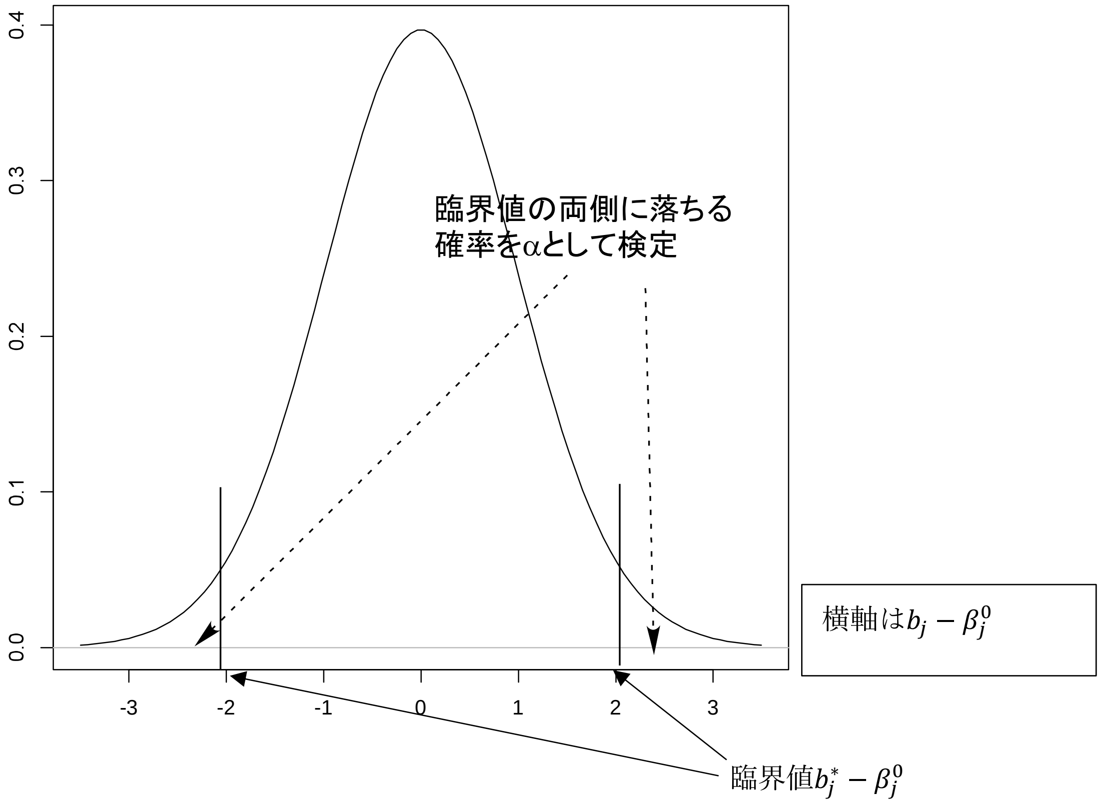
を満たすような b_j^*を**臨界値**とか**境界値**と呼ぶ。また，この式を満たすような b_j の範囲を**棄却域**とよぶ。仮説H_0 は推定された係数が棄却域に落ちるとき棄却するのである。なお，上の式に基づく検定は**両側検定**と呼ばれる。

さて，帰無仮説が H_0:　$\beta_j=\beta_j^0$ であり，対立仮説が H_1:　$\beta_j<\beta_j^0$ のような仮説検定も考えられる。この場合には，

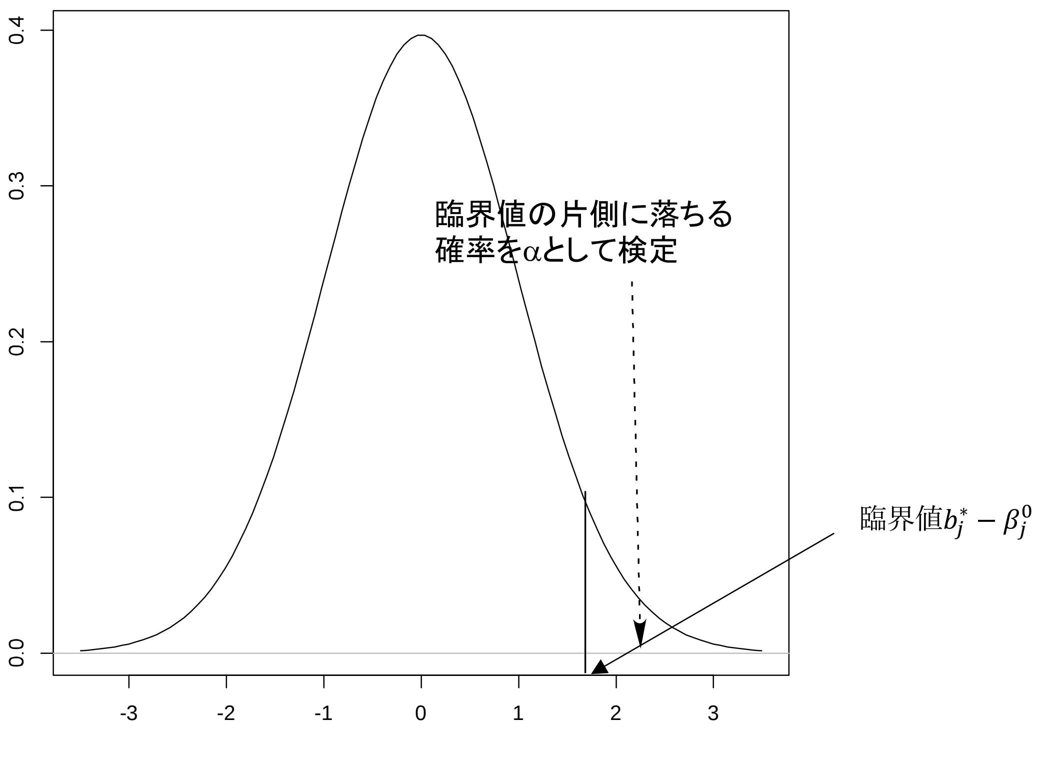
$$Pr\left(b_j-\beta_j^0>\tilde{b}_j-\beta_j^0\right)<\alpha$$

であれば，H_0 を棄却する。このような検定は**片側検定**と呼ばれる。　両側検定，片側検定，臨界値については次の図を参照のこと。

両側検定


臨界値の両側に落ちる
確率をαとして検定
横軸は$b_j - \beta_j^0$
臨界値$b_j^* - \beta_j^0$
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3

片側検定


臨界値の片側に落ちる
確率をαとして検定
臨界値$b_j^* - \beta_j^0$
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3

なお，統計ソフトで回帰分析を行うと，その出力には，回帰係数，標準誤差，t 値，p 値が出力される。この場合の t 値とは，仮説 H_0 において係数の真の値が 0 だとして計算された統計量である。つまり，$\beta_j^0 = 0$ の場合の統計量で，係数の推計値を b_j，その標準誤差を s.e.(b_j) で表せば，b_j/s.e.(b_j)が統計ソフトで出力される t 値である。また，p 値とは，係数/s.e.(b_j)の絶対値が b_j/s.e.(b_j)の絶対値よりも大きくなる確率である。この値がαより小さければ，係数の真の値は 0 であるという仮説は棄却されるのである（仮説 H0 で係数の真の値が 0 以外の値とするときには，自分で$(b_j - \beta_j{}^0)/(s.e.(b_j))$の値を計算しなければならない）。

複数の制約

帰無仮説　H_0: $\mathbf{R\beta}=\boldsymbol{q}$ を考える。対立仮説は H_1: $\mathbf{R\beta}\neq\boldsymbol{q}$ である。例えば，

1)　$\boldsymbol{R}$=[0 1 0 .....0], $\boldsymbol{q}$=[0 0 ... 0]’とすれば，

H_0：β_1=0

という仮説を考えていることになる（1 本の制約と同じである）。

2)　$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ の場合，

H_0：β_1=0,β_2=0

という 2 本の制約に帰着する。

さて，H_0 が真であるとき，

$$\boldsymbol{Rb} = \boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} = \boldsymbol{R\beta} + \boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'u} = \boldsymbol{q} + \boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'u}$$

が成立するので，$\boldsymbol{Rb}$ は次の正規分布に従うことがわかる。

$$\boldsymbol{Rb} \sim N(\boldsymbol{q}, \sigma^2\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'})$$

このとき

$$W = (\boldsymbol{Rb} - \boldsymbol{q})'[\sigma^2\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{Rb} - \boldsymbol{q}) \sim X^2(r)$$

が成り立つ。ここで，r は制約の数（R の階数（ランク））を表す。また，

$$V=\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}=\frac{SSR}{\sigma^2}\sim X^2(n-(k+1))$$

が成り立つ。さらに，W と V は互いに独立であることを示すことができる（残差と b の直交性より）。したがって，

$$\frac{W/r}{V/\big(n-(k+1)\big)}=\frac{\sigma^{-2}\,(\boldsymbol{Rb}-\boldsymbol{q})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb}-\boldsymbol{q})}{r}\cdot\frac{1}{\sigma^{-2}\,SSR/\big(n-(k+1)\big)}$$
$$=\frac{(\boldsymbol{Rb}-\boldsymbol{q})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb}-\boldsymbol{q})/r}{s^2}\sim F\big(r,n-(k+1)\big) \tag{14}$$

が成立する。仮説 H_0 が正しければ，(14)式の値は小さくなり，仮説 H_0 が正しくなければ，(14)式の値は大きくなる[5]。

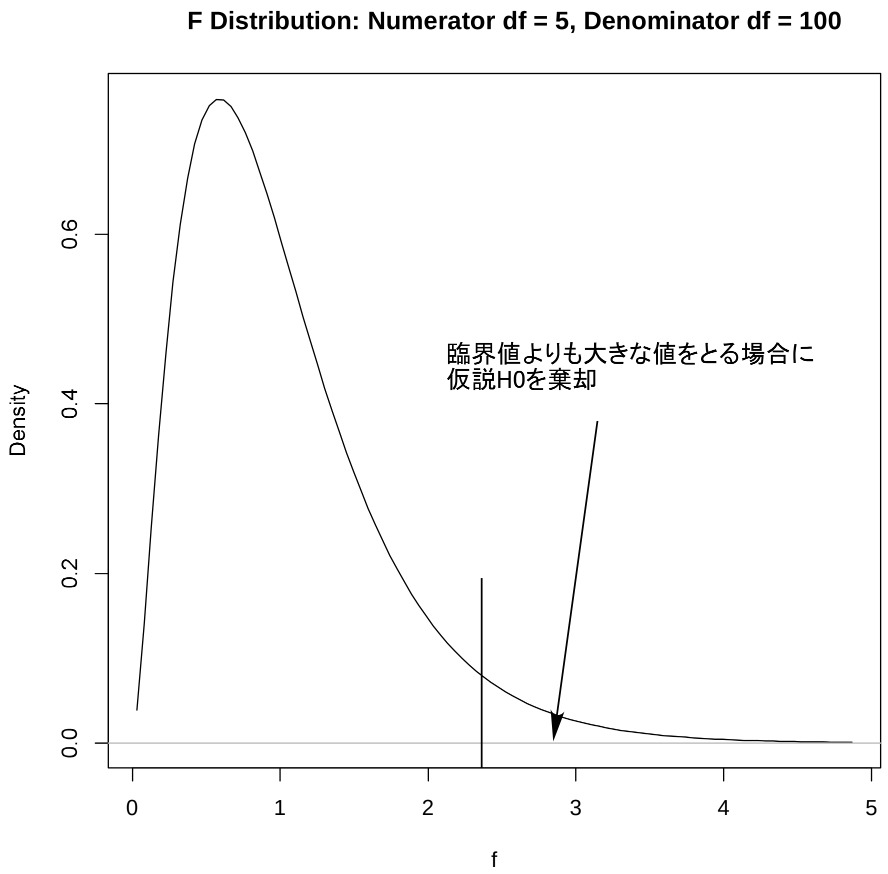
(14)式は自由度(r, $n-(k+1)$)の F 分布に従うが，F 分布は次の図のような形状をしている。図のように，(14)式から計算される統計量が棄却域に入れば，H_0 を棄却すればよい（そうでなければ H_0 を受け入れる）。

(14)式は，例えば，H_0：$\beta_1=0, \beta_2=0$ の場合，すなわち

$$\boldsymbol{R}=\begin{bmatrix}0&1&0&0&\cdots&0\\0&0&1&0&\cdots&0\end{bmatrix},\quad \boldsymbol{q}=\begin{bmatrix}0\\0\end{bmatrix}$$

で，説明変数 x_1, x_2 が無相関の場合を考えるとわかりやすいかもしれない。結論だけを述べると，点(b_1/s.e(b_1), b_2/(s.e.(b_2)) の q'=(0，0)からの距離の平方を表すのである。したがって，(14)式の F 統計量が（棄却域にはいるほど）十分に大きければ仮説 H0 は棄却されるのである。

[5] (14)式は制約が$\beta_1=0, \beta_2=0$ で，説明変数 x_1 と x_2 の相関が 0 の場合を考えるとわかりやすくなる。この場合，F 統計量は推定された(b_1,b_2)の原点(0,0)からの距離の平方であることを示すことができる。



なお，(14)式の W/r と $V/(n-(k+1))$の比を計算して求めるのではなく，次の方法でのテストが同等であることが知られている。まず，係数に何の制約も課さない回帰分析を行い，そのときの残差平方和を URSS (Unrestricted Residual Sum of Squares: 制約無しの残差平方和)とする。次に, H0 の制約を課した上で回帰分析を行い, そのときの残差平方和を RRSS (Restricted Residual Sum of Squares 制約付きの残差平方和)とする。そして，RRSS−URRS を計算し（つまり制約を課すことでどのくらい当てはまりが悪くなるかを計算し），それを制約の数で割った値を分子にする。また，分母は，制約無しの残差平方和を自由度で割った値とする（制約無しの回帰の標準誤差の平方に等しい）。このとき，

$$\frac{(RRSS - URRS)/r}{URRS/(n-(k+1))} \sim F(r, n-(k+1))$$

が成立する。ここでも，棄却域をαとおいて，H_0 が真のとき，上の式の左辺がありそうも無い値をとる場合には H_0 を棄却する。

また，特に，

H_0 : $\beta_1=0, \beta_2=0, ..., \beta_k=0$

という仮説（全ての説明変数に説明力が無い；ただし，定数項を除く）の検定を考えよう。今，制約無しの回帰の全平方和と TSS，回帰変数で説明される部分の平方和を ESS，残差平方和を RSS で表すと，制約付きの残差平方和 RRSS は TSS に等しいので，RRSS−URRS=TSS−RSS=ESS が成り立つ。したがって，（定数項を除く）全ての説明変数の係数が 0 であるという仮説を検定するためには

$$\frac{ESS/k}{RSS/(n-(k+1))} \sim F(k, n-(k+1))$$

であることを用いればよい（通常の統計パッケージでは，回帰分析の標準的な出力にこの F 値が報告される）。