回帰分析（単回帰）

麻生良文

1.回帰分析の前提

回帰分析は，説明変数 x と被説明変数 y の間の関係を統計的にとらえる手法です。説明変数が単一の変数である場合を単回帰(simple regression)，説明変数が複数個ある場合を重回帰(multiple regression)といいます。最初に単回帰モデルを説明します。

まず，次のようなモデルを想定します。

$$y = \alpha + \beta x + u \tag{1}$$

ここで，α，βはある定数で，u は確率変数です。つまり，x が与えられると確定的な項$\alpha + \beta x$にランダムなショック u が加わって y が決まるというモデルを想定します。なお，y，x，u は次のように呼ばれます。

- y：被説明変数(explained variable)，従属変数(dependent variable), regressand
- x：説明変数(explanatory variable)，独立変数(independent variable), regressor
- u: 誤差項(error term)，撹乱項(disturbance term)

例）

(Keynes 型消費関数)　　$C = \alpha + \beta\, YD + u$

C：民間消費，YD:可処分所得，β：限界消費性向

(賃金の決定)　　$\mathrm{wage} = \alpha + \beta\,\mathrm{educ} + u$

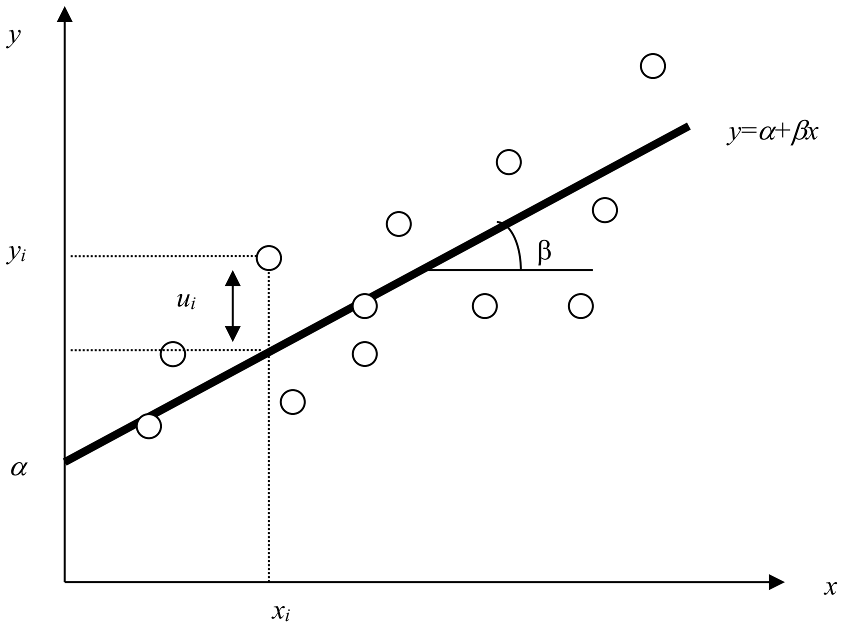
wage:賃金，educ:教育年数

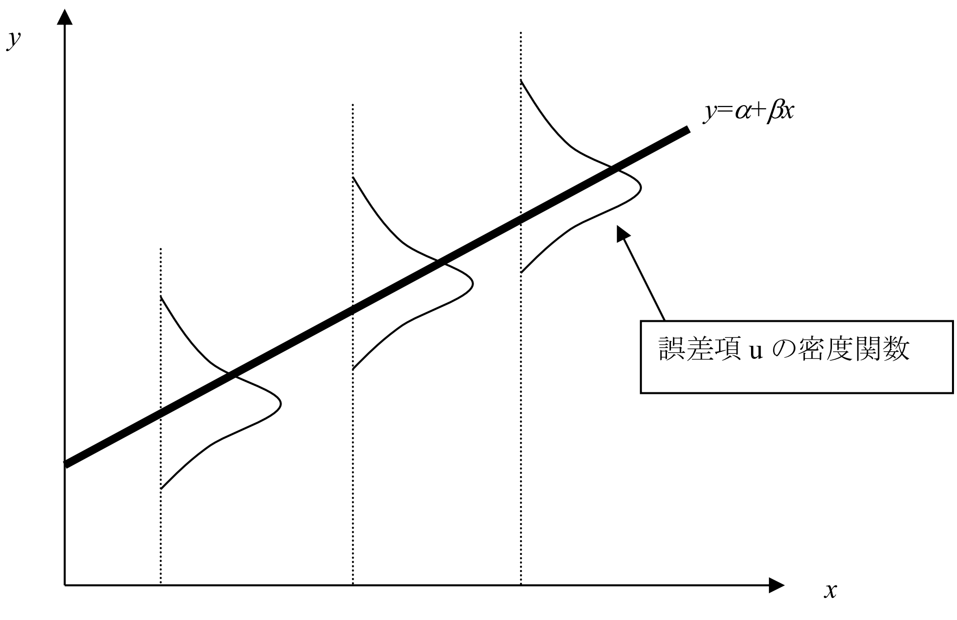
モデルの特徴

- 線型モデルである。
- 線型モデルは制約的に感じられるかもしれないが，x^2 を説明変数に含める，あるいは，ln(x)などの変数変換を行うことで，　x の非線形効果を捉えることは可能である。
- x 以外の効果は誤差項に集約されている。それらは次のようなものである。
  - ✓ 他の変数，モデルで想定していない変数の効果
  - ✓ 観察不可能な変数の影響

- ✓ y の測定誤差

(x,y)の組み合わせが(1)式のような確率モデルにしたがって実現すると考えると，直線$y = \alpha + \beta x$と観測された点(x,y)は次の図のような関係であると考えられます（上の図の○が観測された点を表し，下の図は誤差項の確率分布をイメージした図である）。





なお，重回帰(multiple regression)モデルは，複数の要因を同時に考慮するモデルで，次の式のようなモデルを想定する。

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{2}$$

回帰分析の前提

観測された(x,y)をもとにα,βの値を推定するために，誤差項 u に次の仮定をおく。

（仮定 1） 線型性

真のモデルが次の方程式で表される。α，βは（推定すべき未知の）パラメータで，(x_i, y_i)は i 番目の観測値，u_i は誤差の実現値である。

$$y_i = \alpha + \beta x_i + u_i \tag{A1}$$

（仮定 2） 誤差項の期待値は 0（すべての i について）。

$$\mathrm{E}\,(u_i) = 0 \tag{A2}$$

なお，（A1）式において定数項が存在するため，誤差項の期待値が 0 であるという仮定は何ら制約的でないことに注意せよ（定数項で調整できる）。

（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）。

$$\mathrm{var}\,(u_i) = \sigma^2 \tag{A3}$$

（仮定 4） 誤差項に系列相関は存在しない。すなわち，全ての $i \neq j$ に対し u_i と u_j の共分散は 0 である。

$$\mathrm{cov}\left(u_i, u_j\right) = 0 \quad \text{for all } i \neq j \tag{A4}$$

（仮定 5) 説明変数と誤差項の独立性。

説明変数 x と誤差項 u は独立である。古典的回帰モデルにおいては，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。すなわち，

$$u_i \sim N(0,\sigma^2) \qquad \textit{i.i.d.}$$

が成り立つ（*i.i.d.*は independently identically distributed の略）。

回帰分析は（A1）のようなモデルを仮定して，観測されたデータから，パラメータα,βを求める統計的手法である。観測されたデータは，(x,y)のとりうる値の一部でしかない（標本である）。また,誤差項を含んだ確率モデルを想定しているので，観測されたデータから何らかの方法で推定された a,b が真の値である保証はない。一般には，α,βの推定値は誤差項の確率分布の性質に依存して，ある確率分布に従うと考えられる。

POINT

- どのような方法でパラメータを推定することが望ましいのか。
- 推定されたパラメータがどのような確率分布に従うか。

以下では，最小二乗法による推計を説明する（他にも，モーメント法，最尤法などの推定方法がある）。最小二乗法に基づくパラメータの推計量は，BLUE（Best Linear Unbiased Estimator）という望ましい性質を備えている（説明は省略）。BLUE の性質は上の仮定 1 から仮定 5 までが満たされる場合に成立する。仮定 6 の正規分布の仮定は，この性質を持つためには不要な（強すぎる）仮定である。ただし，推定されたパラメータについての仮説検定のためには，誤差項の確率分布を特定化しなければならず，そのためには通常，誤差項の正規分布を仮定する。この仮定から，最小二乗法で得られたパラメータの推計量の確率分布が求められる。

2. 最小二乗法(method of least square)

パラメータα,βの推定値を a,b で表す。第 i 番目の観測値が(x_i,y_i)であるとき，$\hat{y}_i = a + bx_i$をy_iの推定値，fitted value ，predicted value などと呼ぶ。そして，実際の観測値と推定値の差を残差（residual）と呼ぶ。残差を e_i は次の式で表される。

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i) \tag{3}$$

最小二乗法は残差の平方和を最小にするように，a,b を決定するという手法である。次の式からわかるように，残差平方和は a,b の関数とみなすことができ，それを $S(a,b)$で表すと

$$S(a,b) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{4}$$

となる。$S(a,b)$は(a,b)に関する 2 次関数である。$S(a,b)$を最小化する(a,b)を見つけるためには，(a,b)に関して微分して最小化の必要条件を求めればよい（微分して 0 になることが必要条件だが，(4)式の a^2 や b^2 の項の係数はプラスなので，最小化の十分条件でもある）。

まず，a に関して微分すると

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

を得る。これから次の式が得られる。

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a + b \sum_{i=1}^{n} x_i$$

さらに，この式の両辺を n で割れば

$$\bar{y} = a + b\bar{x} \tag{5}$$

が得られる。ここで，$\bar{x} = \sum_i x_i / n\,, \bar{y} = \sum_i y_i / n$ で，それぞれ x,および y の平均値を表す。

また，b に関して微分すると

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (y_i - a - bx_i)\, x_i = 0$$

を得る。この式を変形すると次の式が求められる。

$$\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} {x_i}^2 \tag{6}$$

(5)式と(6)式は(a,b)に関する連立一次方程式である。この方程式を正規方程式(normal equation)と呼ぶ。正規方程式の解が，最小二乗法による(α,β)の推計値になる。

最小二乗法の推計値を求めるため，(5)式を a について解き，これを(6)式に代入すると

$$\sum_i x_i y_i = (\bar{y} - b\bar{x}) \sum_i x_i + b \sum_i {x_i}^2 = n\bar{x}\bar{y} + b\left(\sum_i {x_i}^2 - n\bar{x}^2\right)$$

を得る。この式から次の式が得られる。

$$S_{xy} = bS_{xx} \tag{7}$$

ただし，S_{xx} は x の平均値の回りの平方和，S_{xy} は x と y の平均値の回りのクロス・モーメントで次の式で与えられる。

$$\mathrm{S}_{xx} = \sum_{\mathrm{i}} (x_i - \bar{x})^2 = \sum_i x_i^2 - n\,\bar{x}^2$$
$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - n\bar{x}\bar{y}$$

さて，(7)式より

$$b = \frac{\mathrm{S}_{xy}}{S_{xx}} \tag{8}$$

が求められ，これがパラメータβの最小二乗法による推定値である。(8)式により b が求まると，αの推定値は(5)式より， $a = \bar{y} - b\bar{x}$で求められる。

なお, 残差（観測された y_i の値と y_i の予測値の差）は $e_i = y_i - (a + bx_i)$で与えられるが，正規方程式の性質から

$$\sum_i e_i = 0 \tag{9}$$

$$\sum_i x_i e_i = 0 \tag{10}$$

が成立することがわかる（残差平方和の一階の条件をみよ）。(9)式および(10)式は，残差の平均が 0（あるいは残差と定数項が無相関），残差と説明変数 x が無相関であることを表している。

当てはまりの良さ

$y_i = a + bx_i + e_i$から $\bar{y} = a + b\bar{x}$を引くと

$$y_i - \bar{y} = b(x_i - \bar{x}) + e_i$$

という関係が得られる。この式の両辺を平方し，i について合計すると，次の式が得られる。

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}{e_i}^2 \tag{11}$$

なお，(11)式の導出において x_i と e_i の交差項が消えるのは，(10)式が成り立つからである。

(11)式の左辺は y の平均値の回りの平方和（全体の平方和），右辺の第 1 項は説明変数 x で説明される部分の平方和，右辺第 2 項は，残差の平方和を表す。全体の平方和を TSS(Total Sum of Squares)，説明変数で説明される部分の平方和を ESS(Explained Sum of Squares)，残差平方和を RSS(Residual Sum of Squares)で表すと，(11)式は

$$\mathrm{TSS} = \mathrm{ESS} + \mathrm{RSS} \tag{12}$$

と書き直すことができる。なお， TSS は y の平均値の回りの平方和なので，(8)式と同様の記号を用いると，S_{yy} と表すことができる。

ESS を TSS で割った値は，全平方和のうち説明変数で説明される平方和の比率を表す。これを**決定係数**(coefficient of determination)といい，通常 R^2 で表す。R^2 は次の式で定義される。

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} \tag{13}$$

R^2 は 0 から 1 の間の値をとる。R^2 が 1 に近いほど，モデルの説明力が高いことになる。

なお，決定係数は x と y の相関係数を平方したものに等しい。これは次のことから確かめられる。まず，(8)式と(11)式から， ESS は次のように変形できる。

$$\mathrm{ESS} = b^2 S_{xx} = \left(\frac{S_{xy}}{S_{xx}}\right)^2 S_{xx} = \frac{{S_{xy}}^2}{S_{xx}}$$

したがって，次の式が成り立つ。

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = \frac{{S_{xy}}^2/S_{xx}}{S_{yy}} = \frac{{S_{xy}}^2}{S_{xx}S_{yy}}$$

この式は，R^2 が x と y の相関係数の平方であることを示している。

最小二乗推定量(least square estimator)の確率的性質

(α,β)の推定値(a,b)は y_i の実現値に依存して決定される（つまり，誤差項の実現値 u_i に依存して決定される）。事前には(a,b)はどの値が実現するかは確定せず，ある確率分布にしたがって実現する。(a,b)が確率変数であることを強調する場合，(a,b)を推定量(estimator)と呼ぶ。

最小二乗推定量の確率分布を求めてみよう。まず，(8)式から

$$b = \frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{S_{xx}}\sum_i (x_i - \bar{x})y_i = \sum_i c_i y_i \tag{14}$$

が成立する。ただし，

$$c_i \equiv \frac{x_i - \bar{x}}{S_{xx}}$$

である。つまり，最小二乗推定量 b は y_i の線型関数である。ここで，回帰モデルの仮定 6 までの全ての仮定が満たされるとすると，b は正規分布に従うことがわかる。一方，$a = \bar{y} - b\bar{x}$より，

$$a = \frac{1}{n}\sum_i y_i - b\bar{x} = \frac{1}{n}\sum_i y_i - \frac{\bar{x}}{S_{xx}}\sum_i (x_i - \bar{x})y_i = \sum_i d_i y_i \tag{15}$$

となる。ただし，

$$d_i \equiv \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}}$$

である。これから，a も y_i の線型関数であり，正規分布に従うことがわかる。

(14)式，(15)式から，(a,b)の期待値，分散が求められる。導出はやや面倒であるので，最初に結果だけを述べておこう。まず，期待値は次の通りになる。

$$\mathrm{E}(a) = \alpha$$
$$\mathrm{E}(b) = \beta$$

これらの式は最小二乗推定量の期待値はパラメータの真の値に一致することを述べている。

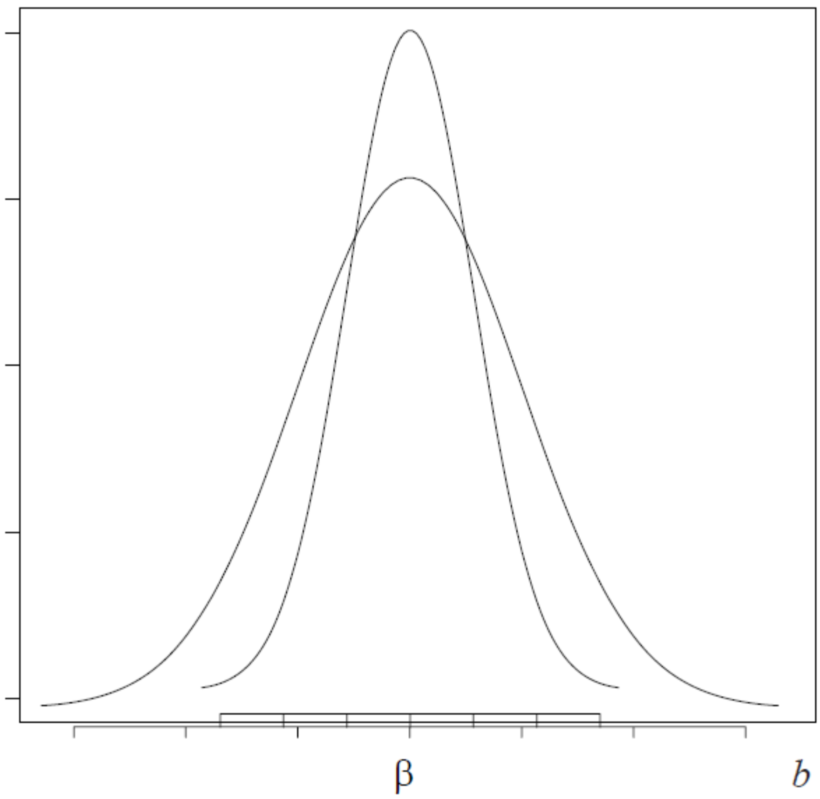
この性質を不偏性(unbiasedness)という。

　また，分散，共分散は次の通りになる。

$$\mathrm{var}(b) = \frac{\sigma^2}{S_{xx}}$$
$$\mathrm{var}(a) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$
$$\mathrm{cov}(a, b) = \sigma^2\left(-\frac{\bar{x}}{S_{xx}}\right)$$

　上の式によれば， S_{xx} が大きいほど（ x のバラつきが大きいほど)，サンプル数 n が大きいほど，パラメータの a,b の分散は小さくなる。逆に言えば，x のバラつきが少なかったり，サンプル数が十分でない場合には，パラメータの推定は不正確なものになる。

図　最小二乗推定量 b の分布



S_{xx} が大きいほど，n が大きいほど，var(b)が小さくなる。

最小二乗推定量の確率分布の導出

（導出はやや面倒）

　まず，真のモデルが

$$y_i = \alpha + \beta x_i + u_i$$

で与えられ，誤差項 u_i は互いに独立で同一な正規分布 $N(0,\sigma^2)$に従うものとする。最小二乗推定量の期待値と分散・共分散を求めるためには，(14)式，(15)式の c_i，d_i の性質について最初に導いておくと都合がよい。まず，c_i については

$$\sum_i c_i = \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) = 0$$

$$\sum_i c_i x_i = \frac{1}{S_{xx}} \sum_i (x_i - \bar{x})\, x_i = \frac{1}{S_{xx}} \sum_i (x_i - \bar{x})^2 = 1$$

$$\sum_i {c_i}^2 = \frac{1}{{S_{xx}}^2} \sum_i (x_i - \bar{x})^2 = \frac{1}{S_{xx}}$$

が成立する。また，d_i については

$$\sum_i d_i = \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}} (x_i - \bar{x}) \right) = 1$$

$$\sum_i d_i x_i = \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}} (x_i - \bar{x}) \right) x_i = \bar{x} - \frac{\bar{x}}{S_{xx}} \sum_i (x_i - \bar{x})^2 = \bar{x} - \bar{x} = 0$$

$$\sum_i {d_i}^2 = \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}} (x_i - \bar{x}) \right)^2 = \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} - \frac{2\bar{x}}{nS_{xx}} \sum_i (x_i - \bar{x}) = \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}$$

が成立する。さらに，

$$\sum_i c_i d_i = \sum_i \left(\frac{x_i - \bar{x}}{S_{xx}} \right) \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}} (x_i - \bar{x}) \right) = -\frac{\bar{x}}{{S_{xx}}^2} \sum_i (x_i - \bar{x})^2 = -\frac{\bar{x}}{S_{xx}}$$

が成立する。

　b の確率分布を求めるため，(14)式に $y_i = \alpha + \beta x_i + u_i$ を代入すると，

$$b = \sum_i c_i (\alpha + \beta x_i + u_i) = \alpha \sum_i c_i + \beta \sum_i c_i x_i + \sum_i c_i u_i = \beta + \sum_i c_i u_i$$

を得る（c_iの性質を参照せよ）。したがって，

$$\mathrm{E}(b) = \beta + \mathrm{E}\left(\sum_i c_i u_i\right) = \beta + \sum_i c_i \mathrm{E}(u_i) = \beta$$

となり，b の期待値が導かれた（2 番目の等式が成立するのは期待値オペレータの線型性の性質より，また 3 番目の等式は誤差項に関する期待値が 0 であるから）。分散を求めると次の通りになる。

$$\mathrm{var}(b) = \mathrm{E}(b-\beta)^2 = \mathrm{E}\left(\sum_i c_i u_i\right)^2 = \sigma^2 \sum_i {c_i}^2 = \frac{\sigma^2}{S_{xx}}$$

3 番目の等号は，誤差項が互いに独立で同一の分布に従うという仮定から導かれる。すなわち，$\mathrm{E}\,({u_i}^2) = \sigma^2$と $\mathrm{E}(u_i u_j) = 0$ が成り立つからである（$i \neq j$）。

a の期待値と分散も同様にして求められる。(15)式に $y_i = \alpha + \beta x_i + u_i$ を代入すると，

$$a = \sum_i d_i(\alpha + \beta x_i + u_i) = \alpha \sum_i d_i + \beta \sum_i d_i x_i + \sum_i d_i u_i = \alpha + \sum_i d_i u_i$$

が成り立つことがわかる（d_iの性質を参照せよ）。したがって，

$$\mathrm{E}(a) = \alpha + \mathrm{E}\left(\sum_i d_i u_i\right) = \alpha + \sum_i d_i \mathrm{E}(u_i) = \alpha$$

$$\mathrm{var}(a) = \mathrm{E}(a-\alpha)^2 = \mathrm{E}(\textstyle\sum_i d_i u_i)^2 = \sigma^2 \sum_i {d_i}^2 = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

が導かれる。さらに

$$\mathrm{cov}(a,b) = \mathrm{E}[(a-\alpha)(b-\beta)] = \mathrm{E}\left[\left(\sum_i d_i u_i\right)\left(\sum_i c_i u_i\right)\right] = \sigma^2 \sum_i c_i d_i = \sigma^2\left(-\frac{\bar{x}}{S_{xx}}\right)$$

が導かれる。なお，以上から最小二乗推定量の分散は誤差項の分散σ^2 に依存することがわかる。ただし，σ^2 は未知のパラメータであることに注意。

誤差項の分散σ^{2}の推定値

サンプル数 n，説明変数 2 個の場合の（単回帰の場合，説明変数 x と定数項で合計 2 個の説明変数があると考える）， σ^2 の最小二乗推定量 s^2 は

$$\boldsymbol{s^2 = \frac{1}{n-2}RSS = \frac{1}{n-2}\sum_i e_i{}^2} \tag{16}$$

で与えられる。n−2 は**自由度**(degrees of freedom)と呼ばれる。なお，s^2 の平方根は**標準誤差**(standard error)と呼ばれる。多くのソフトでは **SER**（回帰の標準誤差:standard error of the regression）として出力される。

残差平方和を n ではなく，(n−2)で割るのは，そうすることで，s^2 が不偏性を持つからである。ここで，

$$\boldsymbol{\frac{RSS}{\sigma^2} = \sum_i \left(\frac{e_i}{\sigma}\right)^2} \tag{17}$$

は**自由度(n−2)のカイ二乗分布**にしたがう（この証明には行列の知識が必要なので省略)。自由度が n−2 であるのは，n 個のサンプルのうち独立に動ける次元が 2 つ（定数項と x）だけ少なくなるからである。行列を用いた証明は，Greene の Econometric Analysis などを参照のこと。

自由度 m のカイ二乗分布に従う確率変数の期待値は m に等しい。これから，RSS/σ^2 の期待値は(n−2)に等しく，したがって，(16)式で定義される誤差項の分散の推定量の期待値は真の分散σ^2 に等しいことがわかる。

βに関する検定

次のような仮説 H_0 を考える。

$$H_0: \quad \beta=\beta_0$$

多くの仮説検定ではβ$_0$=0 を H_0 とすることが多い。例えば, 賃金と教育年数の回帰分析で, 教育年数が賃金に影響しないという仮説は，教育年数の係数が 0 に等しいという仮説検定の問題として考えることができる（もちろん，β$_0$ が 0 という特定の値である必要は無い)。

βの最小二乗推定量 b は，さきほど求めたように，期待値β，分散σ^2/S_{xx} の正規分布に従った。したがって，H_0 が正しいなら

$$\boldsymbol{\frac{b-\beta_0}{\sqrt{\sigma^2/S_{xx}}} \sim N(0,1)} \tag{18}$$

が成り立つ。ただし，σ^2 は未知のパラメータなので，このままでは仮説検定に使えない。そこで，σ^2 を先ほど求めた最小二乗推定量 s^2 で置き換える統計量を考えてみよう。すなわち，

$$\frac{\boldsymbol{b}-\boldsymbol{\beta_0}}{\sqrt{\boldsymbol{s^2}/\boldsymbol{S_{xx}}}}=\frac{\boldsymbol{b}-\boldsymbol{\beta_0}}{\mathrm{s.e.}(\boldsymbol{b})} \tag{19}$$

を考える。(19)式の分母の$\sqrt{s^2/S_{xx}}$は，推定量 b の標準偏差（分散の平方根）の推計値で，b の**標準誤差**（standard error）と呼ばれる（b の真の標準偏差は$\sqrt{\sigma^2/S_{xx}}$であり，未知のパラメータσ^2 に依存していた）。仮説 H_0 が真であれば，(19)式は自由度 (n-2)の t 分布に従う。

----

t 分布の復習

z を標準正規分布，x を自由度 m のカイ二乗分布に従う確率変数とし， z と x は独立であるとしたとき，次の変数は自由度 m の t 分布に従う。

$$\frac{z}{\sqrt{x/m}} \sim t(m)$$

----

(19)式が自由度(n−2)の t 分布に従うことを説明する。まず，(18)式のように，$\frac{b-\beta_0}{\sqrt{\sigma^2/S_{xx}}}$ は標準正規分布に従うので，これを z とする。次に，(17)式の RSS/σ^2 は自由度(n−2)のカイ二乗分布に従うので，これを x とする（$m=n-2$ である）。さらに，この z と x の分布は独立であることを示すことができる（この証明には行列の知識が必要。Greene 等の教科書を参照のこと）。したがって， z と$\sqrt{x/m}$の比を求めると

$$\frac{z}{\sqrt{x/m}}=\frac{\dfrac{b-\beta_0}{\sqrt{\sigma^2/S_{xx}}}}{\sqrt{\dfrac{RSS/\sigma^2}{n-2}}}=\frac{b-\beta_0}{\sqrt{\sigma^2/S_{xx}}}\cdot\frac{\sqrt{\sigma^2}}{\sqrt{RSS/(n-2)}}=\frac{b-\beta_0}{\sqrt{1/S_{xx}}}\cdot\frac{1}{\sqrt{RSS/(n-2)}}=\frac{b-\beta_0}{\sqrt{s^2/S_{xx}}}$$

となり，(19)式に等しいことが示された。したがって(19)式は自由度(n−2)の t 分布に従うことがわかる。つまり，次の式が成立する。

$$\frac{\boldsymbol{b}-\boldsymbol{\beta_0}}{\sqrt{\boldsymbol{s^2}/\boldsymbol{S_{xx}}}}=\frac{\boldsymbol{b}-\boldsymbol{\beta_0}}{\mathrm{s.e.}(\boldsymbol{b})}\sim\boldsymbol{t}(\boldsymbol{n}-\boldsymbol{2}) \tag{20}$$

BLUE

最小二乗推定量は線型推定量(linear estimator)の一種である（(14)式，(15)式で表されるように確率変数 y の線型関数である）。また，最小二乗推定量の期待値は真のパラメータα,βに等しかった。つまり，最小二乗推定量は不偏推定量(unbiased estimator)のクラスに属する。さらに，最小二乗法は，線型不偏推定量(linear unbiased estimator)のクラスに属する推定量の中で，最小の分散を持つ推定量であることが知られている（証明は省略）。この性質をBLUE(best linear unbiased estimator: 最良線型不偏推定量)という。最小二乗法は，このような意味で望ましい推定量である。なお，BLUE という性質は，誤差項が正規分布に従うという仮定なしに導かれる。

その他の推定方法

- モーメント法

$$\sum_i e_i = \sum_i (y_i - a - bx_i) = 0$$
$$\sum_i x_i e_i = \sum_i (y_i - a - bx_i) x_i = 0$$

を満たすようにパラメータ(a,b)を決める（E(u)=0, cov(x,u)=0 に対応）。今考えているケース（(1)式で表されるような線形モデル）では，最小二乗法と同じになる。

- 最尤法(Maximum Likelihood Method)

まず，真のモデルが

$$y_i = \alpha + \beta x_i + u_i$$

で与えられ，誤差項 u_i は互いに独立で同一の正規分布 N(0,σ^2)に従うものとする。さて，今，(x_i,y_i) (i=1,2,..,n)が観測されたとする。この時，x_i が与えられたもとで y_i が実現する確率密度関数（条件付確率密度関数）は次の式で与えられる。

$$f(y_1, y_2, \cdots, y_n | \boldsymbol{x}; \alpha, \beta, \sigma^2) = \prod_{i=1}^{n} \phi(u_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{u_i}{\sigma}\right)^2\right]$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right]$$

上の式で，$\phi(u_i) = \left(\sqrt{2\pi\sigma^2}\right)^{-1} \exp[-(u_i/\sigma)^2/2]$は，期待値 0，分散$\sigma^2$の正規分布の確率密度関数である。また，$\boldsymbol{x}=(x_1,x_2,...,x_n)$である。

上の式を未知パラメータα,β,σ^2の関数とみなしたものは尤度関数(likelihood function)と呼ばれる。未知パラメータα,β,σ^2をa,b,s^2に置き換えて，尤度関数を$L(a,b,s^2)$で表せば，

$$L(a,b,s^2|\boldsymbol{x},\boldsymbol{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - a - bx_i)^2\right]$$

となる。この尤度関数を最大にするように(a,b,s^2)を決める方法が最尤法(maximum likelihood method)である。簡単に言えば，観測された$(x_i,y_i)(i=1,2,...,n)$が実現するのは，確率分布を規定するパラメータがどのような値だともっともらしいかに基づいてパラメータを決定するのが最尤法である。

さて，上の尤度関数の対数をとると

$$\ln L\,(a,b,s^2|\boldsymbol{x},\boldsymbol{y}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln s^2 - \frac{1}{2s^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

と変形できる。これを対数尤度関数(log likelihood function)と呼ぶ。尤度の最大化と対数尤度の最大化の条件は同じなので，通常はこの対数尤度を最大にするようにパラメータを決める。今，考えている線型モデルの場合には，(a,b)の決定は，

$$\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

の最小化条件を求めればよい。これは残差平方和の最小化条件を求めることだから，(a,b)の最尤法による推計は最小二乗法と全く同じになる。s^2については対数尤度関数をs^2で微分して 0 とおくと

$$-\frac{n}{2}\cdot\frac{1}{s^2} + \frac{1}{2}\cdot\frac{1}{(s^2)^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = -\frac{n}{2}\cdot\frac{1}{s^2} + \frac{1}{2}\cdot\left(\frac{1}{s^2}\right)^2 RSS = 0$$

を得る。ただし$RSS = \sum_i {e_i}^2$で残差平方和を表す。上の式を s^2 について解くと

$$s^2 = \frac{RSS}{n}$$

が求められる。

σ^2 についての最小二乗推定量は RSS/(n−2)であったが，最尤推定量は RSS/n である。このため，σ^2 の最尤推定量は不偏性を持たない。普遍性は持たないが，一致性(consistency)という性質は持っている（サンプル数の増加が，推定量の分布を真のパラメータの値に収束させる性質)。なお，サンプル数が大きい場合，n−2 で割ろうが，n で割ろうが，その差はほとんどなくなることにも注意。

誤差項に関する確率分布を特定化すれば，最尤法は，線型モデルだけではなく，非線形モデルにも応用できる。このため，最尤法は広く用いられている。