

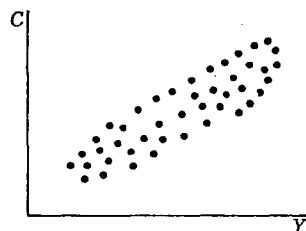
## 第1章 回 帰 分 析

### 1 回帰分析の概要

まず、回帰分析の「回帰」の意味を説明する。統計学の用語として「回帰」という言葉をはじめて用いたのは、生物統計学者ゴールトン (Francis Galton, 1822~1911) である。ゴールトンは、父親の身長と息子の身長データを多数集め、「背の高い父親を持つ息子の平均身長は父親ほどには高くない。逆に背の低い父親を持つ息子の平均身長は父親ほどには低くない。」すなわち、第2世代は平均の方向に「回帰」する事実を発見した。これ以来、「回帰」という語が統計学用語に加えられた訳である。

一般の経済分析に用いられている「回帰」の考え方についてみてみよう。今、所得  $Y$  と消費  $C$  の関係が次の図のようなかたちになっていたとする。

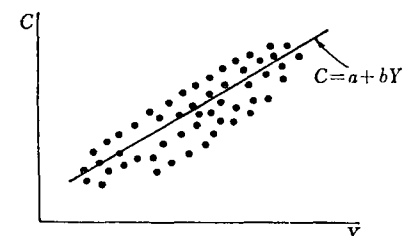
(図表 2.1.1)



問題はこのようなデータが与えられた場合、消費と所得の関係をいかにして簡潔かつ効率的に表現するかである。こうした観点から  $C$  と  $Y$  との関係、たとえば直線で代表させようとするのが回帰分析である。

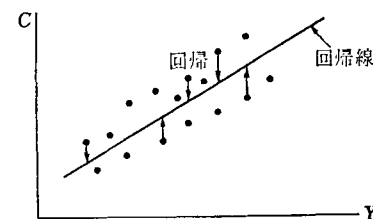
これが「回帰」分析と呼ばれるのは、たとえば  $C$  の動きをどの程度直線

(図表 2.1.2)



で置きかえることができるか、換言すればどの程度直線に戻すこと（直線に回帰すること）ができるか、からきている。

(図表 2.1.3)



また、 $C = a + bY$  のような式、すなわち、原因となる変数（説明変数）と結果となる変数（被説明変数）との間の定量的な関係を示す方程式を「回帰方程式」という。

(線型回帰・非線型回帰、単回帰・重回帰)

被説明変数と説明変数の関係が1次式で表わされる場合を「線型回帰」、1次式でない場合を「非線型回帰」と呼ぶ(注)。

(注) 非線型回帰方程式とは、被説明変数と説明変数の関係が線型関係でないということである。しかし、非線型回帰方程式の中には対数変換等により線型回帰方程式に変換することが可能なものもある。たとえば

$$y = ax^b$$

は非線型回帰方程式であるが、両辺に対数をとると、

$$\log y = \log a + b \log x$$

となり、 $Y = \log y$ ,  $A = \log a$ ,  $X = \log x$  として上式を書き直せば線型回帰方程式となる。また、

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

は  $x$  に関して非線型であるが、 $X_1 = x$ ,  $X_2 = x^2$ ,  $X_3 = x^3$  とすれば、線型回帰方程式となる。

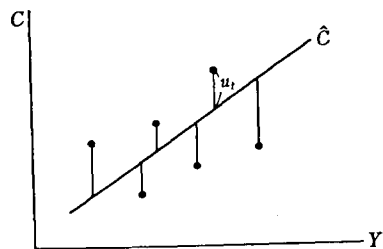
また、2つ以上の説明変数に対する回帰を考える場合（たとえば  $Y = a + bX_1 + cX_2$ ）を「重回帰」と呼び、これに対し、1つの説明変数の場合を特に「単回帰」と呼ぶことがある。以下では、線型回帰の場合について考えることとする。

## 2 最小2乗法 (ordinary least squares, 略称 OLS)

### (イ) 概要

最小2乗法は、回帰方程式のパラメータ（前例  $C = a + bY$  では  $a$  と  $b$  に相当）を求める最も一般的な手法である。その考え方を概念的に言えば、下図において、各点から回帰直線までの長さ ( $u_i$ ) の2乗和を最小にするようにパラメータを設定することにほかならない(注)。

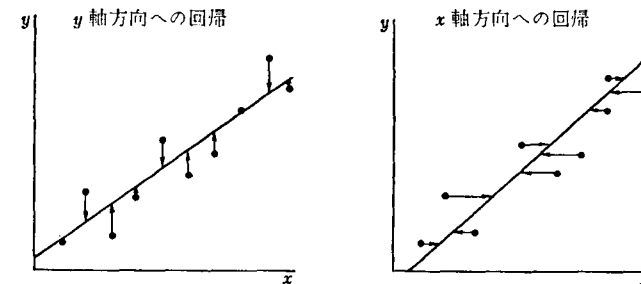
(図表 2.1.4)



(注)  $x$  と  $y$  についてデータが与えられたとき、 $y$  軸方向への回帰と  $x$  軸方向への回帰の両方考えることができる。

たとえば  $y$  軸を消費、 $x$  軸を所得とすると、 $y$  軸方向への回帰とは、消費額を所得額によって説明することであり、逆に  $x$  軸方向への回帰とは、消費額から所得額を説明することである。両者は分析しようとしている対象は同じでも、その理論仮説が違っており、因果関係の観点からはまったく逆の立場に立っている。従って回帰を考える場合には、変数間の因果関係を十分考慮する必要がある。

(図表 2.1.5)



すなわち、同手法は現実の値 ( $C$ ) と回帰直線上の値 ( $C$  の推計値  $\hat{C}$ ) との差 ( $C - \hat{C}$ ) の2乗の合計を最小にする方法である。これが「最小2乗法」と呼ばれる所以である。また、最小2乗法(OLS)によるパラメータの推定値を「最小2乗推定値」(注)(least squares estimate, 略称 LSE) という。

(注) 推定量 (estimator) と推定値 (estimate)

厳密には、推定量とは、推定の方式(公式)をいい、推定値とは、推定量に実際のデータを与えて得られる値をいう。

### (ロ) 最小2乗法の前提と最小2乗推定量の性質

$y_i = a + bx_i$  で表わされるモデルに最小2乗法を適用する場合、一般的には誤差が含まれる。誤差は確率的に発生するものとして、これを  $u_i$  で表わせば上記モデルは、

$$y_i = a + bx_i + u_i \dots \dots \dots (2.1.1)$$

と表現できる。誤差項  $u_i$  の確率分布に関する前提は、最小2乗推定量の統計的性質に大きな影響を与えるので重要である。最小2乗法では、誤差項  $u_i$  の確率分布について、次のような前提をおく。

仮定 1. 誤差項の期待値は常にゼロ。

$$E(u_i) = 0$$

仮定 2. 誤差項の分散は観測時点  $t$  とは無関係な定数である。

$$\text{Var}(u_i) = \sigma^2$$

仮定3. 異なる時点の誤差項は互いに無相関（系列相関なし）.

$$E(u_i u_{i-k}) = 0 \quad k \neq 0$$

仮定4. 説明変数と誤差項は互いに無相関.

$$E(u_i (x_i - E(x_i))) = 0$$

仮定5.  $u_i$  は正規分布に従う.

この仮定はときによりはずされることもある.

仮定1から仮定4までが満たされている場合を「標準的な回帰モデル」、さらに仮定5も満たされている場合を「標準的な正規回帰モデル」という.

以上の仮定のうち、特に1から4が重要である。仮定1から仮定4が満たされた場合、線型回帰の最小2乗推定量はガウス・マルコフの定理<sup>(注)</sup>より「最小分散線型不偏推定量 (best linear unbiased estimator, 略称 BLUE)」と呼ばれる(42ページ)。なお、一般的には以上の仮定1から仮定5のうちいくつかは成立しない場合が多いが、こうしたケースについては後段の「最小2乗法の修正」(98ページ)で扱うこととする。

(注) ガウス・マルコフの定理とは次のことをいう。

誤差項が仮定1から仮定4の条件を満たす場合、最小2乗推定量は線型不偏推定量の中で最小分散を持つ。

ここで注意すべきことは、ガウス・マルコフの定理はあくまで線型不偏推定量の中で最小2乗推定量の分散が最も小さいことを意味している点である。従って、広範囲の(a, b)に対して最小2乗推定量より分散は小さいが偏りのある推定量や、最小2乗推定量より分散が小さい非線型推定量が存在する可能性がある（しかし仮定5.を加えれば最小分散になる）。

### 3 回帰分析の基本的概念

#### (イ) 決 定 係 数

回帰方程式において被説明変数の全変動のうち、説明変数の変動によって説明される部分の割合を「決定係数」と呼ぶ。従って、決定係数は方程式の説明力の大小を示す指標といえよう。

決定係数は次の式から導かれる。

$y$  の全分散 ( $s_y^2$ ) (=  $y$  の全変動)

=  $y$  の全分散のうち回帰式によって説明された分散 ( $s_r^2$ )

+  $y$  の全分散のうち回帰式によって説明されない分散 ( $s_e^2$ )

決定係数を  $R^2 = \frac{s_r^2}{s_y^2}$  と定義すれば上式より、

$$R^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2} \dots\dots\dots (2.1.2)$$

一般に  $s_y^2 \geq s_e^2$  であるから  $0 \leq R^2 \leq 1$  となる。

#### (ロ) 自由度修正済決定係数

次に自由度(84ページ)で調整した決定係数(自由度修正済決定係数)についてみてみよう。通常用いられる決定係数は次のように定義される「自由度修正済決定係数」( $\bar{R}^2$ )である。

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2) \dots\dots\dots (2.1.3)$$

ただし、 $n$  : データ数、 $k$  : 説明変数の数(含む定数項)

従って、 $\bar{R}^2 \leq R^2$  ( $R^2=1$ の場合のみ  $\bar{R}^2=R^2=1$ )。

決定係数( $R^2$ )および自由度修正済決定係数( $\bar{R}^2$ )は1に近いほど回帰方程式の説明力が高いことを示す。しかし、決定係数および自由度修正済決定係数の大きさだけから回帰方程式の優劣を評価することはできない。

たとえば、被説明変数  $y$  を説明変数  $x_1, x_2, x_3$  で回帰した場合の  $\bar{R}^2$  が0.988,  $x_4, x_5, x_6$  で回帰した場合の  $\bar{R}^2$  が0.978であった場合、どちらがより優れた定式化であるかを  $\bar{R}^2$  だけによって判断することはできない。

(決定係数と相関係数の関係)

単回帰式  $y_i = a + bx_i$  において用いられる決定係数( $R^2$ )と相関係数( $r$ , 10ページ)の間には次のような関係がある。

$$R^2 = r^2$$

すなわち、相関係数( $r$ )の2乗は決定係数に等しい。

## (ハ) 自由度

(一般的概念)

$n$  個のデータが互いに独立 (すなわち, 自由な値をとり得る) であるとす。今,  $n$  個のデータの平均が  $\bar{x}$  で与えられた場合, 最初の  $(n-1)$  個は自由な値をとりうるが, 最後の 1 個は, 全体の平均値が  $\bar{x}$  でなければならないという制約のために, 自動的に決まることとなる。すなわち, この場合には,  $n$  個のうち,  $(n-1)$  個のデータしか自由な値をとれない訳で, このような自由な値をとり得るデータの数に相当するものを「自由度」という。

たとえば, 母集団から選ばれた  $n$  個のデータ  $(x_i, i=1, \dots, n)$  について標本平均  $\bar{x}$  および標本分散  $s^2$  を求めることを考えよう。

$$\begin{cases} \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \\ s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \end{cases} \dots\dots\dots (2.1.4)$$

母集団の平均を  $\mu$ , 分散を  $\sigma^2$  とする。この時,  $\bar{x}$  の期待値  $E(\bar{x})$  は  $\mu$  と等しいが,  $s^2$  の期待値  $E(s^2)$  は  $\sigma^2$  と等しくない。すなわち, 標本平均は母平均の不偏推定量であるが, 標本分散は母分散の不偏推定量とはならない。母分散の不偏推定量を求めるためには,  $s^2$  の分母  $n$  を自由度  $(n-1)$  で置き換えて, 以下の値を算出することが必要である (これを不偏分散という)。

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots\dots\dots (2.1.5)$$

ここで  $(n-1)$  で除しているのは,  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  という制約によって,  $n$  個のうち  $(n-1)$  個までは自由な値をとることができるが, 残りの 1 個は自動的に決められる ( $n\bar{x} - \sum_{i=1}^{n-1} x_i = x_n$  で求められる) ことになり, この  $n$  個目の値については自由な値がとれないためである。そこで,  $s^2$  の自由度は  $(n-1)$  となる。

(回帰分析における自由度の意味)

$n$  個のデータから回帰モデル  $y = a + bx$  を最小 2 乗法により計測する場合を考える。残差の 2 乗和を最小にするという条件から次式が導かれる。

$$\begin{cases} \sum_{i=1}^n \{y_i - (\hat{a} + \hat{b}x_i)\} = 0 \\ \sum_{i=1}^n x_i \{y_i - (\hat{a} + \hat{b}x_i)\} = 0 \end{cases} \dots\dots\dots (2.1.6)$$

残差を  $\hat{u}_i = y_i - (\hat{a} + \hat{b}x_i)$  とすれば

$$\begin{cases} \sum_{i=1}^n \hat{u}_i = 0 \\ \sum_{i=1}^n x_i \hat{u}_i = 0 \end{cases} \dots\dots\dots (2.1.7)$$

すなわち, 残差項  $\hat{u}_i$  は (2.1.7) 式の 2 つの関係を満たさねばならない。従って  $n$  個の残差のうち  $(n-2)$  個までしか自由な値をとることができない。つまり, 残差項の自由度は  $(n-2)$  となる。

一般には, データの数 ( $n$ ) から推定されなければならないパラメータの数 ( $k$ ) を引いた値  $(n-k)$  (注) が自由度となる。

(注) ① 重回帰  $y = \underbrace{a + b_1x_1 + b_2x_2 + \dots + b_{k-1}x_{k-1}}_{k \text{ 項}} + u$  の場合

最小 2 乗法の条件より,

$$\begin{cases} \sum_{i=1}^n \hat{u}_i = 0 \\ \sum_{i=1}^n x_{ij} \hat{u}_i = 0 \quad (j=1, 2, \dots, k-1) \end{cases}$$

従って残差  $\hat{u}_i$  は  $\{1 + (k-1)\}$  本の独立な関係式を満たさなければならない。この時推定すべきパラメータは  $a, b_1, b_2, \dots, b_{k-1}$  の  $\{1 + (k-1)\}$  個。従って, 自由度は  $(n-k)$  となる。

② 定数項のない  $y_i = bx_i$  の場合

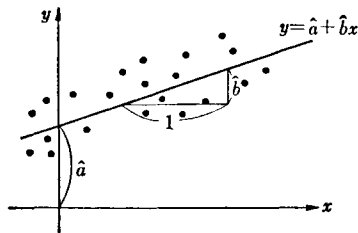
最小 2 乗法の条件より,

$$\sum_{i=1}^n x_i \hat{u}_i = 0$$

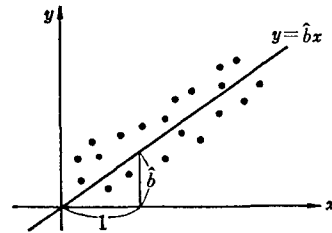
$\hat{u}_i$  はこの関係のみを満たせばよく, 自由度の減少は 1 つ。つまり自由度は  $(n-1)$  (推定すべきパラメータは  $b$  のみであるため一般論は成り立つ)。

以上の点を視覚的にとらえるため、単回帰の場合についてグラフ上で考えてみよう。

(図表 2.1.6)



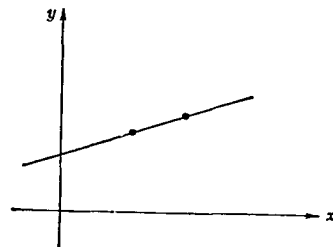
(図表 2.1.7)



図表(2.1.6)の場合、回帰線をひくためには、 $y$ 切片と傾きを与えなければならない。すなわち、この2つが制約となって自由度は $(n-2)$ となる。これに対し、定数項のない図表(2.1.7)の場合、回帰線は、観測値とは無関係に原点を通るため、傾きだけを与えればよいこととなり、自由度は、 $(n-1)$ である。それでは自由度は回帰分析の結果にいかなる影響を与えるのであろうか。

データ数( $n$ )に比べ推定すべきパラメータの数( $k, (n \geq k)$ )が多い時、すなわち、自由度 $(n-k)$ が低い時、決定係数 $R^2$ (82ページ)は高くなる。たとえば、データ数が2個の単回帰では、図表(2.1.8)のように残差がゼロとなり、 $R^2$ は最大値の1をとる。

(図表 2.1.8)



この時、直線のあてはまりはよくなるが、推定値の信頼度は低くなってしまう。一方、自由度修正決定係数(83ページ)をみると、自由度が小さい時、 $R^2$ と $\bar{R}^2$ の差は大きくなり、自由度が大きい時、 $R^2 \approx \bar{R}^2$ となる。つまり、みかけの“あてはまり”を“信頼度”で調整した値を得るこ

とができる(消費を所得で説明することを例にとると、 $n=3$ (昭和50/I ~ 50/III)の時、 $R^2=0.952$ ,  $\bar{R}^2=0.903$ ( $R^2 - \bar{R}^2=0.049$ )であるのに対し、 $n=22$ (昭和50/I ~ 55/II)の時、 $R^2=0.948$ ,  $\bar{R}^2=0.946$ ( $R^2 - \bar{R}^2=0.002$ )となる)。従って、回帰においては、自由度修正決定係数を用いて判断する方が良い(ただし、自由度がゼロの時、自由度修正決定係数は定義されない)。

また、自由度はパラメータの検定( $t$ 検定(87ページ)、 $F$ 検定(90ページ)等)においても重要である。

## (二) $t$ 値

$t$ 値は、最小2乗法によって計測されたパラメータが有意であるかどうかを判定するために用いられる統計量である。ここでいう「有意」とは、パラメータが統計的にみてゼロでないと判断されるという意味である。

$t$ 値を用いたパラメータの有意性の検定( $t$ 検定)の原理は次のとおり。

$$y_i = a + bx_i + u_i \quad (2.1.8)$$

において $b=0$ すなわち、「 $x$ は $y$ に全く影響を与えない」という帰無仮説(42ページ)をたてると、

$$t_i = \frac{\hat{b}}{\hat{\sigma}_b} \quad (2.1.9)$$

(ただし $\hat{b}$ は $b$ の最小2乗推定値、 $\hat{\sigma}_b$ は $\hat{b}$ の分散の推定値)

で定義される $b$ の $t$ 値 $t_i$ は、自由度 $(n-2)$ ( $n$ はデータ数、2はパラメータの個数)の $t$ 分布(36ページ)に従うことが知られている。従って、 $t$ 分布表により帰無仮説 $b=0$ を検定することができる。

たとえば消費 $y_i$ を所得 $x_i$ で回帰する場合( $y_i = a + bx_i + u_i$ )を考えてみよう。昭和52年Ⅳ期から昭和58年Ⅰ期までの22個のデータを用いて次のような式が計測されたとする。

$$\hat{y}_i = -37147.2 + 1.1146x_i$$

所得のパラメータの標準誤差（分散の不偏推定量の平方根）は、

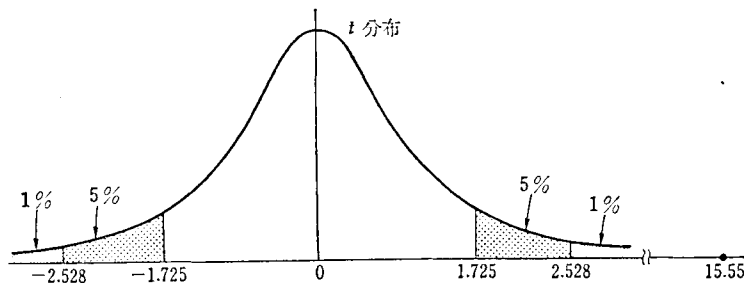
$$\hat{\sigma}_b = \sqrt{\frac{1}{\sum_{i=1}^{22} (x_i - \bar{x})^2}} \hat{\sigma}_u = 0.0717$$

である。従って、

$$t \text{ 値} = \frac{1.1146}{0.0717} = 15.55$$

となり、データ数が22個であるから  $t$  値は自由度20の  $t$  分布に従うが、このように大きな値が得られる確率は非常に少ない。すなわち  $b=0$  と仮定して計算した確率は非常に小さく、「所得は消費に影響を与えない」という仮説は棄却され、 $\hat{b}$  は統計的にみて有意と判断される(注)。

(図表 2.1.9)



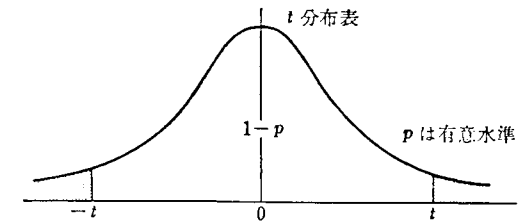
(注) ここで注意すべきことは、前述の  $t$  検定で棄却された仮説は、 $b=0$  である点である。すなわち、ここでの検定結果は  $b=0$  ではないことを示すにすぎず、 $b=1.1146$  の信頼性を示すものではない。従って「 $t$  値の大きさ（絶対値）＝パラメータの信頼性」と誤解してはならない。 $t$  値の大きさは当該パラメータがゼロとは異なる度合いを示すにすぎない。

$t$  検定において問題となるのは、有意水準（帰無仮説が正しいときそれを間違えて棄却する確率）をどのように設定するかである。通常は5%あるいは1%が用いられるが、絶対的基準はなく、分析者の判断に委ねられている。

一般的には  $t$  値が2以上（学者によっては  $\sqrt{2} \approx 1.4$  以上）あればパラメータは有意であるとする人が多いが、 $t$  分布表はデータ数( $n$ )によ

て大きく値が変わり、データが少ない時には、 $t$  値の絶対値がかなり大きくないと有意水準を満たさないことに注意すべきである。

(図表 2.1.10)



| 自 由 度    | 有意水準 5 % | 有意水準 1 % |
|----------|----------|----------|
|          | $t$ の 値  | $t$ の 値  |
| 1        | 12.706   | 63.657   |
| 2        | 4.303    | 9.925    |
| 5        | 2.571    | 4.032    |
| 10       | 2.228    | 3.169    |
| 15       | 2.131    | 2.947    |
| 20       | 2.086    | 2.845    |
| 30       | 2.042    | 2.750    |
| 40       | 2.021    | 2.704    |
| 120      | 1.980    | 2.617    |
| $\infty$ | 1.960    | 2.576    |

(出典) Biometrika Tables for Statisticians, Vol. 1 第3版 (1966).

#### (ホ) 方程式の標準誤差

次式で与えられる  $s$  を回帰方程式の標準誤差という。

$$s = \sqrt{\frac{1}{n-k} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots\dots\dots (2.1.10)$$

$n$  : データ数,  $k$  : 独立変数の数 (含む定数項),  $y_i$  : 実績値,