

平成 26 年度 中級計量経済学・応用計量経済学
講義ノート 2: 線形回帰モデル

このノートでは、線形回帰モデルの理論を紹介する。線形回帰モデルは、経済分析で最も良くつかわれているモデルであると同時に、他の分析手法の基礎となっているため、計量経済学において最も重要なトピックである。線形回帰モデルを使う目的は、ある変数の別の変数への影響を調べることである。こうした変数間の関係を線形に表現して、確率論的な仮定を置くことで統計学的な分析を可能にしたものが線形回帰モデルである。このモデルを最小二乗法を用いて推定し、その結果を用いて、検定や統計的推測を行う。そうした統計手法の理論的背景を学習することが、このノートの目的となる。

2.1 線形回帰モデルと最小二乗法

(y_i, \mathbf{X}_i) という変数の組を $i = 1, \dots, n$ 観測するとする。ある変数 y_i と、別の変数の組 \mathbf{X}_i との関係を線形に表現する。つまり

$$y_i = \beta' \mathbf{X}_i + u_i \quad (1)$$

というモデルを立てる。 u_i は誤差項と呼ばれ、 y_i を決める \mathbf{X}_i 以外のすべての要素をまとめたものである。上のモデルは、線形モデルと呼ばれる。線形回帰モデルと呼ぶ際には、通常、暗黙に \mathbf{X}_i と u_i が無相関であるという仮定、または u_i の \mathbf{X}_i の条件付き期待値が 0 であるという追加的な仮定をおいていることが多い。

変数の組のベクトル \mathbf{X}_i は通常、定数項である 1 を含む。 $\mathbf{X}_i = (1, x_{i2}, \dots, x_{ik})'$ と $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ という、二つのベクトルを定義する。すると

$$\beta' \mathbf{X}_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (2)$$

と書ける。

用語 以下の用語は、講義を通して繰り返し使用する。

- β_1 : 切片と呼ばれる。
- $(\beta_2, \dots, \beta_k)$: 係数と呼ばれる。
- $y = \beta' \mathbf{X}$: 母回帰線。
- β : 母回帰線の係数 (母数)。
- y_i : 従属変数、被説明変数。
- \mathbf{X}_i : 独立変数、説明変数、共変量、回帰変数など、いろいろな呼び方があり、どの呼び方も使用されている。各名称ごとに少しずつニュアンスは異なるが、その違いはそれほど気にすることは無いと思われる。
- u_i : 誤差項と呼ばれるが、経済分析においては、文字通りの誤差であるという理解は必ずしも正しくない。 y_i を決める要素のうち \mathbf{X}_i 以外のものすべてを含んだ量という理解が適切。 y_i への影響度で考えると、 \mathbf{X}_i よりも重要な要素を含む可能性もある。

線形回帰モデルの係数の推定 母数のベクトル β を推定する。

\mathbf{X}_i を使って y_i を予測するときの誤差の2乗を最小化するやり方を最小二乗推定という。また、その推定量を、最小二乗推定量という。これを OLS(ordinary least squares) と表記する。 β の OLS 推定量を $\hat{\beta}$ とすると、それは、

$$\sum_{i=1}^n (y_i - b' \mathbf{X}_i)^2 \quad (3)$$

を最小化する b の値である。1次の条件は、

$$-\sum_{i=1}^n \mathbf{X}_i (y_i - b' \mathbf{X}_i) = 0 \quad (4)$$

となる。目的関数は凹関数であるので、1次の条件は最小化のための必要十分条件になっている。OLS 推定量 $\hat{\beta}$ は、

$$\hat{\beta} = \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \sum_{i=1}^n \mathbf{X}_i y_i \quad (5)$$

と明示的に書くことができる。

以下の用語を講義を通して使用する。

- $y = \hat{\beta}' \mathbf{X}$: OLS 回帰直線。
- $\hat{y}_i = \hat{\beta}' \mathbf{X}_i$: OLS 回帰線による $X = X_i$ の時の y_i の予測値。
- $\hat{u}_i = y_i - \hat{y}_i$: i 番目の観測値の残差。なお、誤差項 (u_i) との違いを理解しておくこと。

2.2 OLS 推定量の標本分布

推定量の標本分布を求めることが、このノートでの最も重要な点である。そのために以下の仮定を置く。

最小二乗法の性質を調べるための仮定

1. $E(u_i | \mathbf{X}_i) = 0$ あるいは $E(y_i | X_i) = \beta' \mathbf{X}_i$ 。ここで、 $E(u_i | \mathbf{X}_i) = 0$ であるとき $\text{corr}(\mathbf{X}_i u_i) = 0$ となる。(逆は必ずしも成り立たない。)
2. $(\mathbf{X}_i, y_i), i = 1, \dots, n$ は i.i.d. (独立同一分布)。もし、ある母集団から無作為抽出によって観測値を集めた場合、この仮定は満たされる。
3. \mathbf{X}_i と u_i は4次のモーメントを持つ。つまり、すべての m について $0 < E((x_{im} - E(x_{im}))^4) < \infty$ となり、また $E(u_i^4) < \infty$ である。これは異常に大きい X_i や u_i の値をそれほど頻繁には取らないことを意味する。
4. $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)'$ という行列は、列フルランクである。完全な多重共線性がないことを仮定している。

これらの仮定の役割

- 数学的に、OLS 推定量が一致で不偏で漸近正規であることを示すのに使われる。
- OLS による回帰分析がうまくいかない場合を考える際に役に立つ。最初の仮定 ($E(u_i|\mathbf{X}_i) = 0$) が、実証研究においては、もっともよく議論される。

OLS 推定の漸近的性質 OLS 推定量 $\hat{\beta}$ は、ランダムに抽出された標本から計算している。したがって、推定量は、ある確率分布を持つ確率変数である。

OLS 推定量の分布を知ることは、母数の値に関する仮説検定や、信頼区間の構築に必要となる。

誤差項 u_i の分布を決めてしまわずに議論するために、大標本理論 (n を無限に大きくした極限で考える統計理論) を使って推定量の標本分布を近似する。

OLS 推定量は、

- 不偏 ($E(\hat{\beta}) = \beta$) で、
- 一致 ($\hat{\beta} \rightarrow_p \beta$) で、
- 漸近正規:

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N\left(\mathbf{0}, (E(\mathbf{X}_i\mathbf{X}_i'))^{-1} E(u_i^2\mathbf{X}_i\mathbf{X}_i') (E(\mathbf{X}_i\mathbf{X}_i'))^{-1}\right). \quad (6)$$

である。

証明

Proof. 不偏性: まずはじめに、 $\hat{\beta}$ の式に $y_i = \beta'\mathbf{X}_i + u_i$ を代入すると

$$\hat{\beta} = \beta + \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i \quad (7)$$

が成り立つことがわかる。ここで、i.i.d. の仮定と、条件付き平均が 0 である仮定から、

$$E\left\{\left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i | \mathbf{X}\right\} = \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \sum_{i=1}^n \mathbf{X}_i E(u_i | \mathbf{X}_i) = 0 \quad (8)$$

と示せる。つまり、繰り返し期待値の法則から、 $E(\hat{\beta}) = E\{E(\hat{\beta}|\mathbf{X})\} = E(\beta) = \beta$ となる。

一致性: まず、

$$\hat{\beta} = \beta + \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \sum_{i=1}^n \mathbf{X}_i u_i = \beta + \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i u_i \quad (9)$$

と表現できる。大数の法則より、

$$\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i' \rightarrow_p E(\mathbf{X}_i\mathbf{X}_i') \quad (10)$$

となり、右辺は正値定符号である。また、

$$\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i u_i \rightarrow_p E(\mathbf{X}_i u_i) = 0 \quad (11)$$

も示せる。つまり、 $\hat{\beta} \rightarrow_p \beta$ となる。

漸近正規性: まず、

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i u_i. \quad (12)$$

と表現する。中心極限定理により、

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i u_i \rightarrow_d N(\mathbf{0}, E(u_i^2 \mathbf{X}_i \mathbf{X}_i')) \quad (13)$$

となる。したがって、Slutsky の補題により

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d (E(\mathbf{X}_i \mathbf{X}_i'))^{-1} N(\mathbf{0}, E(u_i^2 \mathbf{X}_i \mathbf{X}_i')) \quad (14)$$

$$= N\left(\mathbf{0}, (E(\mathbf{X}_i \mathbf{X}_i'))^{-1} E(u_i^2 \mathbf{X}_i \mathbf{X}_i') (E(\mathbf{X}_i \mathbf{X}_i'))^{-1}\right). \quad (15)$$

□

漸近分散の推定 OLS 推定量の漸近分散は、次のように推定できる。

$$\hat{\mathbf{V}} = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \mathbf{X}_i \mathbf{X}_i' \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1}. \quad (16)$$

この推定量は、上で述べた仮定の下で一致性 ($\hat{\mathbf{V}} \rightarrow_p (E(\mathbf{X}_i \mathbf{X}_i'))^{-1} E(u_i^2 \mathbf{X}_i \mathbf{X}_i') (E(\mathbf{X}_i \mathbf{X}_i'))^{-1}$) を持つ。証明は省略する。

分散均一の場合 もし $E(u_i^2 | \mathbf{X}_i) = \sigma^2$ となるなら、分散均一であるという。そのとき、上の結果はすべて成立するが、漸近分散が簡単になり、

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N\left(\mathbf{0}, \sigma^2 (E(\mathbf{X}_i \mathbf{X}_i'))^{-1}\right) \quad (17)$$

となる。また、漸近分散の推定も $s^2 = \sum_{i=1}^n \hat{u}_i^2 / n$ として、

$$\hat{\mathbf{V}} = s^2 \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \quad (18)$$

となる。

2.3 仮説検定と信頼区間

ある一つの回帰係数に関する検定 ある回帰係数 β_m (ここで、 β_m は、 β の m 番目の要素である) に関する帰無仮説 $H_0: \beta_m = \beta_{m,0}$ は、 t 統計量を使って検定できる。

1. OLS 推定値 $\hat{\beta}_m$ を計算する。
2. $\hat{\beta}_m$ の標準誤差を計算する。 \hat{V}_{mm} を、 \hat{V} の (m, m) 要素とする。 $\hat{\beta}_m$ の標準誤差 $SE(\hat{\beta}_m)$ は

$$SE(\hat{\beta}_m) = \sqrt{\frac{\hat{V}_{mm}}{n}} \quad (19)$$

である。

3. t 統計量を計算する。

$$t = \frac{\hat{\beta}_m - \beta_{m,0}}{SE(\hat{\beta}_m)}. \quad (20)$$

4. p 値を $2\Phi(-|t|)$ として計算する。あるいは、有意水準を、たとえば5%と決め、もし、 $|t| > 1.96$ ならば、 H_0 を棄却する。

片側検定 帰無仮説を $H_0: \beta_m = \beta_{m,0}$ とし、対立仮説を $H_1: \beta_m < \beta_{m,0}$ とする。両側検定と片側検定の違いは、手順の4番にある。

- p 値: $\Phi(t)$.
- 有意水準が5%なら $t < -1.645$ のとき帰無仮説を棄却する。

回帰係数の信頼区間 β_m の95%信頼区間とは、

- 両側検定をしたときに、5%の有意水準では棄却できない帰無仮説のもとでの係数の値の集合。
- 95%の確率で、 β_m の真の値を含む区間。ここで、区間が確率変数である。

95%信頼区間は、

$$(\hat{\beta}_m - 1.96 \times SE(\hat{\beta}_m), \hat{\beta}_m + 1.96 \times SE(\hat{\beta}_m)) \quad (21)$$

として、計算できる。

x を変化させたときの y の変化分の予測値の信頼区間 x_m を Δx_m だけ変化させると、 y の変化分の予測値は、 $\beta_m \Delta x_m$ である。

つまり、 $\beta_m \Delta x_m$ の信頼区間が必要となる。これは

$$(\hat{\beta}_m \Delta x_m - 1.96 \times SE(\hat{\beta}_m) |\Delta x_m|, \hat{\beta}_m \Delta x_m + 1.96 \times SE(\hat{\beta}_m) |\Delta x_m|) \quad (22)$$

として計算できる。

2.4 複合仮説の検定

複合仮説とは、二つ以上の制約のある仮説である。主に、二つ以上の係数がある特定の値であるという仮説を考える。

Wald 統計量と F 統計量 帰無仮説を、 $R\beta - r = 0$ とする。ここで、 R は $q \times k$ の行列で、行フルランクであり、 r は $q \times 1$ のベクトルとする ($q < k$)。

例えば、 $\beta = (\beta_1, \dots, \beta_4)$ であり、帰無仮説が $H_0 : \beta_1 = \beta_2, \beta_3 = 0$ であるとする、それに対応する R と r は、

$$R = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (23)$$

である。

Wald 統計量は、

$$W = n(R\hat{\beta} - r)' (R\hat{V}R')^{-1} (R\hat{\beta} - r) \quad (24)$$

である。 H_0 のもとで、 $W \rightarrow_d \chi_q^2$ である。

F 統計量は、 $F = W/q$ である。

- 注：分散不均一に頑健な Wald あるいは F 統計量を使うこと。多くの統計ソフトでは、特に指定しない限り、分散均一の場合のみ使用できる Wald あるいは F 統計量が計算してくる。
- q は、帰無仮説を成り立たせる最小の制約の数である。例えば、 $\beta_1 = \beta_2 = \beta_3$ という帰無仮説の場合は、制約の数は、 $q = 2$ となる。
- それぞれの係数ごとに検定を行うと、検定の有意水準、あるいは棄却域の設定が難しくなる。仮に、 t_1 か t_2 のどちらかの絶対値が、1.96 を超えたときに、複合仮説を棄却するとどうなるであろうか。このとき、検定の有意水準は、 t_1 と t_2 の相関に依存してしまう。

例えば、 t_1 と t_2 が完全に相関しているとする、有意水準は 5% となる。一方で、仮に t_1 と t_2 が独立であるとする、

$$\Pr(|t_1| < 1.96, |t_2| < 1.96) = \Pr(|t_1| < 1.96) \times \Pr(|t_2| < 1.96) = 0.95^2 = 0.9025 \quad (25)$$

となり、 $\Pr(\text{棄却} | H_0) = 1 - 0.9025 = 0.0975$ となる。帰無仮説は、必要以上によく棄却されることになる。

- β_1 を切片とする。次の帰無仮説の F 統計量を考える。

$$H_0 : \beta_2 = 0, \dots, \beta_k = 0. \quad (26)$$

この場合、 $R = [\mathbf{0}_{(k-1) \times 1} : I_{k-1}]$ で $r = \mathbf{0}_{(k-1) \times 1}$ である。この統計量は、回帰全体の F 統計量と呼ばれ、多くの統計ソフトで回帰をすると自動的に計算される。

複数の係数に関する信頼集合 二つ以上の係数に関する 95% の信頼集合とは、95% の確率で真の係数値を含む集合である。

- これは、5% の有意水準で、係数の組がある値であるという F 検定が棄却できない、係数の値の集合である。
- 2 つの係数の時には、信頼集合は楕円になる。
- 近年では、多くの統計ソフトで計算できるようになってきた。

2.5 R^2

決定係数 R^2 とは、 y_i の標本分散のうち、 \mathbf{X}_i で説明できる割合である。

まず、次のような y_i の分割を考える: $y_i = \hat{y}_i + \hat{u}_i$ 。

- $\hat{y}_i = \hat{\beta}'\mathbf{X}_i$: y_i のうち、モデルによって説明できる部分。
- $\hat{u}_i = y_i - \hat{y}_i$: y_i のうち、モデルで説明できない部分。

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (27)$$

$0 \leq R^2 \leq 1$ である。 R^2 は、回帰変数を増やすと増加する。

R^2 が 1 に近いということは、回帰変数が Y_i の値を予測する精度が高いということである。一方で、 R^2 が 0 に近いということは、回帰変数が Y_i を予測するのにあまり役に立たないということである。

修正済み R^2 R^2 が大きくなったからといって、変数を加えることがモデルの当てはまりを改善したとは一概には言えない。 R^2 は回帰変数を増やすと常に大きくなる。この問題に対処するために、修正済み R^2 を使う。

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (28)$$

回帰変数を増やすと $\sum_{i=1}^n \hat{u}_i^2$ は減るが、 $1/(n-k-1)$ は大きくなる。

- $\bar{R}^2 \leq R^2$.
- \bar{R}^2 は負になることもある。

注意すべき点

- R^2 が増えたからといって、追加した変数が統計的に有意とは限らない。
- R^2 や \bar{R}^2 が大きいからと言って、回帰変数が、被説明変数を決める真の要因になっていないとはいえない。
- R^2 が高いからといって、欠落変数問題 (後ほど解説する) がないとは限らない。
- R^2 が高いからといって、適切な回帰変数の組が選ばれているとは限らない。

2.6 分散均一性と不均一性

- 分散均一性: u_i の条件付き分散が \mathbf{X}_i に依存していないということ。 ($E(u_i^2 | \mathbf{X}_i) = \sigma^2$)
- 分散不均一性: u_i の条件付き分散が \mathbf{X}_i に依存していないということ。

例として、次のような回帰モデルを考える。*Earnings* は所得であり、*MALE* は男性なら 1、女性なら 0 をとる 2 項変数とする。

$$Earnings_i = \beta_1 + \beta_2 MALE_i + u_i. \quad (29)$$

分散均一性とは、所得の分散が男性と女性で同じであるということ。

分散均一性の仮定の功罪

- 制約が強い。
- OLS は、分散均一性が成り立っていれば、有効推定量（最小分散）になる。“The Gauss-Markov theorem”。
- OLS 係数推定値の標準誤差を簡単な式で計算できる。

ここまでで紹介してきた標準誤差の式は、分散不均一に対して頑健なものである。その標準誤差を使用することで、誤差項が分散均一かそうでないかに関わらず、適切に統計的推測ができる。

どちらを使うべきか？ 多くの経済分析では、分散均一性は、非現実的な仮定である。したがって、分散不均一に頑健な標準誤差を使うのがよいであろう。

なお、多くの統計ソフトでは、分散均一の場合のみ使える標準誤差をデフォルトで計算するので、注意すること。

2.7 どの変数を回帰に含めるべきか

経済の実証分析においては、二つの変数の関係、特にある変数が他の変数に与える影響に興味があることが多い。しかし、実際に回帰を行うときは、興味のある変数以外の変数もモデルに入れて、推定を行うことが多く、またそうすることが必要となってくる。そうした他の変数を入れる目的のうち、最も重要なものは、欠落変数のバイアスを回避することである。

例えば、小学校において学級の大きさが学力テストの点数に与える影響を調べたい時に、次のような線形回帰モデルを使ったとする。

$$TestScore = \beta_1 + \beta_2 STR + u \quad (30)$$

として、 $TestScore$ が学力テストの点数で、 STR が教師一人当たりの児童数であるとする。誤差項の u は教師一人当たりの児童数以外に学力テストの成績に影響を与えるよう要素すべてを含んだものである。したがって、 u には、以下のようなものが含まれると考えられる。

- 教師の資質;
- コンピューターを導入しているか;
- 児童の家庭環境。

これらの要素と学級の大きさは、相関していることが多い。この相関が、問題のある結果を出すこともある。

例えば、ある学校は学級の大きさも小さく、学力テストの成績もよかったとする。しかし、その学校に教師の数が多ことから、教育熱心な家庭が校区に居住するようになっているかもしれない。その時には、成績のよさが、学級の大きさから来るのか、それとも家庭環境からくるのかよくわからなくなってしまう。

欠落変数の定義 もし、回帰変数が、回帰に含まれていない変数と相関しており、またその回帰に含まれていない変数が被説明変数に影響を与えているのならば、OLS 推定量は欠落変数のバイアスを被っているという。

欠落変数のバイアスは次のときにおこる。

1. 欠落変数が、回帰変数と相関している。
2. 欠落変数が、被説明変数に影響を与えている。

欠落変数のバイアスは

$$E(u_i | \mathbf{X}_i) \neq 0 \quad (31)$$

ということの意味している。 u_i と X_i は、欠落変数がある時相関している。欠落変数のバイアスは、OLS 推定量の不一致にしてしまう。

説明変数が一つの場合を考える。仮に u_i と X_i が相関しているとする。そのとき、OLS 推定量 $\hat{\beta}_1$ の極限は

$$\hat{\beta}_1 \rightarrow_p \beta_1 + \rho_{Xu} \frac{\sigma_u}{\sigma_X} \quad (32)$$

である。ここで、 $\rho_{Xu} = \text{corr}(X_i, u_i)$ である。

- 欠落変数のバイアスは標本数を増やしても解決しない。
- $|\rho_{Xu}|$ が大きいなら、バイアスも大きい。
- もし ρ_{Xu} が正なら、上向きにバイアスがかかる。

欠落変数のバイアスは、追加で変数を回帰に含めることで解決できる。

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i. \quad (33)$$

β_2 は x_2 という変数の値を固定したときの、 x_1 の y に与える影響である。

条件付き平均に関する独立性 欠落変数のバイアスを回避するためには、回帰変数を追加すればよいということがわかった。 X を興味のある回帰変数、 W_1, W_2, \dots, W_k をバイアス回避のために追加した回帰変数としよう。上で見た、OLS 推定量が一致になるための条件は、

$$E(u | X, W_1, \dots, W_k) = 0 \quad (34)$$

である。この条件のもとでは、 X の影響はもちろんのこと、 W_1, \dots, W_k の影響も正しく推定できる。

しかし、この条件を満たすように、 W_1, \dots, W_k を選ぶのは、難しいかもしれず、また興味のあるのは X の影響であって、 W_1, \dots, W_k の影響はそれほど興味もないことも多い。実は、 X の影響を一致推定するためには、もっと弱い条件で十分である。それが次にあげる条件付き平均に関する独立性である。

$$E(u | X, W_1, \dots, W_k) = \gamma_0 + \gamma_1 W_1 + \dots + \gamma_k W_k. \quad (35)$$

ここで重要なのは、この条件付き平均が X に依存していないことである。なぜこの条件で十分なのかをみるために、 X が 2 項変数の場合を考える。この時、 X の係数を β_X とすると X を 0 から 1 に変えた時の y への影響は β_X で表現でき、それは、

$$E(y|X = 1, W_1, \dots, W_k) - E(y|X = 0, W_1, \dots, W_k) = \beta_X \quad (36)$$

と書ける。左辺はデータから推定できるので、 β_1 を推定することができる。

ただし、条件付き平均に関する独立性しか仮定しない場合は、 u と W_1, \dots, W_k は相関しているので、 W_1, \dots, W_k の y への影響を正しく推定することはできない。つまり、 W_1, \dots, W_k の係数の解釈には注意が必要である。

2.8 回帰分析の解釈

この節では説明を容易にするために、切片の係数を β_0 とし、最初の回帰変数の係数を β_1 としている。

回帰変数が 2 項変数の場合 2 項変数とは、0 か 1 の二つの値しかとらない変数のことである。

基本的に、回帰変数が 2 項変数でも、そうでないときと回帰の仕組みは変わらない。

しかし、 β_1 の解釈が異なってくる。2 項変数の場合の回帰は、平均の差の分析と同じことになる。

D_i を 2 項変数の回帰変数とする。

$$y_i = \beta_0 + \beta_1 D_i + u_i. \quad (37)$$

ここで $E(u_i|D_i) = 0$ とすると、

- $E(y_i|D_i = 0) = \beta_0 \Rightarrow \beta_0$ は、 $D_i = 0$ の時の y_i の平均である。
- $E(y_i|D_i = 1) = \beta_0 + \beta_1 \Rightarrow \beta_0 + \beta_1$ は、 $D_i = 1$ の時の y_i の平均である。
- β_1 はこれらの平均の差である。

β_1 の OLS 推定量は、これらの二つのグループから計算した標本平均の差である。

多項回帰モデル 多項回帰モデルとは

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_r X_i^r + u_i \quad (38)$$

のようなモデルである。 r を次数といい、回帰に含まれている X の最大の乗数である。

- $r = 2$ なら、2 次回帰モデル
- $r = 3$ なら、3 次回帰モデルである。

回帰線が線形かどうかは、 $H_0: \beta_2 = 0, \dots, \beta_r = 0$ という帰無仮説を F 統計量を使って検定することで検定できる。

多項回帰モデルは、回帰変数の影響が、回帰変数の値に依存するという状況をモデル化するために使われる。2 次モデルを考える。

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + u, \quad (39)$$

$$y + \Delta y = \beta_0 + \beta_1 (X + \Delta X) + \beta_2 (X + \Delta X)^2 + u. \quad (40)$$

このとき、

$$\Delta y = \beta_1 \Delta X + 2\beta_2 X \Delta X + \beta_2 (\Delta X)^2, \quad (41)$$

あるいは、

$$\frac{\Delta y}{\Delta X} = \beta_1 + 2\beta_2 X + \beta_2 \Delta X. \quad (42)$$

X の変化が y に与える影響は、

$$\widehat{\Delta y} = \hat{\beta}_1 \Delta X + 2\hat{\beta}_2 X \Delta X + \hat{\beta}_2 (\Delta X)^2 \quad (43)$$

として推定できる。注意すべきは、この影響は、 X の初期値と変化の大きさ ΔX に依存していることである。

推定された影響の標準誤差も計算することができる。例えば、 X を 10 から 11 に増やした時の y の変化分は、

$$\widehat{\Delta y} = (\hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2) \quad (44)$$

$$= \hat{\beta}_1 + 21\hat{\beta}_2 \quad (45)$$

である。したがって、 $\widehat{\Delta y}$ の標準誤差は $SE(\hat{\beta}_1 + 21\hat{\beta}_2)$ であり、それは

$$SE(\Delta y) = \sqrt{\frac{1}{n} \begin{pmatrix} 0 & 1 & 21 \end{pmatrix} \hat{V} \begin{pmatrix} 0 \\ 1 \\ 21 \end{pmatrix}} \quad (46)$$

として計算できる。

- β_1 を、 X の、 X^2 を固定した時の、 y に与える影響と考えるのは、あまり意味がない。非線形モデルにおいては、回帰関数は、回帰関数の変化が y に与える影響を考えることによって、解釈すべきである。

対数を使った回帰 y あるいは X の対数を使った非線形回帰も重要である。対数の単位当たりの変化は、割合の変化として解釈できる。経済分析においては、変数の値そのものの変化よりも、割合の変化のほうに興味があることも多い。

例:

- 賃金格差。
- 所得の変化としては、10 万円の変化よりも 1%のほうに興味があることもある。
- ある変数の値を 1%変えた時の、他の変数の%で表した変化の大きさを、“弾力性”という。

対数と、割合の関係:

$$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}, \quad (47)$$

ここで、 $\Delta x/x$ は十分に小さいとする。つまり、 $\ln(x)$ が 0.01 変わったなら、それは、 x が 1%変わったのと大体同じになる。

- 線形対数モデル: $y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$.
 - X の 1% の変化は、 y を $0.01\beta_1$ 分増加させる。
- 対数線形モデル: $\ln(y_i) = \beta_0 + \beta_1 X_i + u_i$.
 - X を 1 単位変えた時、 y は $100\beta_1\%$ だけ増加する。
- 対数対数モデル: $\ln(y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$.
 - X を 1% 変えると、 y は $\beta_1\%$ だけ変わる。係数 β_1 は y の X に対する“弾力性”である。

回帰変数の相互作用 相互作用をあらわす項をモデルに入れることによって、ある変数の効果が、他の変数の値に依存する状況を表現することができる。

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i. \quad (48)$$

交差項 $X_{1i} \times X_{2i}$ は相互作用項と呼ばれる。 X_1 の変化が、 X_2 を固定したときに y に与える影響は X_2 の関数になっている。

$$\frac{\Delta y}{\Delta X_1} = \beta_1 + \beta_3 X_2. \quad (49)$$

係数 β_3 は、 X_1 と X_2 を両方とも単位分だけ変化させたときの影響の、それぞれ片方だけ変化させたときの影響の和との差を表現している。 X_1 を ΔX_1 だけ変化させ、 X_2 を ΔX_2 だけ変化させると、

$$\Delta y = (\beta_1 + \beta_3 X_2) \Delta X_1 + (\beta_2 + \beta_3 X_1) \Delta X_2 + \beta_3 \Delta X_1 \Delta X_2. \quad (50)$$

回帰変数は、2 項変数の場合もあり、その時には別の解釈も成り立つ。

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i. \quad (51)$$

- X_i : 連続な回帰変数。
- D_i : 2 項変数。

X の変化の y への影響は、

$$\begin{cases} \beta_1 & \text{もし } D = 0, \\ \beta_1 + \beta_3 & \text{もし } D = 1. \end{cases} \quad (52)$$

- この回帰は、 D の値によって、標本を二つに分け、それぞれの標本で、 y の X に対する回帰を行うことと、同じである。

また、2 つの 2 項変数の相互作用も考えることができる。

$$y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i. \quad (53)$$

このモデルの場合、

$$E(y_i | D_{1i} = 0, D_{2i} = 0) = \beta_0 \quad (54)$$

$$E(y_i | D_{1i} = 1, D_{2i} = 0) = \beta_0 + \beta_1 \quad (55)$$

$$E(y_i | D_{1i} = 0, D_{2i} = 1) = \beta_0 + \beta_2 \quad (56)$$

$$E(y_i | D_{1i} = 1, D_{2i} = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \quad (57)$$

となる。またこの回帰をすることは、標本を D_1 と D_2 の値によって4つにわけ、それぞれの標本の標本平均を計算することと同じになる。なお β_3 は差分内差分法において、興味あるパラメーターになっている。