第 4 章　1 説明変数の線形回帰分析

James H. Stock 著・Mark W. Watson 著・宮尾 龍蔵 訳

『入門計量経済学』（共立出版、2016）
https://www.kyoritsu-pub.co.jp/book/b10003746.html

2025-09-22

1. はじめに

線形回帰分析の問い (Questions in Linear Regression Analysis)

線形回帰分析は、ある変数の変化が別の変数にどのような影響をもたらすかを定量的に分析するための手法である。

具体的な問いの例を以下に示す。

- 飲酒運転に対する罰則の厳格化は、高速道路での事故死亡率にどのような影響を与えるか。
- 小学校の 1 クラスの人数減少は、共通テストの成績にどのような影響を与えるか。
- 大学講義の履修年数増加は、将来の賃金収入にどのような影響を与えるか。これらの問いはすべて、ある変数 X (説明変数 - independent variable) が変化したときに、別の変数 Y (被説明変数 - dependent variable) へ与える効果を問うものである。

X と Y の関係の定量化

本章では、X と Y の線形な関係を分析する。

X と Y を関係付ける直線の傾き (slope) は、X が 1 単位変化したときの Y への影響に相当する。

- 例えば、クラス規模の変化に対するテスト成績の変化は、ギリシャ文字 β を用いて次のように表される。

$$\beta_{\text{ClassSize}} = \frac{\text{TestScore の変化}}{\text{ClassSize の変化}} = \frac{\Delta\ \text{TestScore}}{\Delta\ \text{ClassSize}} \tag{4.1}$$

- ここで、Δ（デルタ）は「〜の変化」を意味する。したがって、$\beta_{\text{ClassSize}}$ は、クラス規模の変化によって生じたテスト成績の変化をクラス規模の変化で割ったものである。

傾き

$$\Delta\ \mathsf{TestScore}\ = \beta_{\mathsf{ClassSize}}\ \times \Delta\ \mathsf{ClassSize} \tag{4.2}$$

もし $\beta_{\mathsf{ClassSize}} = -0.6$ であると仮定すると、クラス規模を 2 名分減少させると（$\Delta\mathsf{ClassSize} = -2$）、テスト成績への影響は $(-0.6) \times (-2) = 1.2$ となる。これは、成績が 1.2 ポイント上昇すると予想されることを意味する。

2. 線形回帰モデル (Linear Regression Model)

母集団の回帰線 (Population Regression Line)

X と Y の線形関係は、以下の式で表現される母集団の回帰線 (population regression line) あるいは 母集団の回帰関数 (population regression function) によって表される。

$$\mathsf{TestScore} = \beta_0 + \beta_{\mathsf{ClassSize}} \times \mathsf{ClassSize} \quad (4.3)$$

- ここで、β_0 は直線の 切片 (intercept)、$\beta_{\mathsf{ClassSize}}$ は直線の 傾き (slope) である。
- この式は、母集団の Y と X の間で平均的に成立する関係を表しており、もし X の値が分かれば、被説明変数 Y の値を $\beta_0 + \beta_1 X$ と予測できる。

回帰モデル

現実には、クラス規模が同じでも、教師の質、教科書、生徒の背景（移民の割合、経済状況）、試験当日の運など、様々な他の要因によってテスト成績は異なりうる。これらの「その他の要因」を考慮するため、回帰モデルは以下のように表される。

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (4.5)$$

- この式は、それぞれの観測値（例えば学区）$i = 1, \ldots, n$ について成り立つ。
- Y_i は 被説明変数 (dependent variable)、X_i は 説明変数 (independent variable または regressor) と呼ばれる。
- β_0 は 母集団回帰線の切片 (intercept)、β_1 は母集団回帰線の傾き (slope) である。

誤差項

- u_i は 誤差項 (error term) と呼ばれる。
- 誤差項は、特定の観測値 i に関して、Y を説明する X 以外のすべての要因（教師の質、生徒の経済環境、テスト当日の運不運、採点ミスなど）を含む。

傾きと切片の解釈

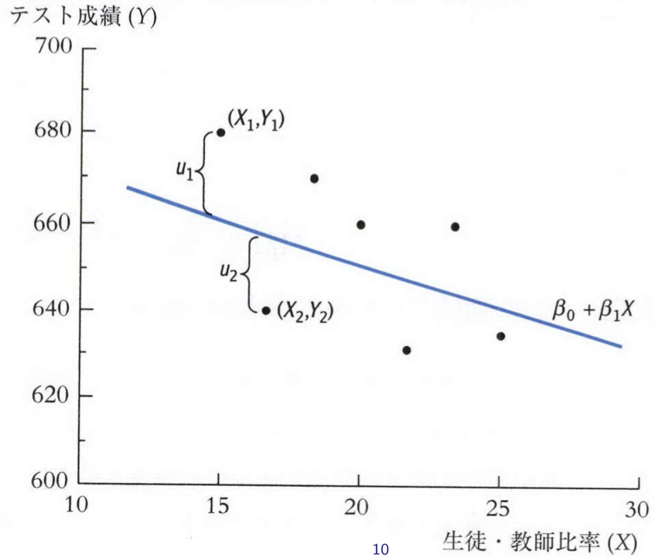
- 傾き β_1 は、X が 1 単位変化したときの Y の変化を表す。
- 切片 β_0 は、X=0 のときの母集団回帰線の値であり、その点で回帰線は Y 軸と交わる。切片が経済学的な意味を持つ場合もあれば、現実的な意味を持たない場合もある（例：クラスに生徒がいないときのテスト成績）。現実的な意味がない場合は、回帰線の水準を決定する係数として数学的に理解すればよい。

散布図と母集団の回帰線 (Scatter Plot and Population Regression Line)

図 4.1 は、テスト成績 (Y) とクラス規模 (X) に関する仮想データと母集団の回帰線を示している。

- この回帰線の傾きは右下がりで ($\beta_1 < 0$)、クラス規模が小さいほどテスト成績が高い傾向を意味する。
- 個々の観測値は、誤差項が存在するため、回帰線上にぴったり位置するわけではない。例えば、回帰線よりも上に位置する観測値 ($u_i > 0$) は予測よりも良い成績を、下に位置する観測値 ($u_i < 0$) は予測よりも悪い成績を示す。

図 4-1


テスト成績 (Y)
700
680
660
640
620
600
10
15
20
25
30
生徒・教師比率 (X)
(X_1,Y_1)
(X_2,Y_2)
u_1
u_2
$\beta_0+\beta_1X$

3. 回帰係数の推定

OLS 推定量 (OLS Estimators)

実際の応用例では、母集団の回帰線における切片 β_0 と傾き β_1 の値は未知である。そのため、手元にあるデータを使ってこれらの未知の係数を推定する。

- この推定には、最小二乗法 (Ordinary Least Squares: OLS) と呼ばれる手法が用いられる。
- OLS は、観察されるデータに推定される回帰線ができるだけ「近く」なるような回帰係数を選ぶ方法である。この「近さ」は、与えられた X に基づいて Y を予測したときの誤りを二乗した値の合計 (sum of squared prediction errors) で測られる。

OLS

- OLS は、以下の予測ミスの二乗和を最小にするような β_0 と β_1 の推定量である。

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 \quad (4.6)$$

OLS 推定量の公式と用語 (OLS Formulas and Terminology)

β_0 と β_1 の OLS 推定量 (OLS estimators) は、それぞれ $\hat{\beta}_0$ と $\hat{\beta}_1$ と表記される。これらの推定量は、以下の公式によって計算される。

$$\hat{\beta}_1 = \frac{S_{XY}}{S_X^2} \quad (4.7)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (4.8)$$

OLS の予測値と残差

- OLS 回帰線 (OLS regression line) は、OLS 推定量を使って得られた直線 $\hat{\beta}_0 + \hat{\beta}_1 X$ である。
- 与えられた X の下で計算される Y の 予測値 (predicted value) は $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ である。
- 第 i 観測値に関する 残差 (residual) は、Y_i とその予測値の差、$u_i = Y_i - \hat{Y}_i$ である。

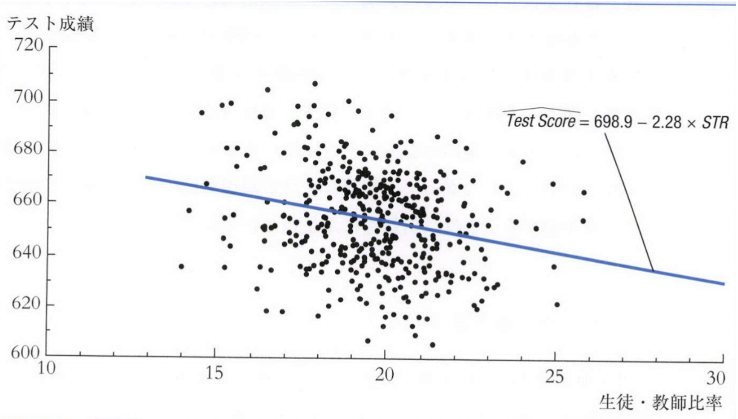
カリフォルニア州テスト成績データへの応用

1998 年のカリフォルニア州 420 学区におけるテスト成績とクラス規模のデータ（表 4.1）を用いて OLS 推定を行った結果、以下の OLS 回帰線が得られた。

$$\widehat{\text{TestScore}} = 698.9 - 2.28 \times \text{STR} \quad (4.11)$$

- 傾き -2.28 は、生徒・教師比率 (STR) が 1 名分増加したとき、その学区のテスト成績は平均で 2.28 ポイント低下することを意味する。
- したがって、生徒・教師比率が 2 名分減少すれば、テスト成績は平均して $(-2) \times (-2.28) = 4.56$ ポイント上昇すると予想される。
- この回帰式から、例えば生徒・教師比率が 20 の学区では、テスト成績は $698.9 - 2.28 \times 20 = 653.3$ と予測される。

図 4-3


テスト成績
720
700
680
660
640
620
600
10
15
20
25
30
生徒・教師比率
$\widehat{Test\ Score} = 698.9 - 2.28 \times STR$

なぜ OLS 推定量を使うのか？

OLS 推定量を使うことには、実践的および理論的な理由がある。

- 実践的な理由:
  - 経済学、ファイナンス、社会科学全般において、回帰分析を行う際の共通の言語となっている。
  - ほとんどすべての統計プログラムや表計算ソフトに組み込まれており、簡単に使用できる。
- 理論的な理由:
  - 特定の仮定（4.4 節で議論）の下で、OLS は 不偏推定量 (unbiased estimator) かつ 一致推定量 (consistent estimator) である。
  - 不偏推定量のある種のグループの中で最も 効率的 (efficient)（分散が最小）となる。

4. 回帰式の当てはまりの指標

R 二乗 (R^2) (R-squared、決定係数)

OLS 回帰線がデータにどの程度うまく当てはまっているかを測る重要な指標の一つに、回帰の R 二乗 (regression R-squared、決定係数) がある。

- R^2 は、被説明変数 Y_i の標本分散が X_i によってどれだけ説明されるか、その割合を測る。

決定係数の意味

被説明変数 Y_i は、その予測値 $\hat{Y}_i$ と残差 $\hat{u}_i$ の和として表現できる。

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{4.13}$$

- この式で考えると、R^2 は $\hat{Y}_i$ の標本分散と Y_i の標本分散との比率。

ESS と TSS の定義

- 説明された二乗和（explained sum of squares, ESS ）とは, Y_i の予測値 $\hat{Y}_i$ とその平均との差の二乗和。

$$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 \tag{4.14}$$

- 全体の二乗和（total sum of squares, TSS ）とは, Y_i とその平均との差の二乗和。

$$TSS = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 \tag{4.15}$$

R 二乗の公式

$$R^2 = \frac{\mathsf{ESS}}{\mathsf{TSS}} = 1 - \frac{\mathsf{SSR}}{\mathsf{TSS}} \quad (4.16), (4.18)$$

ここで、残差の二乗和 (sum of squared residuals: SSR) は OLS 残差の二乗の和である。

R 二乗の解釈

- R^2 は 0 から 1 の間の値を取る。◦ R^2 が 1 に近いときは、説明変数が Y をうまく予測している場合である。◦ R^2 が 0 に近いときは、Y をうまく予測できていない場合である。
- また、Y に対する 1 説明変数 X への回帰における R^2 は、Y と X の 相関係数 (correlation coefficient) の二乗にもなる。

回帰の標準誤差 (Standard Error of the Regression: SER)

回帰の標準誤差 (standard error of the regression: SER) は、回帰誤差 u_i の標準偏差の推定量である。* SER は、観測値が回帰線の周囲に散らばる度合いを被説明変数の単位で表す指標となる。* SER は以下の公式で計算される。

$$\mathrm{SER} = s_u = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} \quad (4.19)$$

ここで、分母に $n-2$ を使用するのは、2 つの回帰係数 (β_0 と β_1) が推定されるため、自由度 (degrees of freedom) の修正を行うためである。

カリフォルニア州テスト成績データへの応用

カリフォルニア州のテスト成績データを用いた回帰式の R^2 は 0.051、SER は 18.6 である。

- R^2 が 0.051 ということは、説明変数 STR は被説明変数 TestScore の分散のわずか 5.1%しか説明できないことを意味する。
- SER が 18.6 ということは、回帰残差の標準偏差がテスト成績の点数単位で 18.6 であり、回帰線を中心とした散らばり度合いが大きいことを意味する。
- R^2 が低く、SER が大きいことは、回帰が「良い」か「悪い」かを意味するものではなく、他の重要な要因もテスト成績に影響を及ぼすことを示唆している。

5. OLS の仮定 (Assumptions of OLS)

OLS が β_0, β_1 の適切な推定量となるためには、以下の 3 つの最小二乗法の仮定 (least squares assumptions) が満たされる必要がある。

1. 誤差項の条件付平均はゼロ ($E(u_i|X_i) = 0$)
2. (X_i, Y_i) は独立かつ同一の分布に従う (i.i.d.)
3. 大きな異常値はほとんど起こりえない (Large Outliers are Rare)

仮定 1

仮定 1: 誤差項の条件付平均はゼロ ($E(u_i|X_i) = 0$) (Conditional Mean of Error Term is Zero)

- X_i が与えられた下で、u_i の条件付分布の平均はゼロであるという仮定である。
- これは、u_i に含まれる「その他の要因」と X_i とは特定の意味で無関係であることを表す。すなわち、X_i の値が与えられている下で、その他の要因に関する分布の平均はゼロである、ということである。
- 数学的に、$E(u_i|X_i = x) = 0$ と表現される。
- この仮定が満たされれば、X_i と u_i は無相関 (uncorrelated) であることも意味する。

仮定 2

仮定 2: (X_i, Y_i) は独立かつ同一の分布に従う (Independent and Identically Distributed: i.i.d.)

- (X_i, Y_i) は、すべての観測値の間で 独立かつ同一の分布に従う (i.i.d.) という仮定である。
- これは、標本データが母集団からの 単純なランダム・サンプリング (simple random sampling) によって得られた場合、妥当な仮定となる。
- この仮定が妥当でない例としては、実験の一部として研究者が任意に X の値を設定する場合や、時系列データ (time series data) のように、時点の近い観測値が独立でない場合が挙げられる。

仮定 3

仮定 3: 大きな異常値はほとんど起こりえない (Large Outliers are Rare)

- X_i もしくは Y_i の観測値が、通常の変動幅を超えてしまうような大きな異常値 (large outliers) がほとんど起こりえないという仮定である。
- OLS は極端な異常値の影響を受けやすいため、この仮定が重要となる。
- 数学的には、X と Y はゼロではない有限の 4 次のモーメント (fourth moment) を持つ、つまり有限の 尖度 (kurtosis) を持つと仮定される。
- 異常値の原因としては、データ入力の誤り（タイプミス、単位の誤りなど）が考えられる。

6. OLS 推定量の標本分布 (Sample Distribution of OLS Estimators)

OLS 推定量 $\hat{\beta}_0, \hat{\beta}_1$ は、ランダムに抽出されたサンプルから計算されるため、確率変数 (random variables) である。

その確率分布は、標本分布 (sample distribution) と呼ばれる。

不偏性 (Unbiasedness)

最小二乗法の仮定（基本概念 4.3）が成り立つとき、OLS 推定量は不偏 (unbiased) である。

- すなわち、OLS 推定量の標本分布の平均は、母集団の真の係数に等しい。

$$E(\hat{\beta}_0) = \beta_0 \quad \text{および} \quad E(\hat{\beta}_1) = \beta_1 \quad (4.22)$$

大標本における正規分布への近似

標本数が十分に大きい場合、**中心極限定理 (Central Limit Theorem)** により、$\hat{\beta}_0, \hat{\beta}_1$ の標本分布は近似的に 正規分布 (normal distribution) に従う。

- $\hat{\beta}_1$ の大標本正規分布は $N(\beta_1, \sigma^2_{\hat{\beta}_1})$ と表される。
- この分布の分散 $\sigma^2_{\hat{\beta}_1}$ は、

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\mathrm{var}\left[(X_i - \mu_X) u_i\right]}{\left[\mathrm{var}\left(X_i\right)\right]^2} \tag{4.20}$$

- この正規近似は、通常 $n > 100$ 程度であれば信頼できるとみなされる。

一致性 (Consistency)

OLS 推定量は 一致性 (consistency) も持つ。

- 標本数が大きくなると、$\hat{\beta}_0$ と $\hat{\beta}_1$ は、母集団の真の係数 β_0, β_1 に高い確率で近づく。
- この性質は、標本数 n が大きくなるにつれ、推定量の分散 $\sigma^2_{\hat{\beta}_0}, \sigma^2_{\hat{\beta}_1}$ がゼロに向かって減少することから示される。

X の分散と $\hat{\beta}_1$ の分散の関係

一般に、X の分散が大きいほど、$\hat{\beta}_1$ の分散 $\sigma^2_{\hat{\beta}_1}$ は小さくなる。

- これは、X の分散が大きければ、$\hat{\beta}_1$ に関する公式の分母が大きくなり、結果として $\sigma^2_{\hat{\beta}_1}$ がより小さくなるためである。
- X の分散が大きいほど、回帰線をより正確に推定できることを意味する。

7. 結論

- 本章では、被説明変数 Y と 1 つの説明変数 X の観測値を用いて、母集団回帰線の切片と傾きを推定する手法として、最小二乗法 (OLS) を解説した。
- OLS は、予測ミスの二乗和を最小化するという基準で回帰線を特定する。
- 最小二乗法の 3 つの仮定（誤差項の条件付平均はゼロ、i.i.d. 標本、大きな異常値は稀）が成り立つならば、OLS 推定量は不偏性 (unbiasedness)、一致性 (consistency) を持ち、標本数が大きいときには正規分布 (normal distribution) に従うという望ましい性質を持つ。

まとめ

- これらの仮定は、OLS の統計的推論の数学的基盤を提供するとともに、回帰分析に伴う問題点を整理する役割を果たす。
- OLS 推定量の標本分布の正規分布への近似は、母集団における真の回帰係数について、標本データだけを使って推論するための強力な分析ツールとなる。
- 次のステップでは、仮説検定や信頼区間の作成のために必要な OLS 推定量の標準誤差について議論される。