回帰分析I

8. 単回帰モデル

1. 単回帰モデルとは？

- 「変数XとYがどのように関係しているか、特にXがYにどのような影響を与えているか」を知りたいとしましょう。

例1：「警察官の数が犯罪率に与える影響」
Y ---犯罪率
X --- 警察官の数

例2：「学校に一年多く行くことの賃金への影響」
Y ---賃金 (例: 時間給)
X --- 教育 (例: 学校に行った年数)

- そのためには「XがYを説明するようなモデル」を設定する必要があります。

- このとき少なくても3つの問題があります。
- 問題1：
  - 現実は複雑。
  - 一つの要因が結果を「完全に」決める、なんてことは一般的にはない。
  - 今の文脈では、XがYを「完全に」決定する、なんてことは一般的にはない。
  - X以外のさまざまな要因がYに影響を与えることをどのようにモデル化すればよい？
- 問題2：
  - XがYと関係しているとして、、、
  - どのように関係している？
  - 数学的に言うと、関数形は？
- 問題3：
  - 知りたいのは、「他の条件を一定としたときの、XがYに与える影響」。
  - すなわち、XのYへの「因果効果」。
  - この効果、うまく抽出できる？

- 次のモデルを考えましょう。

$$Y = \beta_0 + \beta_1 X + U$$

- このモデルを「単純線形回帰モデル」といいます。
  - Yは**従属変数**または**被説明変数**、
  - Xは**独立変数**または**説明変数**、
  - 変数Uは**誤差項** といいます。
- 誤差項Uは、Yに影響を与える X以外のさまざまな要因を捉えるもの。

問題1：一般的には、XがYを「完全に」決定することはない。X以外のさまざまな要因がYに影響を与えることをどのようにモデル化する？

モデルに誤差項を入れることで対処する。

- 一般的に、誤差項は「Yに影響を与えるけれども、観測されない（データとして存在しない）さまざまな要因」と考えるとよいです。
- 従って、単純線形回帰モデルでは、X以外のYに影響を与えるさまざまな要因を、観測されないもの（データとして存在しないもの）として取り扱っている、と考えるとよいです。

5

$$Y = \beta_0 + \beta_1 X + U$$

- 単純線形回帰モデルでは、YとXの間に線形の関係があると仮定しています。
- これを見るために、誤差項Uが捉えるさまざまな要因を固定しまう。すなわちUの変化が0（$\Delta U = 0$）とします。
- すると、Yの変化（ΔY）とXの変化（ΔX）の間には以下のような関係があることが分かります：

$$\Delta Y = \beta_1 \Delta X$$

Uを一定にしたとき、Yの変化はXの変化にβ_1にかけたもの。

6

- Uを一定にしたとき、YとXの関係においてβ_1 は「傾き」を表します。
- そのためβ_1は**傾きの係数**と呼ばれることもあります。
- 単に「**Xの係数**」と呼ばれることもあります。
- また、β_0は**切片**または**定数項**といいます。

問題2：XとYの関係はどのようなものか（関数形）？

単純線形回帰モデルでは、XとYの間に線形関係があると仮定している。

7

- なぜ「XとYの間に線形の関係がある」と仮定する？

  （1）計算が簡単なため。

  （2）現実のデータを線形関係でうまく捉えられる（少なくとも近似できる）ことがよくあるため。

- XとYの関係が非線形だとしたら？

  XとYの間に非線形の関係があったとしても、それを線形の関係に変換し線形回帰モデルのフレームワークで分析できることがあります。

  次のモデルを考えましょう。

$$\ln(Y) = \beta_0 + \beta_1 X + U$$

  左辺は対数、右辺はレベルですから、XとYの関係は線形ではありません。

8

$$\ln(Y) = \beta_0 + \beta_1 X + U$$

- ここでトリックを使います。
- $\ln(Y)$それ自身を一つの変数として考えましょう。
- ここではその変数をZと呼びます。すなわち$Z = \ln(Y)$です。
- モデルを書き直します:

$$Z = \beta_0 + \beta_1 X + U$$

- はい、単純線形回帰モデルになりました。
- ということで、線形の仮定は見た目ほど強いものではありません。

9

- 問題3は最も難しい問題。

「他の条件を一定としたときのXがYに与える影響（因果効果）」を抽出することができるか？

- 先に説明したように、β_1は「他の条件を一定としたときに、XがYに与える影響」を表すもの。
- 従って、問題3は、

データを使ってモデルを推定したときに、信用に値するようなβ_1の推定値を得ることができるか?

と言い換えることができます。

10

- この問題の答えはモデルをどのように推定するかに依存します。
- ここでは最も基本的な「**最小二乗法**」と呼ばれる推定方法について考えます。
- それでは、「モデルを最小二乗法で推定したときに、信用に値するようなβ_1の推定値を得ることができる？」
- 答えは、「ある条件の下ではできる」です。その条件とは、、、
- 次の2つの仮定が満たされるときに、最小二乗法は信用に値するようなβ_1の推定値を与えます。

（1）　$E(U) = 0$
（2）　$E(U|X) = E(U)$

11

仮定1: $E(U) = 0$

- 仮定1:　誤差項の期待値はゼロ。
- この仮定は、Yに影響を与える要因の中で観測されないもの、すなわち誤差項に関する仮定。
- この仮定は定数項β_0がモデルに入っている限り満たされます。
  - 定数項をモデルに入れないと、仮定1が満たされないことあり得ます。
  - よって、モデルには必ず定数項をいれること！

12

仮定1: $E(U)=0$

- なぜ仮定1は定数項がモデルに入っている限り満たされる?
- これを見るためは以下のモデルを考えましょう:

$$Y=\beta_0+\beta_1 X+U,\ E(U)=\alpha_0(\neq 0)$$

- このモデルにおいて、誤差項の期待値はゼロではありません。
- このモデルを、定数項は異なるが、傾きのパラメーターは同じ(すなわちβ_1)で、誤差項の期待値がゼロ のモデルに書き換えることができます。
- 見てみましょう。

13

$$Y=\beta_0+\beta_1 X+U=\underbrace{(\alpha_0+\beta_0)}_{\delta_0}+\beta_1 X+\underbrace{(U-\alpha_0)}_{V}=\delta_0+\beta_1 X+V$$

$$Y=\delta_0+\beta_1 X+V$$

- このモデルの誤差項はVでその期待値は

$$E(V)=E(U-\alpha_0)=E(U)-\alpha_0=\alpha_0-\alpha_0=0$$

- 傾きの係数はβ_1のまま。
- 従って、定数項を入れている限り、「誤差項の期待値はゼロ」という仮定は満たされることになります。
- すなわち、大した仮定ではありません。
- 重要なのは次の仮定です。

14

仮定2: $E(U|X)=E(U)$

- 仮定2: Uの期待値はXの値に依存しない。すなわち、Uの条件付期待値はUの期待値と等しい。
- 仮定1より、仮定2が示すことは、

$$\boxed{E(U|X)=0}$$

- この仮定は、観測できない要因と説明変数に関する<u>非常に重要な仮定</u>です。
- $E(U|X)=0$より、XとUの間に

$$\boxed{COV(X,U)=0}$$

がなりたっている必要がある、ということになります。

15

これまでのまとめ

- 最小二乗法が信用に値するようなβ_1の推定値を与えるためには、(1) $E(U)=0$ (2) $E(U|X)=E(U)$ という2つの仮定が満たされなければならない。
- 仮定(1)は定数項さえ入れれば成り立ちます(従って大した仮定ではありません)。
- 仮定(2)は、仮定(1)とともに、

「XとUの間に相関がない」

ことを必要とします。

- なぜこの仮定が満たされないとき、最小二乗法が与えるβ_1の推定値は信頼できないものになるのか、理由は後ほど詳しく説明します。

16

- コメント：「XとUの間に相関がない」という仮定が成り立っているかいないかは、モデルを見て判断します。
- やってみましょう：
$$WAGE = \beta_0 + \beta_1 EDU + U$$
- このモデルにおいて、EDUとUは相関してそうですか？
- 説明変数と誤差項の相関はいろいろなメカニズムで発生しますが、もっとも分かりやすいものは、「欠落変数」によるものです。
- モデルを見たら、まず、説明変数以外で従属変数に影響を与える要因を考えましょう。
  - その要因は誤差項の中に入ることになります。
  - 誤差項の一部、といってもいいです。
- 次に、その要因が説明変数と相関するかどうか考えます。
  - これこれこういう理由で相関しそうだ、てな感じでいいです。
  - その理由がもっともらしいものであれば、「**XとUの間に相関がない**」という仮定は疑わしいと判断します。

17

- 「**XとUの間に相関がない**」という仮定が<u>成り立ってなさそう</u>であれば、最小二乗法によるβ_1の推定値は、因果効果の推定値としては使い物にはなりません。
- <u>XとUが相関しているのかしていないのか、本当のところは分からないかもしれません</u>。
- ポイントは、相関がありそうなのか、相関はなさそうなのか、というところにあります。
- 相関が<u>ありそう</u>という時点で、因果効果の推定値としては適切ではないという可能性が生まれます。
- その可能性がある以上、因果効果の推定値としては信用に足りうるものではない、ということになります。

18

- もう一個行きます。次の問題を考えましょう
  - Stock and Watson（2007）の第4章の練習問題からです。
- ある教授が期末試験に対する時間プレッシャーの効果を測るための実験を行った。
- 講義に出ている400人の生徒それぞれに同じ試験問題。
- 一部の生徒は試験時間が90分。それ以外の生徒は120分。
- 各生徒はコイントスによってどちらの試験時間かランダムに振り分けられる。
- モデル：$Y = \beta_0 + \beta_1 X + U$　Y: 試験のスコア　X: 試験時間
- 「XとUの間に相関がない」、成り立ちそうですか？

19

母集団の回帰線

$$Y = \beta_0 + \beta_1 X + U$$

- $E(U|X) = 0$の仮定が成り立っているとすると、

$$E(Y \mid X) = E(\beta_0 + \beta_1 X + U \mid X) = \beta_0 + \beta_1 X + E(U \mid X) = \beta_0 + \beta_1 X$$

すなわち

$$E(Y \mid X) = \beta_0 + \beta_1 X$$

となります。

- $\beta_0 + \beta_1 X$ を「母集団の回帰線」と呼びます。
- これは母集団のYとXの間で平均的に成立する関係を示します。
- 従って、もしXの値が分かれば、この母集団の回帰線に従ってYの値を$\beta_0 + \beta_1 X$ と予想します。

20

2. 最小二乗法

- これまでは単純線形回帰モデルを最小二乗法で推定する際の「重要な」仮定について説明しました（その理由は後でします）。
- ここでは最小二乗法がどのような方法なのかを見ていきましょう。
- ただし、その前に仮定（1）と（2）をベースとした推定方法について説明します。
- この方法はモーメント法と言って、みなさんはお気づきではないかもしれませんがすでにお使いになられている方法です。

21

- これから式（1）の推定の仕方を考えます。
- 母集団からのサイズNのランダム・サンプルを

$$\{(X_i, Y_i), i = 1, \ldots, N\}$$

とします。

- 注意：
  - (X_i, Y_i)は実現値ではなくて、選ばれる前のものです。
  - なので、ここでは(X_i, Y_i)は確率変数です。
  - 選ばれて、初めて値が確定します（実現値）。
- ここで式（1）をそれぞれのiについて書き直しておきます。

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

後でこの式を使います。

22

- 仮定1･2は、YとXに関して、どのようなことを示している？

仮定1 $E(U) = 0$が示すもの

$$Y = \beta_0 + \beta_1 X + U \Rightarrow U = Y - \beta_0 - \beta_1 X$$

従って、仮定1は

$$\boxed{E(Y - \beta_0 - \beta_1 X) = 0}$$

と等しいことになります。

- すなわち、仮定1をおくということは、

「母集団において$Y - \beta_0 - \beta_1 X$の期待値はゼロである」

と分析者が考えているということです。

23

仮定2 $E(U|X) = E(U) = 0$が示すもの

$$E(U \mid X) = 0 \Rightarrow COV(X, U) = E(XU) = 0$$

- すなわち、XとUの共分散はゼロ（ XとUは相関しない）。
- したがって仮定2は以下のことを示すことになります。

$$E(XU) = 0 \Rightarrow \boxed{E(X(Y - \beta_0 - \beta_1 X)) = 0}$$

- すなわち仮定2をおくということは、

「母集団において$X(Y - \beta_0 - \beta_1 X)$の期待値はゼロである」

と分析者が考えているということです。

24

ここまでのまとめ：

$$Y = \beta_0 + \beta_1 X + U$$

- 仮定1: $E(U) = 0 \rightarrow E(Y - \beta_0 - \beta_1 X) = 0$ (1)
- 仮定2: $E(U|X) = E(U) = 0 \rightarrow E(UX) = 0 \rightarrow E(X(Y - \beta_0 - \beta_1 x)) = 0$ (2)
- 仮定1と2より導かれた式は、母集団においてXとYがどのように関係しているかを示している。

- ならば、β_0とβ_1の推定値$(\hat{\beta}_0, \hat{\beta}_1)$を、「**サンプルにおいても式(1)・(2)のような関係が成り立つ**」ように選ぼう。
- このアイデアは、母集団において平均的にこれこれこうなっているならば、サンプルでも平均的にはそうなっているはずでしょ、という極めて自然なもの。
- 分析者がおく仮定をダイレクトに使った推定方法と言えます。

25

- $E(Y - \beta_0 - \beta_1 X) = 0$ （$Y - \beta_0 - \beta_1 X$の期待値はゼロ）

$$\Updownarrow$$

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$ （$Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$の標本平均はゼロ）

- $E(X(Y - \beta_0 - \beta_1 X)) = 0$ （$X(Y - \beta_0 - \beta_1 x)$の期待値はゼロ）

$$\Updownarrow$$

$$\frac{1}{N}\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$ （$X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$の標本平均はゼロ）

ここでしたこと：期待値のオペレーター$E(\bullet)$を、標本平均のオペレーター$(1/N)\sum_{i=1}^{N}(\bullet)$に変えただけ。

26

- $E(Y - \beta_0 - \beta_1 X) = 0$ （$Y - \beta_0 - \beta_1 X$の期待値はゼロ）

この推定方法（モーメント法）のアイデア：

母集団で平均的にある関係が成り立っているなら、
データにおいても平均的にはその関係が成り立っている（はず）。

ここでしたこと：期待値のオペレーター$E(\bullet)$を、標本平均のオペレーター$(1/N)\sum_{i=1}^{N}(\bullet)$に変えただけ。

27

- 以下の連立方程式を解きます：

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\frac{1}{N}\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

- (Y_i, X_i) $(i = 1,2,\ldots,N)$には、サンプルが選ばれれば、具体的な数が入ることになります。
- 2本の方程式で、2つの未知数（パラメーター）。
- 解ける（一意に決まる）！

28

- ちなみに、これはモーメント法と呼ばれる推定のアプローチの応用です。
- モーメント法の最も単純な例は、、、
  - 確率変数Yの期待値がμだとする: $E(Y)=\mu$
  - このμ を母集団からのサイズNのランダム・サンプル を使って推定したいとする。
  - $E(Y)=\mu$より
$$E(Y-\mu)=0 \qquad (3)$$
  - この式は<u>母集団において</u>Yがどのような条件を満たすかを示している。
  - ここで μの推定値$\hat{\mu}$を「**<u>サンプルにおいて</u>も式(3)のような関係が成り立つ**」ように選ぼう。

- $E(Y-\mu)=0$ （$Y-\mu$の期待値はゼロ）

$$\Updownarrow$$

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i-\hat{\mu})=0$$ （ $Y_i-\hat{\mu}$ の標本平均はゼロ）

ここでしたこと: 期待値のオペレーター$E(\bullet)$を、標本平均のオペレーター $(1/N)\sum_{i=1}^{N}(\bullet)$ に変えただけ。

- 以下の式を解く:
$$\frac{1}{N}\sum_{i=1}^{N}(Y_i-\hat{\mu})=0 \Rightarrow \hat{\mu}=\frac{1}{N}\sum_{i=1}^{N}Y_i$$
- これは標本平均。従って標本平均は**モーメント法**の一例。

- それでは、先ほどの連立方程式を解きます。
$$\frac{1}{N}\sum_{i=1}^{N}(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0 \qquad \text{(A)}$$
$$\frac{1}{N}\sum_{i=1}^{n}X_i(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0 \qquad \text{(B)}$$
- まずから式(A)から考えます。左辺は、
$$\frac{1}{N}\sum_{i=1}^{N}(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=\frac{1}{N}\sum_{i=1}^{N}Y_i-\hat{\beta}_0\frac{1}{N}\sum_{i=1}^{N}1-\hat{\beta}_1\frac{1}{N}\sum_{i=1}^{N}X_i$$
$$=\overline{Y}-\hat{\beta}_0-\hat{\beta}_1\overline{X}$$

- これより、式(A)は
$$\overline{Y}-\hat{\beta}_0-\hat{\beta}_1\overline{X}=0 \Leftrightarrow \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}$$
- ここで$\overline{Y}$と $\overline{X}$は標本平均。データから簡単に求められる。
- よって、$\hat{\beta}_0$は $\hat{\beta}_1$が分かれば簡単に求められる。
- 次に、式(B)を考えよう。
$$\frac{1}{N}\sum_{i=1}^{N}X_i(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0 \Rightarrow \sum_{i=1}^{N}X_i(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0$$
- この式に$\hat{\beta}_0$を代入すると、、、、

$$\sum_{i=1}^{N} X_i[Y_i - (\overline{Y} - \hat{\beta}_1 \overline{X}) - \hat{\beta}_1 X_i] = 0$$

$$\Rightarrow \sum_{i=1}^{N} X_i[(Y_i - \overline{Y}) - \hat{\beta}_1 (X_i - \overline{X})] = 0$$

$$\Rightarrow \sum_{i=1}^{N} X_i(Y_i - \overline{Y}) - \hat{\beta}_1 \sum_{i=1}^{N} X_i(X_i - \overline{X}) = 0$$

$$\Rightarrow \sum_{i=1}^{N} X_i(Y_i - \overline{Y}) = \hat{\beta}_1 \sum_{i=1}^{N} X_i(X_i - \overline{X})$$

ここで

$$\sum_{i=1}^{N} X_i(X_i - \overline{X}) = \sum_{i=1}^{N} (X_i - \overline{X})^2$$

$$\sum_{i=1}^{N} X_i(Y_i - \overline{Y}) = \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})$$

を使うと、

$$\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y}) = \hat{\beta}_1 \sum_{i=1}^{N} (X_i - \overline{X})^2$$

よって

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}$$

33

まとめ（モーメント法による推定）

モデル: $Y = \beta_0 + \beta_1 X + U$　ランダムサンプル: $\{(X_i, Y_i), i = 1, \ldots, N\}$
仮定1: $E(U) = 0$
仮定2: $E(U|X) = E(U) \rightarrow E(U|X) = 0 \rightarrow E(XU) = 0$

仮定1より $E(Y - \beta_0 - \beta_1 X) = 0 \rightarrow$ $\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$

仮定2より$E(X(Y - \beta_0 - \beta_1 X)) = 0 \rightarrow$ $\frac{1}{N}\sum_{i=1}^{N} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$

これら2本の連立方程式を解いて、

$$\hat{\beta}_0^{MM} = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1^{MM} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N} (X_i - \overline{X})^2}$$

34

- このモーメント法による推定値は、通常、「**最小二乗推定値**」と呼ばれている。
- なぜ？

理由:

$(\hat{\beta}_0, \hat{\beta}_1)$が「**<u>残差</u>**の二乗和を最小にする」ように求められたものと等しくなっているため。

35

- これを理解するために、まず**推計値**(fitted value)と**残差**(residual)を定義します。
- (β_0, β_1)の推定値を、$(\hat{\beta}_0, \hat{\beta}_1)$と表すと、「$X = X_i$のときの$Y$の推計値（モデルが予測する値）」は

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

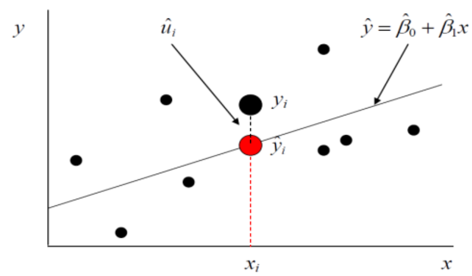
と表せます。

- 残差は「現実の値(Y_i)と予測された値($\hat{Y}_i$)との差」と定義されます:

$$U_i = Y_i - Y_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

36

推計値と残差



$$y_i = \hat{y}_i + \hat{u}_i$$

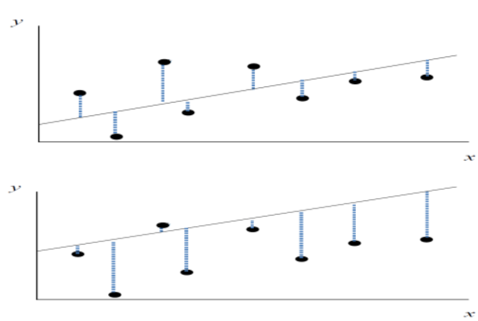
37

- ここで次の推定方法を考えましょう：

残差の二乗和が最小になるように推定値を選ぶ

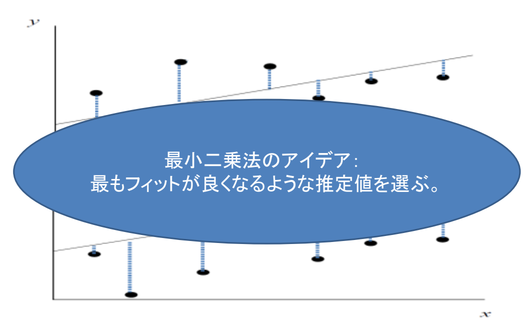
- ノート：モーメント法と最小二乗法は推定のアイデアが全く違うこと、覚えておいてください。

38



どちらの方がデータに“フィット”している？

39



40

- 「**残差**の二乗和が最小になるように$(\hat{\beta}_0, \hat{\beta}_1)$を選ぶ」とき、$(\hat{\beta}_0, \hat{\beta}_1)$は次の最小化問題の解になります。

$$\min_{\hat{\beta}_0,\hat{\beta}_1}\left(\sum_{i=1}^{N} U_i^2\right) \quad \text{または} \quad \min_{\hat{\beta}_0,\hat{\beta}_1}\left(\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2\right)$$

- $\hat{\beta}_0$の一階の条件: $-2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$
- $\hat{\beta}_1$の一階の条件: $-2\sum_{i=1}^{N}X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$
- この2つの式は、モーメント法の時の条件(31ページ参照)と同じ。
- 従って、モーメント法による推定値は、最小二乗法の推定値と同じになり、**残差**の二乗和を最小にしていることになります。

41

- 基礎的な統計や計量経済の教科書では、「残差の二乗和が最小になるように推定値を選ぶ」ことが強調されます。
- しかし、この講義ではモーメント法の考え方を強調しておきます。
- 理由:
  - 最小二乗法は、$E(u|x) = 0$の仮定が満たされないときには、適切な推定値を与えません。
  - 一方、モーメント法は、$E(u|x) = 0$の仮定が満たされない場合でも応用することができます。
  - よってモーメント法はより高度な回帰分析をする際にも使われます。
  - 最小二乗法はモーメント法の特殊ケースと考えるのがいいでしょう。

42

3. モデルの適合度(決定係数)

- ここではモデル適合度の尺度して使われる「**決定係数**(R–squared)」を紹介します。
- 決定係数は「データの従属変数の散らばりのうち、モデルによって説明できた割合」を表すもの。
- 決定係数は以下のように定義されます:

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$$SST \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2;\ SSE \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2;\ SSR \equiv \sum_{i=1}^{n}\hat{u}_i^2$$

43

- $SST \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2$ は「y_i がどれくらい散らばっているか」を示す。
  - これはy_iの標本分散みたいなもの。
- $SSE \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ は「$\hat{y}_i$ がどれくらい散らばっているか」を示す。
  - $\hat{y}_i$の平均は$\bar{y}_i$であることを示せる。
  - 従ってSSEは$\hat{y}_i$ の標本分散みたいなもの。
- $SSR \equiv \sum_{i=1}^{n}\hat{u}_i^2$ は「$\hat{u}_i$ がどれくらい散らばっているか」を示す。
  - $\hat{u}_i$の平均はゼロであることを示せる。
  - 従ってSSRは$\hat{u}_i$ の標本分散みたいなもの。

44

総平方和（SST）は、モデルで説明される平方和（SSE）とモデルで説明されない平方和（SSR）に分解できる。

$SST = SSE + SSR$の証明：

$$
\begin{aligned}
SST &\equiv \sum_{i=1}^{n}(y_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}[(y_i-\hat{y}_i)+(\hat{y}_i-\overline{y})]^2 \\
&= \sum_{i=1}^{n}[\hat{u}_i+(\hat{y}_i-\overline{y})]^2 \\
&= \sum_{i=1}^{n}\hat{u}_i^2+\boxed{2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\overline{y})}+\sum_{i=1}^{2}(\hat{y}_i-\overline{y})^2 \\
&= SSR+0+SSE
\end{aligned}
$$

=0になることが示せる。

45

- $SST = SSE + SSR$の両辺をSSTで割ると

$$1=\frac{SSE}{SST}+\frac{SSR}{SST}\Rightarrow (R^2\equiv)\frac{SSE}{SST}=1-\frac{SSR}{SST}$$

- SSEはSSTより大きくなれない（なぜ？）ので、決定係数は0以上1以下です。
- データの散らばりをモデルが説明していればしているほど、決定係数は大きくなります。

決定係数の解釈上の注意点

- 決定係数が低いときに、「モデルが役に立たない」と結論づけるのは、必ずしも正しくはない。
- 特に、モデリングの目的が「因果効果の推定」である場合は、決定係数を重要視する必要はない。
- なぜなら因果効果の良い推定値を得られるかどうかは決定係数の大きさには直接的には関係しておらず、決定係数が低くても良い推定値を得ることは可能だから。

46

まとめ

- 単純線形回帰モデル

$$Y=\beta_0+\beta_1X+U$$

- Y：従属変数（被説明変数） X：独立変数（説明変数） U：誤差項
- （重要な）仮定：
  - ランダムサンプル
  - $E(U)=0$
  - $E(U|X)=E(U)$

  上の二つの仮定が意味する重要なこと → $E(XU)=0$（XとUの間に相関なし）
- モーメント法：母集団において成り立っている関係（モーメント）は、データにおいても平均的には成り立っている（はず）。そのように推定値を選ぶ。
- 母集団のモーメント　　サンプルのモーメント

$E(Y-\beta_0-\beta_1X)=0$　　$(1/N)\sum_{i=1}^{N}(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0$

$E(X(Y-\beta_0-\beta_1X))=0$　　$(1/N)\sum_{i=1}^{N}X_i(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0$

47

- 推計値（モデルが予測する値）： $Y_i=\hat{\beta}_0+\hat{\beta}_1X_i$
- 残差（現実の値（Y_i）と予測された値（$\hat{Y}_i$）との差）： $U_i=Y_i-Y_i=Y_i-\hat{\beta}_0-\hat{\beta}_1X_i$
- 最小二乗法：残差の二乗和が最小になるように推定値を選ぶ。

$$\min_{\hat{\beta}_0,\hat{\beta}_1}\left(\sum_{i=1}^{N}(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2\right)$$

- モーメント法、最小二乗法ともに推定値は以下の式で与えられる：

$$
\begin{aligned}
\hat{\beta}_0 &= \overline{Y}-\hat{\beta}_1\overline{X} \\
\hat{\beta}_1 &= \sum_{i=1}^{N}(X_i-\overline{X})(Y_i-\overline{Y})/\sum_{i=1}^{N}(X_i-\overline{X})^2
\end{aligned}
$$

- 実際の推定は、データを得たら、変数XとYの標本平均（$\bar{X}$と$\bar{Y}$）を計算、あとはX_iとY_iにそれぞれ数字を入れて足し算、掛け算、割り算なんかをするだけ。
- 決定係数は「データの従属変数の散らばりのうち、モデルによって説明できた割合」を表すもの。0以上1以下の値をとる。

48