# 回帰分析I

9. 最小二乗推定量の期待値と分散
最小二乗推定量の不偏性・一致性



## イントロダクション

- 前回、単純線形回帰(単回帰)モデル、その推定方法について説明しました。
- その際に「重要な仮定」についても説明しました。
- そこで、「重要な仮定」が満たされないときには、最小二乗法が与える推定結果は信用に足りうるものではない(または適切なものではない)と強調しました。
- ここでは、信用に足りうるもの(または適切なもの)、とはどういうことかを定義します。
- そして、「重要な仮定」が満たされないときに、どのようなことが起こるのかを理論的に見ていきます。
- その後、Stataを使ってシミュレーションをし、体感してみましょう。



## 推定量と推定値

- 最初に「推定量」という言葉について説明します。
- 少し一般的な話になります、、、
- 未知のパラメーター$\theta$を持つ母集団からのランダムサンプルを$\{Y_1, \dots, Y_N\}$としましょう。
- このとき、$\theta$の**推定量**とは、

サンプルの実現値が与えられたとき、$\theta$の値を推定する方法

のことを言います。

- これを理解するための簡単な例は以下のものです。



- 例:平均$\mu$(未知のパラメーター)の母集団からのランダムサンプルを$\{Y_1, \dots, Y_N\}$とする。
- $\mu$の推定方法のひとつは標本平均。
  - 標本平均は$\mu$の推定方法(のひとつ)

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

- これの別の言い方は、$\mu$の推定量のひとつは標本平均。
  - 標本平均は$\mu$の推定量(のひとつ)
- $\bar{Y}$は確率変数$Y_1, \dots, Y_N$ のいかなる実現値に対しても、「$\mu$を推定する際に、$Y_1, \dots, Y_N$を足して$N$で割るという方法」ということです。

- それでは<u>推定値</u>とは？
- 確率変数$Y_1, \dots, Y_N$の実現値、すなわち実際のデータ$y_1, \dots, y_N$に、ルールを当てはめることで得られた値を「推定値」といいます。
- 上の例では、

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

が推定値ですね。



- より一般的には、パラメーター$\theta$の推定量$W$は

$$W = h(Y_1, \dots, Y_N)$$

で与えられます。ここで、$h$は<u>既知</u>の関数です。

- ここで重要なことは、<u>推定量$W$は確率変数</u>であるということです。
  - なぜなら$W$は確率変数$Y_1, \dots, Y_N$の関数だから。
- 具体的な値、$y_1, \dots, y_N$を代入して得られた値

$$w = h(y_1, \dots, y_N)$$

は$\theta$の推定値。これは確率変数ではなく「数」（確率変数$W$の実現値）です。

- 少し難しいですが、例を使って考えましょう。



- サイコロを投げて出た目を10倍したお金がもらえるゲームを考えよう。
- サイコロの出目: 確率変数$Y$　（まだ投げていない）
- もらえるお金: $Y$を10倍するよ。$Z(=10Y)$（まだいくらもらえるか分からない：確率変数）
- サイコロを投げた → 6が出た（実現値） → 10・6 = 60円もらえた（実現値）

---

- $N$個の観測値が得られたら全部足して$N$で割るというゲームを考えよう。
- ランダムサンプル: 確率変数$Y_1, \dots, Y_N$（母集団からどれが選ばれるかはまだ決まってない）
- 求める値：$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$（まだどういう値になるかは分からない：確率変数）
- サンプルを選んだ → $y_1, \dots, y_N$（実現値） → $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$（実現値 ）



- 最小二乗法の文脈で、推定量と推定値は何かを考えてみましょう。
- 母集団からのランダムサンプルを$\{(X_i, Y_i): i = 1, \dots, N\}$とします。
- このとき

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

は, $\beta_1$の<u>推定量</u>です。

  - 最小二乗法によるものなので、特に「最小二乗推定量」と言います。
- 従って、最小二乗推定量$\hat{\beta}_1$は確率変数です（よって確率分布を持ちます）
- サンプルの実現値（データ）$\{(x_i, y_i): i = 1, \dots, N\}$を式に代入して得られた値、

$$\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^{N} (x_i - \bar{x})^2$$

は<u>推定値</u>（確率変数$\hat{\beta}_1$の実現値）です。

- もう少し理解を深めるために、、、
- 母集団からのランダムサンプルを$\{(X_i, Y_i): i = 1, \dots, N\}$とします。

- 試行1回目：$\{(x_i^{(1)}, y_i^{(1)}): i = 1, \dots, N\}$

$$\hat{\beta}_1^{(1)} = \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})(y_i^{(1)} - \bar{y}^{(1)}) \,/ \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})^2$$

9

- もう少し理解を深めるために、、、
- 母集団からのランダムサンプルを$\{(X_i, Y_i): i = 1, \dots, N\}$とします。

- 試行1回目：$\{(x_i^{(1)}, y_i^{(1)}): i = 1, \dots, N\}$

$$\hat{\beta}_1^{(1)} = \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})(y_i^{(1)} - \bar{y}^{(1)}) \,/ \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})^2$$

通常はこれで終わりです。ここで、仮想的ですが、$M$回（$M$は大きな数）サンプルをとれるとしましょう。

- 試行2回目：$\{(x_i^{(2)}, y_i^{(2)}): i = 1, \dots, N\}$

母集団からまたランダムに選んでいるので、これは試行一回目のデータとは一般的には同じではありません。

10

- もう少し理解を深めるために、、、
- 母集団からのランダムサンプルを$\{(X_i, Y_i): i = 1, \dots, N\}$とします。

- 試行1回目：$\{(x_i^{(1)}, y_i^{(1)}): i = 1, \dots, N\}$

$$\hat{\beta}_1^{(1)} = \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})(y_i^{(1)} - \bar{y}^{(1)}) \,/ \sum_{i=1}^{N} (x_i^{(1)} - \bar{x}^{(1)})^2$$

通常はこれで終わりです。ここで、仮想的ですが、$M$回（$M$は大きな数）サンプルをとれるとしましょう。

- 試行2回目：$\{(x_i^{(2)}, y_i^{(2)}): i = 1, \dots, N\}$

母集団からまたランダムに選んでいるので、これは試行一回目のデータとは一般的には同じではありません。

$$\hat{\beta}_1^{(2)} = \sum_{i=1}^{N} (x_i^{(2)} - \bar{x}^{(2)})(y_i^{(2)} - \bar{y}^{(2)}) \,/ \sum_{i=1}^{N} (x_i^{(2)} - \bar{x}^{(2)})^2$$

よって$\hat{\beta}_1^{(2)}$は $\hat{\beta}_1^{(1)}$と一般的には同じではない。

11

- 試行$M$回目：$\{(x_i^{(M)}, y_i^{(M)}): i = 1, \dots, N\}$

$$\hat{\beta}_1^{(M)} = \sum_{i=1}^{N} (x_i^{(M)} - \bar{x}^{(M)})(y_i^{(M)} - \bar{y}^{(M)}) \,/ \sum_{i=1}^{N} (x_i^{(M)} - \bar{x}^{(M)})^2$$

- この仮想的な実験で、$\hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \dots, \hat{\beta}_1^{(M)}$が得られます。

- これらをヒストグラムで表したもの、それが最小二乗推定量$\hat{\beta}_1$の確率分布です。
  - ➢ $\hat{\beta}_1$は連続的な確率変数ですから確率密度関数になります。

- この実験、後ほど実際にやってみましょう。

12

## 推定量の不偏性

- これから、「良い」推定方法とは何かについて説明を始めます。
- 最初に、色々な推定の方法がありうることを理解しましょう。
  - 例えば、「あてずっぽう」も一つの推定方法と言えます。
- 様々な推定方法があるとして何をもって良い、悪いを決めたらよいでしょうか？
- 基準は何か？ということです。
- 代表的な基準は3つあります：「不偏性」、「一致性」、「効率性」です。
  - それぞれ推定量の分布の性質に関するものです。
  - ここでは「不偏性」と「一致性」について詳しく説明します。

13

## 不偏性

パラメーター$\theta$の推定量$W$は、もしその期待値が$\theta$ならば、すなわち

$$E(W) = \theta$$

ならば不偏性を持つという。

- 最小二乗法の文脈では、もし

$$E(\hat{\beta}_0) = \beta_0$$
$$E(\hat{\beta}_1) = \beta_1$$

なら最小二乗推定量は不偏性を持つことになります。

14

- ある推定量が不偏性を持つというのは、ざっくり言えば次のようなこと。
- 仮想的ですが、母集団から$M$回（ここでは1000回にしましょう）同じサイズ（$N$）のサンプルが取れるとします。
- それぞれのサンプルについて、その推定方法を使ってパラメーターの推定値を求めます。
  - よってパラメータの推定値が1000個あることになります。
- それらの1000個のパラメーターの推定値の平均が、真のパラメータの値と同じになるなら、その推定量は不偏性を持つといいます。
- そうでないならば不偏性を持たないと言い、その推定方法は「良くない」と判断する材料になります。
- 先ほどの例を使えば、$\hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \dots, \hat{\beta}_1^{(M)}$の平均が、真の値$\beta_1$と同じになるなら最小二乗推定量は不偏性を持つ、ということです。

15

## 推定量のバイアス

推定量のバイアスは以下のように定義される：

$$Bias(W) = E(W) - \theta$$

もし推定量$W$が不偏推定量ならば、バイアスはゼロとなる。

- バイアスがプラスなら「**上にバイアスする**」、バイアスがマイナスなら「**下にバイアスする**」とも言います。
  - 特殊な場合を除いて、いずれの場合も好ましくないです。
- 母集団から1回しかサンプルがとれないとき（通常はこのケース）、推定値が真のパラメーターの値と同じになることは、サンプルサイズ（$N$）が無限大でない限り確率的にはほぼありません。
  - そのため、「平均的には当たるのか？」を一つの基準として使います。
- 以下では、最小二乗推定量が不偏性を持つかどうか（言い換えればバイアスしていないか）どうかを見ていきます。

16

## 最小二乗推定量の不偏性

定理：次の仮定の下で、最小二乗推定量は不偏性を持つ。

仮定1：母集団において、$Y$、$X$、$U$は以下のように関係している

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

仮定2：母集団からのランダムサンプル$\{(X_i, Y_i): i = 1, \dots, N\}$ がある。

仮定3： $X$がいかなる値であっても、 $U$の期待値はゼロである。すなわち$U$の条件付期待値はゼロ。

$$E(U|X) = 0$$

17

- 仮定1と仮定2より、

$$Y_i = \beta_0 + \beta_1 X_i + U_i, i = 1, \dots, N$$

- 仮定3は、仮定2と合わせると

$$E(U_i|X_i) = 0, i = 1, \dots, N$$

さらにすべての$i$について、

$$E(U_i|X_1, \dots, X_i, \dots, X_N) = 0, i = 1, \dots, N$$

18

- **証明**：

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

- $\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{N}(X_i - \bar{X})Y_i$（講義ノート3参照）だから

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

- この式に$Y_i = \beta_0 + \beta_1 X_i + U_i$を代入。

$$\hat{\beta}_1 = \sum_{i=1}^{N} (X_i - \bar{X})(\beta_0 + \beta_1 X_i + U_i) \Big/ \sum_{i=1}^{N} (X_i - \bar{X})^2$$

- この式の分子は

$$\sum_i (X_i - \bar{X})\beta_0 + \sum_i (X_i - \bar{X})\beta_1 X_i + \sum_i (X_i - \bar{X})\bar{U}$$
$$= \beta_0 \sum_i (X_i - \bar{X}) + \beta_1 \sum_i (X_i - \bar{X})X_i + \sum_i (X_i - \bar{X})\bar{U}$$

19

- **証明**：

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

- $\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{N}(X_i - \bar{X})Y_i$（講義ノート3参照）だから

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

- この式に$Y_i = \beta_0 + \beta_1 X_i + U_i$を代入。

$$\hat{\beta}_1 = \sum_{i=1}^{N} (X_i - \bar{X})(\beta_0 + \beta_1 X_i + U_i) \Big/ \sum_{i=1}^{N} (X_i - \bar{X})^2$$

- この式の分子は

$$\sum_i (X_i - \bar{X})\beta_0 + \sum_i (X_i - \bar{X})\beta_1 X_i + \sum_i (X_i - \bar{X})\bar{U}$$
$$= \beta_0 \sum_i (X_i - \bar{X}) + \beta_1 \sum_i (X_i - \bar{X})X_i + \sum_i (X_i - \bar{X})\bar{U}$$

$= 0$（講義ノート3参照）　$= \sum_i (X_i - \bar{X})^2$（講義ノート3参照）

20

- 従って、

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

は、

$$\boxed{\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})U_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}$$

とも表すことができる（この表現とても大切なので覚えておいてください）。

これで証明の準備完了です。

- $\hat{\beta}_1$の条件付き期待値を考えましょう:

21

$$E(\hat{\beta}_1|X_1,\dots,X_N) = \beta_1 + E\left(\frac{\sum_{i=1}^{N}(X_i - \bar{X})U_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\middle| X_1,\dots,X_N\right)$$

$$= \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})E(U_i|X_1,\dots,X_N)}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$= \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\cdot 0}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$= \beta_1$$

- 繰り返し期待値の法則を使うと、

$$E(\hat{\beta}_1) = E(E(\hat{\beta}_1|X_1,\dots,X_N)) = E(\beta_1) = \beta_1$$

- もう一つの方、$E(\hat{\beta}_0) = \beta_0$もさほど難しくなく証明できます。

22

もし$E(U|X) = 0$が成り立たないと$E(U_i|X_1,\dots,X_N) = 0$が成り立たないので、

$$E(\hat{\beta}_1|X_1,\dots,X_N) = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})E(U_i|X_1,\dots,X_N)}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

この式の$\beta_1$の後ろにあるごてごてした項が消えません。

従って、 $E(\hat{\beta}_1|X_1,\dots,X_N) \neq \beta_1$、よって$E(\hat{\beta}_1) \neq \beta_1$となり、不偏性を満たさないということになります。

23

- 仮定1から3の中で、特に大切なのは仮定3です。
- 仮定3の$E(U|X) = 0$ですが、これが成り立っているなら$COV(X,U) = 0$も成り立ちます（逆は必ずしも成り立ちませんが）。
- 従って、

最小二乗推定量が不偏性を持つためには、
少なくとも「$X$と$U$が相関していない」ことが必要！！

- 別の言い方をすれば、

もし$X$と$U$が相関するなら最小二乗推定量はバイアスする！

- よって、説明変数と誤差項が相関するなら、不偏性という観点からは、最小二乗推定量は「適切ではない」といえます。
- そういう方法を使って得られた推定値は、（当たっている可能性はあるかもしれませんが）信用できるものではないと判断します。

24

- 最小二乗推定量の不偏性に関して、もう一つ重要なことがあります。
- 最小二乗推定量の不偏性の証明では、誤差項に特定の分布(例えば正規分布など)は仮定していません。
- 誤差項に関して重要な仮定は$E(U|X) = 0$ だけです。
- 従って、この条件さえ満たしていれば、誤差項はどんな分布であれ、最小二乗法は不偏性を持つことになります。
- 逆に言えば、$E(U|X) = 0$ が満たされないのであれば、誤差項はどんな分布であれ、最小二乗法は不偏性を持たないということになります。



## シミュレーション

- 前述の仮定が満たされれば最小二乗法は本当に不偏性を持つ？
- 前述の仮定が満たされないと最小二乗法はバイアスするっていうけど、実際どの程度のもの？
- シミュレーションして体感してみましょう。
- シミュレーション?



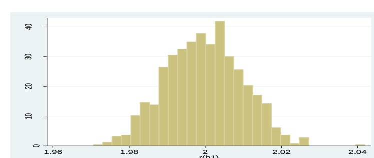
- $N$人いるとします。
- それぞれの人に、乱数を発生させ、$X$を作ってあげます。
  - $x_1, \ldots, x_N$
- それぞれの人に、乱数を発生させ、$U$も作ってあげます。
  - $u_1, \ldots, u_N$
- $y_i = \beta_0 + \beta_1 x_i + u_i$をそれぞれの人に計算します。（$\beta_0, \beta_1$）はあなたが決めたものです。
- すなわち、あなたはパラメータの真の値を知っています。
- $N$人分の$(y_i, x_i)$ができます。
- $N$人分の$(y_i, x_i)$をデータとして使い、真の値（$\beta_0, \beta_1$）を知らない体で、モデル$Y = \beta_0 + \beta_1 X + U$を最小二乗法で推定します。
- 最小二乗推定値$(\hat{\beta}_0, \hat{\beta}_1)$が得られます。
- これを$M$回繰り返します → $\hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_1^{(M)}$が得られます。
- これらの平均が、真の値である$\beta_1$と同じになるかどうかを見ます。



## シミュレーション1

- $N = 200$（サンプルサイズは200）
- $Y = 1 + 2X + U$
- $X$は正規分布、平均5 標準偏差2
- $U$は一様分布、平均0
- $X$と$U$は独立に生成。だから$E(U|X) = 0$ （$X$と$U$に相関無し）
- $M = 1000$ （1000回繰り返す）

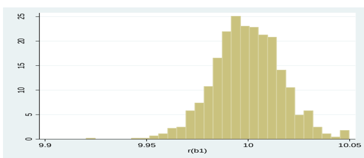
| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | 1.99979 | .0099956 | 1.97064 | 2.041977 |



## シミュレーション2

- $N = 200$（サンプルサイズは200）
- $Y = 4 + 10X + U$
- $X$は正規分布、平均8 標準偏差6
- $U$は平均をゼロに調整した自由度1のカイ二乗分布 （非対称な分布）
- $X$と$U$は独立に生成。だから$E(U|X) = 0$ （$X$と$U$に相関無し）
- $M = 1000$ （1000回繰り返す）

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | 9.99964 | .0166448 | 9.920229 | 10.04983 |

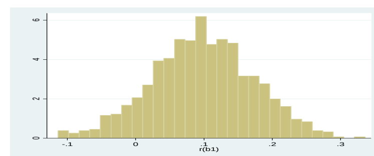



29

## シミュレーション3

- $N = 200$（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3、標準偏差1
- $U$は標準正規分布（平均0）
- $X$と$U$は独立に生成。だから$E(U|X) = 0$ （$X$と$U$に相関無し）
- $M = 1000$ （1000回繰り返す）

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | .098096 | .0744555 | -.1136763 | .335998 |



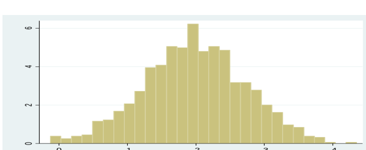

30

## シミュレーション4

- $N = 200$（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。<u>$X$と$U$の相関は0.1</u>
- $M = 1000$ （1000回繰り返す）

本当の効果（0.1）の約2倍！
効果を過大に評価

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | .1981056 | .0740823 | -.0126052 | .434815 |



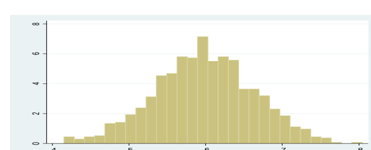

31

## シミュレーション5

- $N = 200$（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。<u>$X$と$U$の相関は0.5</u>
- $M = 1000$ （1000回繰り返す）

本当の効果（0.1）の約6倍！
効果を過大に評価

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | .5983511 | .0644803 | .4149509 | .8043802 |



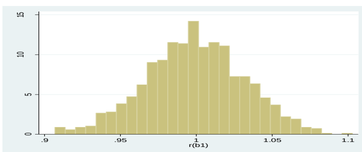

32

## シミュレーション6

- $N$ = 200（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は0.9
- $M$ = 1000 （1000回繰り返す）

本当の効果（0.1）の約10倍！
効果を過大に評価

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | .9991701 | .0324544 | .9068606 | 1.102869 |




33

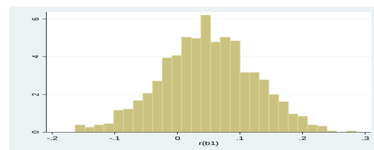
## シミュレーション7

- $N$ = 200（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は-0.05
- $M$ = 1000 （1000回繰り返す）

本当の効果（0.1）の約半分
効果を過小に評価

ランダムサンプルの実現値
によっては、効果をマイナス
と推定

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | .0480984 | .0743624 | -.163409 | .2857028 |



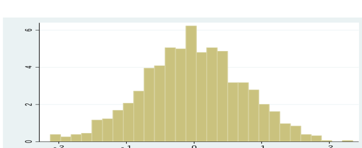

34

## シミュレーション8

- $N$ = 200（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は-0.1
- $M$ = 1000 （1000回繰り返す）

符号が逆！
本当は増やすのにもかかわらず
減らすと推定

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | -.0018944 | .0740823 | -.2126052 | .234815 |




35

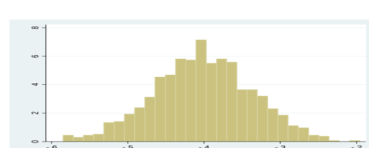
## シミュレーション9

- $N$ = 200（サンプルサイズは200）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は-0.5
- $M$ = 1000 （1000回繰り返す）

符号が逆！
本当は増やすのにもかかわらず
減らすと推定

ランダムサンプルの実現値
がいかなるものであっても、
効果をマイナスと推定

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| b1f | 1000 | -.4016489 | .0644804 | -.5850491 | -.1956198 |




36

- このシミュレーションで分かったこと:
- 真の関係:$X$が$Y$にプラスの因果効果を与える。
- このとき、最小二乗推定量は、、、、
  - $X$と$U$の間に正の相関がある場合、効果を過大に推定する傾向がある。
  - 正の相関が強くなればなるほど、より過大に推定してしまう傾向がある。
  - $X$と$U$の間に負の相関がある場合、効果を過小に推定する傾向がある。
  - 負の相関が強い場合は、因果効果をマイナスと推定してしまうこともありえる。
- 最小二乗法を使った結果を見るとき、このことは絶対に頭に入れておいた方がいいです。
  - 「仮説と整合的な実証結果が得られた」、、、、
  - 本当ですか、バイアスによる可能性はないですか？
  - この姿勢を持つことは本当に大切です。

37

## 推定量の一致性

- 「良い」推定量かどうかを判断する不偏性とは別の基準があります。
- まずはざっくり説明です。
- あなたはある推定方法（推定量）を使おうと思っています。
- ここで、仮想的な話ですが、あなたは好きなだけデータを使えるとします。
- データは多ければ多い方がいいですよね。では、サンプルサイズは無限大としましょう。
- そのとき、その推定方法を使って、真のパラメーターの値をどんぴしゃで当てることができますか？
- そんなに素晴らしい環境にあるにも関わらず、もし当てられないのなら、その推定方法はダメですね。

38

## 推定量の一致性

- 当てられないのなら、その推定量は一致性を持たないと言います。
- 当てられるのなら、一致性を持つと言います。
- 一般的に、「一致性を持つこと」は、推定量として<u>最低限</u>満たさなければならない基準として考えられています。
- 以下では、一致性をフォーマルに定義します。
- その前に「確率収束」という概念を定義します。

39

## 確率収束

<u>確率収束</u>の定義:

確率変数の数列$X_1, \ldots, X_n$を考えよう。もし任意の$\varepsilon > 0$に対して、$n \to \infty$ のときに

$$Pr(|X_n - c| > \varepsilon) \to 0$$

なら、この確率変数の数列は$c$に確率収束するという。また、この数$c$を、確率極限（*probability limit*）または$plim$と呼び、次のように表す。

$$plim X_n = c \text{ または} X_n \xrightarrow{p} c$$

この定義の理解: $\underrightarrow{X_1, X_2, \ldots, X_{1000}, \ldots}$

$n$を十分に大きくとれば、$X$が$c$と異なる確率はゼロ、ということ。

- 基礎数学でやった「極限」（$lim$）の確率変数バージョンと考えればいいです。
- 極限（$lim$）に関するルールは、ほとんどそのまま確率極限（$plim$）にも使えます。

40

## 推定量の一致性

- それでは、一致性をフォーマルに定義しましょう。

一致性（定義）

> 推定量の数列の確率極限（$plim$）が真のパラメータと等しいとき、その推定量は一致性を持つという。

- この定義の理解：
  - 真のパラメータの値を$\theta$とする。
  - ある推定量$\hat{\theta}$を考える。
  - $\hat{\theta}_{(n)}$はサンプルサイズが$n$の時の推定量。

> 大雑把に言えば、「もしデータを好きなだけ（サンプルサイズ＝無限大）集めることができるなら、真のパラメータの値を確率1で当てることができる」なら、その推定量は一致性を持つ、ってこと。

$\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \dots, \hat{\theta}_{(1000)}, \dots \hat{\theta}_{(n)}$はサンプルサイズが$n$の時の推定量。

$\longrightarrow$

- $n$を十分大きくとれば、$\hat{\theta}_{(n)}$が$\theta$と異なる確率はゼロ。



推定量の漸近的バイアス

> パラメーター$\theta$の推定量$W$の漸近的バイアスは以下のように定義される：
>
> $$Asymptotic\ bias(W) = plimW - \theta$$

- もし推定量$W$が一致推定量ならば、漸近的バイアスはゼロとなる。

- 漸近的バイアスが正（$plimW > \theta$）なら、その推定方法は、観測値数が無限大のときに、

確率1で、本当のパラメータの値よりも大きい値を推定値として与える、ということ。

- 漸近的バイアスが負（$plimW < \theta$）なら、その推定方法は、観測値数が無限大のときに、

確率1で、本当のパラメータの値よりも小さい値を推定値として与える、ということ。



- それでは最小二乗推定量は一致性を持つ？

- 答えは、

> $X$と$U$が相関しないなら、最小二乗推定量は一致性を持つ。

- 言い換えると、

> $X$と$U$が相関するなら、最小二乗推定量は一致性を持たない

- ということで、$X$と$U$が相関する可能性がありそうなら、最小二乗法による推定値は因果効果の推定値としては信頼に足りうるものでない、ということになります。



- これを証明するために、一つの有名な法則を使います。

大数の法則（定理）：

> $X_1, X_2, X_3 \dots$を独立で同一の分布（IID）をもつ確率変数の数列とする。またその期待値は$\mu$とする（$\mu = E(X_i), i = 1,2,3, \dots$）。このとき、
>
> $$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

これは$plim$を使って書くと、

$$plim\frac{1}{n}\sum_{i=1}^{n} X_i = E(X_i) = \mu$$

- 母集団からのランダムサンプル＝ IID です。

- なので大雑把に言うと、「サンプルサイズが無限大なら、標本平均（推定量）は母集団の平均をドンピシャで当てる」ですね。

## 最小二乗推定量の一致性

定理：次の仮定の下で、最小二乗推定量は一致性を持つ。

仮定1：母集団において、$Y$、$X$、$U$は、以下のように関係している

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

仮定2：母集団からのランダムサンプル$\{(X_i, Y_i): i = 1, \dots, N\}$ がある。

仮定3：$X$と$U$は相関しない。すなわち$COV(X, U) = 0$

ノート1：ここの仮定3は、不偏性の時の仮定3$E(U|X) = 0$よりも弱い仮定です。 なぜなら$E(U|X) = 0$であれば必ず$COV(X, U) = 0$が成り立ちますが、$COV(X, U) = 0$でも$E(U|X) \neq 0$ はあり得るので。

ノート2：ここでも、誤差項に特定の分布（例：正規分布）は仮定していないことに注意しましょう。

45

- 証明：21ページで導出したように

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})U_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- 講義ノート3のテクニック$\sum_{i=1}^{N}(X_i - \bar{X})U_i = \sum_{i=1}^{N}(X_i - \bar{X})(U_i - \bar{U})$を使う、また分母をともに$N$で割ると、

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(U_i - \bar{U})}{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- 両辺に$plim$を取ります

$$plim\hat{\beta}_1 = plim\left(\beta_1 + \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(U_i - \bar{U})}{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}\right)$$

$$plim\hat{\beta}_1 = \beta_1 + \frac{plim\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(U_i - \bar{U})}{plim\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$plim\hat{\beta}_1 = \beta_1 + \frac{E((X_i - \bar{X})(U_i - \bar{U}))}{E((X_i - \bar{X})^2)}$$

分母分子、ともに
大数の法則を利用

$$plim\hat{\beta}_1 = \beta_1 + \frac{COV(X_i, U_i)}{VAR(X_i)}$$

46

- よって、

$$\boxed{plim\hat{\beta}_1 = \beta_1 + \frac{COV(X, U)}{VAR(X)}}$$

- $COV(X, U) = 0$なら

$$plim\hat{\beta}_1 = \beta_1$$

だから一致性を持つ（証明終わり）。

- 上の式は超重要です（暗記してください）。上の式が意味することは次の通り。$X$が確率変数である限り$VAR(X) > 0$なので、

もし$COV(X, U) > 0$なら、$plim\hat{\beta}_1 > \beta_1$すなわち最小二乗推定量$\hat{\beta}_1$の（漸近）バイアスは正（$\beta_1$を過大に評価するということ）。

もし$COV(X, U) < 0$なら、$plim\hat{\beta}_1 < \beta_1$すなわち最小二乗推定量$\hat{\beta}_1$の（漸近）バイアスは負（$\beta_1$を過小に評価するということ）。

47

## シミュレーション

- 前述の仮定が満たされれば最小二乗法は本当に一致性を持つ？
- 前述の仮定が満たされないと最小二乗法は（漸近）バイアスするっていうけど、実際どの程度のもの？
- シミュレーションして体感してみましょう。
- 基本的にはさっきと同じ。
- 違うところは$N$（サンプルサイズ）を大きくするところです。
- 無限大はPC的に無理なので、サンプルサイズ＝百万個、で勘弁してください。
- このシミュレーションでは繰り返す必要はありません。

48

## シミュレーション10

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 1 + 2X + U$
- $X$は正規分布、平均5 標準偏差2
- $U$は一様分布、平均0
- $X$と$U$は独立に生成。$X$と$U$に相関無し。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  15968062.5      1  15968062.5           Prob > F      =  0.0000
    Residual |  83081.9968999998  .083082163           R-squared     =  0.9948
-------------+------------------------------           Adj R-squared =  0.9948
       Total |  16051144.5999999  16.0511605           Root MSE      =  .28824

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.999903   .0001443  1.4e+04   0.000      1.99962    2.000186
       _cons |   1.000529    .000777  1287.63   0.000     .9990057    1.002052
------------------------------------------------------------------------------
```



## シミュレーション11

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 4 + 10X + U$
- $X$は正規分布、平均8 標準偏差6
- $U$は平均をゼロに調整した自由度1のカイ二乗分布 （非対称な分布）
- $X$と$U$は独立に生成。$X$と$U$に相関無し。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  3.5992e+09      1  3.5992e+09           Prob > F      =  0.0000
    Residual |  2011096.11999998  2.01110013           R-squared     =  0.9994
-------------+------------------------------           Adj R-squared =  0.9994
       Total |  3.6012e+09999999   3601.1712           Root MSE      =  1.4181

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   9.999993   .0002364  4.2e+04   0.000     9.999529    10.00046
       _cons |   4.002407   .0023621  1694.44   0.000     3.997777    4.007036
------------------------------------------------------------------------------
```



## シミュレーション12

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。<u>$X$と$U$の相関は0.1</u>

本当の効果（0.1）の約2倍!
効果を過大に評価

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =40046.14
       Model |  39665.3077      1  39665.3077           Prob > F      =  0.0000
    Residual |  990488.272999998  .990490253           R-squared     =  0.0385
-------------+------------------------------           Adj R-squared =  0.0385
       Total |  1030153.58999999  1.03015461           Root MSE      =  .99523

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .1991847   .0009953   200.12   0.000     .1972339    .2011356
       _cons |   3.703719   .0031462  1177.22   0.000     3.697553    3.709886
------------------------------------------------------------------------------
```



## シミュレーション13

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。<u>$X$と$U$の相関は0.5</u>

本当の効果（0.1）の約6倍!
効果を過大に評価

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  359065.326      1  359065.326           Prob > F      =  0.0000
    Residual |  750369.903999998  .750371404           R-squared     =  0.3236
-------------+------------------------------           Adj R-squared =  0.3236
       Total |  1109435.23999999  1.10943634           Root MSE      =  .86624

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .5992904   .0008663   691.75   0.000     .5975924    .6009884
       _cons |   2.503237   .0027384   914.13   0.000      2.49787    2.508605
------------------------------------------------------------------------------
```

## シミュレーション14

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は-0.05

効果を過小に評価

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) = 2423.13
       Model |  2418.26854     1  2418.26854           Prob > F      =  0.0000
    Residual | 997991.972999998  .997993967           R-squared     =  0.0024
-------------+------------------------------           Adj R-squared =  0.0024
       Total | 1000410.24999999  1.00041124           Root MSE      =    .999

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .0491816   .0009991    49.23   0.000     .0472234    .0511398
       _cons |   4.153734   .0031581  1315.28   0.000     4.147544    4.159923
------------------------------------------------------------------------------
```



## シミュレーション15

- $N$ = 1000000（観測値の数、百万個！です）
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。$X$と$U$の相関は-0.5

効果の方向を逆！に推定

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  160530.816     1  160530.816           Prob > F      =  0.0000
    Residual | 750369.903999998  .750371404           R-squared     =  0.1762
-------------+------------------------------           Adj R-squared =  0.1762
       Total | 910900.719999999   .91090163           Root MSE      =  .86624

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -.4007096   .0008663  -462.53   0.000    -.4024076   -.3990116
       _cons |   5.503237   .0027384  2009.67   0.000      5.49787    5.508605
------------------------------------------------------------------------------
```



- 以下の結果は、回帰分析の中で最も大切なもののひとつです。

$$plim\hat{\beta}_1 = \beta_1 + \frac{COV(X,U)}{VAR(X)}$$

もし$COV(X,U) > 0$なら、$plim\hat{\beta}_1 > \beta_1$すなわち最小二乗推定量$\hat{\beta}_1$の（漸近）バイアスは正（$\beta_1$を過大に評価するということ）。

もし$COV(X,U) < 0$なら、$plim\hat{\beta}_1 < \beta_1$すなわち最小二乗推定量$\hat{\beta}_1$の（漸近）バイアスは負（$\beta_1$を過小に評価するということ）。

- きちんとした計量のトレーニングを受けた人は、この関係性の重要さを理解し記憶し頻繁に使います。
- 他人の分析結果を見たときに、独立変数と誤差項の相関がありうるか、またありうるとすればどちらの方向がもっともらしいかを考え、推定量のバイアスの方向を推測します。
- そして仮説と分析結果が整合的なのかどうかの判断に使います。



- バイアスの重要性、いまいちぴんと来ないかもしれません。でも、重要です！
- 本当に様々なところで回帰分析は使われています。
- 例えば、コンサルが「こんな分析の結果が出ています。こうした方が良いでしょう」といってくること、あると思います。
- 在庫管理のために需要予測をすることあります。その際に回帰分析の結果を使うこともあります。
- 政策決定の一つの材料として、回帰分析の結果が使われることもあります。
- これらの分析の結果が誤ったものであったらどうでしょうか？
  - 効果が小さく（大きく）推定されていたら、、、
  - もっというと、効果が逆に推定されていたら、、、
  - そういう結果に基づいて意思決定したら、、、、結構損しちゃうかもしれませんね。
  - こういうこと現実に起きているでしょうね、程度の差はあれ。

# 決定係数について

- 前回の講義ノートで「決定係数」という指標を紹介しました。
- 「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」ですね。
- そこで、決定係数の解釈の注意点として、
  - 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。
- これをシミュレーションで実感してみましょう。



- $N = 1000000$
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3　標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は独立。
- 従って、最小二乗推定量は一致性を持ちます（因果効果の推定、OKです）。

| Source | SS | df | MS |
|---|---|---|---|
| Model | 10361.0654 | 1 | 10361.0654 |
| Residual | 1002023.64999998 | | 1.00202564 |
| Total | 1012384.7999999 | | 1.01238571 |

| | |
|---|---|
| Number of obs | = 1000000 |
| F( 1,999998) | =10340.12 |
| Prob > F | = 0.0000 |
| R-squared | = 0.0102 |
| Adj R-squared | = 0.0102 |
| Root MSE | = 1.001 |

| y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| x | .1016872 | .001 | 101.69 | 0.000 | .0997272 | .1036472 |
| _cons | 3.994273 | .0031609 | 1263.63 | 0.000 | 3.988078 | 4.000469 |

- 決定係数、見て下さい。1%（笑）です。
- でも、因果効果の推定値いいですよね。



- $N = 1000000$
- $Y = 4 + 0.1X + U$
- $X$は正規分布、平均3 標準偏差1
- $U$は標準正規分布（平均0）
- ただし$X$と$U$は相関有り。<u>$X$と$U$の相関は0.5</u>
- このケースは最小二乗法は一致性を持ちません（因果効果の推定はダメ）

| Source | SS | df | MS |
|---|---|---|---|
| Model | 359065.326 | 1 | 359065.326 |
| Residual | 750369.903999998 | | .750371404 |
| Total | 1109435.23999999 | | 1.10943634 |

| | |
|---|---|
| Number of obs | = 1000000 |
| F( 1,999998) | = . |
| Prob > F | = 0.0000 |
| R-squared | = 0.3236 |
| Adj R-squared | = 0.3236 |
| Root MSE | = .86624 |

| y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| x | .5992904 | .0008663 | 691.75 | 0.000 | .5975924 | .6009884 |
| _cons | 2.503237 | .0027384 | 914.13 | 0.000 | 2.49787 | 2.508605 |

- 推定値、真の値の6倍！です。
- 決定係数、見て下さい。32%です。
- $Y$の「散らばり」は結構説明していますが、因果効果の推定としてはひどいですね。



- 回帰分析を浅く学習した人に限って、決定係数が―、決定係数が―、とことさらに決定係数を重要視します。
- しかし、シミュレーションで見たように、決定係数が低くとも因果効果の推定が上手くいっていることもあります。
- 逆に、決定係数が高くとも、因果効果の推定が上手くいっていないこともあります。
- 結論としては、、、、
  - モデルが複数あるとして、それぞれのモデルにおいて、説明変数と誤差項が相関していない（してそうにない）、という状況があったとしましょう。
  - そういうときには、決定係数が高いモデルの方がいいでしょう。

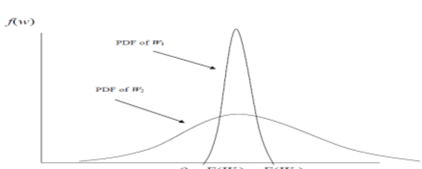
## 最小二乗推定量の分散

- 推定量の分布が不偏性や一致性を持つかどうかにくわえて、「推定量が真の値から平均的にはどのくらい離れているか」を知ることは大切です。
- 通常、その尺度として分散（または標準偏差）を使います。
- 推定量の分散は、後で取り扱う「仮説検定」において重要な役割を果たすことになります。

61

- おまけ：
  - 不偏性、一致性を持つ推定方法が1つではないときに、推定量の分散の情報を使うことによって、最も「良い」推定方法を選ぶことが可能になります。
  - 「効率性」の考え方です。
  - 一致性を持った複数の推定量があるときは、分散の小さい推定量を使った方がいいということです。
  - このコースでは「効率性」についてはあまり取り上げませんが、以下で少しだけ触れておきます。

62




- 推定量が二つあるとします：$W_1$と$W_2$
- ともに不偏性、一致性を持つとします。
- どちらの推定量のほうがより「正確」？
- $W_1$（推定量の分散が小さい）の方は外れたとしても、真の値$\theta$の周りに推定値がでやすいですね。
- 一方、$W_2$（推定量の分散が大きい）はひどい外れ方をすることがありますね。
- ということで、$W_1$の方が推定量としてベター（効率的）です。

63

- 不偏性の証明でした仮定1－3の下で最小二乗推定量の分散を計算することは、可能だけれどもそこそこ大変です。
- ここでは、横断面分析でよく用いられる以下の仮定を置きます（後で取っ払いますが）。

仮定4：誤差項$U$の$X$についての条件つき分散は一定（**均一分散**といいます）。すなわち、

$$VAR(U|X) = \sigma^2$$

- $\sigma^2$が大きいならば、「従属変数に影響を与えるが観測できないもの」のちらばりが大きいということ。
- 仮定4は、仮定2と合わせると

$$VAR(U_i|X_i) = \sigma^2, i = 1, \ldots, N$$

また、いかなる$i$についても

$$VAR(U_i|X_1, \ldots, X_i, \ldots, X_N) = \sigma^2, i = 1, \ldots, N$$

64

コメント：

- 最小二乗推定量の不偏性や一致性に仮定4は必要ありません。すなわち、最小二乗推定量の不偏性や一致性には誤差項が均一分散であるかどうかは関係ありません。
- 仮定4（均一分散）は最小二乗推定量の分散の計算を単純にするためです。
  - いきなり難しいのをやらない、という教育的配慮の側面があります。
- 誤差項$U$の$X$についての条件つき分散が、$X$の値によって変わるとき、誤差項は「**不均一分散**」を持つといいます。
- 近年の実証分析では、「不均一分散」を仮定することが一般的です。
  - 均一と仮定して、実際均一じゃなかったら、なんかひどそうですよね。
  - 実際、ひどいことが起こり得ます（後に説明します）。
  - 「均一」は「不均一」の特殊ケースですから、「不均一」を仮定しておく方が安全ですよね、っていうのが最近の計量分析の考え方です。



- 定理：仮定1－4の下で、

$$\boxed{VAR(\hat{\beta}_1|X_1,\dots,X_N)=\frac{\sigma^2}{\sum_{i=1}^N(X_i-\bar{X})^2}}$$

- $\hat{\beta}_0$の分散もありますが省略します。
- 以下では、条件の部分である$X_1,\dots,X_N$をまとめて$\boldsymbol{X}$と書きます。
- 先ほど導出した

$$\hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^N(X_i-\bar{X})U_i}{\sum_{i=1}^N(X_i-\bar{X})^2}$$

を使います。



- 以下の証明では分散の2つのルールを使います。
- ルール1：$a$と$b$が非確率変数なら、

$$VAR(aX+b)=a^2VAR(X)$$

- ルール2：$a$と$b$が非確率変数で、$X_1$と$X_2$が独立なら

$$VAR(aX_1+bX_2)=a^2VAR(X_1)+b^2VAR(X_2)$$

- ルール2のより一般的なものは、$a_i$（$i=1,\dots N$）が非確率変数で、$X_i$（$i=1,\dots N$）が独立なら

$$VAR(\sum_{i=1}^N a_iX_i)=\sum_{i=1}^N a_i^2VAR(X_i)$$



$$VAR(\hat{\beta}_1\mid\boldsymbol{X})$$

$$=VAR\left(\beta_1+\frac{\sum_{i=1}^N(X_i-\bar{X})U_i}{\sum_{i=1}^N(X_i-\bar{X})^2}\middle|\boldsymbol{X}\right)$$

$$=\left(\frac{1}{\sum_{i=1}^N(X_i-\bar{X})^2}\right)^2VAR\left(\sum_{i=1}^N(X_i-\bar{X})U_i\mid\boldsymbol{X}\right)\longleftarrow$$

$X$が条件なので、$X_i$は定数扱い。ルール1使用。

$$=\left(\frac{1}{\sum_{i=1}^N(X_i-\bar{X})^2}\right)^2\left(\sum_{i=1}^N(X_i-\bar{X})^2VAR(U_i\mid\boldsymbol{X})\right)\longleftarrow$$

ルール2使用。

$$=\left(\frac{1}{\sum_{i=1}^N(X_i-\bar{X})^2}\right)^2\left(\sum_{i=1}^N(X_i-\bar{X})^2\sigma^2\right)\longleftarrow$$

仮定4使用。

$$=\sigma^2\left(\frac{1}{\sum_{i=1}^N(X_i-\bar{X})^2}\right)^2\left(\sum_{i=1}^N(X_i-\bar{X})^2\right)\longleftarrow$$

$\sigma^2$は定数。シグマの外に出る。

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

証明終わり

$$VAR(\hat{\beta}_1|X_1 = x_1, \ldots, X_N = x_N) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

この式は次のことを意味します。

（1） $\sigma^2$が大きくなればなるほど、最小二乗推定量$\hat{\beta}_1$の分散は大きくなる

- これは、「$Y$に影響を与えるが、観測できないもの（誤差項）」のちらばりが大きくなればなるほど、$\beta_1$を正確に推定することが難しくなる、ということです。

（2） $\sum_{i=1}^{N}(x_i - \bar{x})^2$が小さくなればなるほど、最小二乗推定量$\hat{\beta}_1$の分散は大きくなる

- これは、データの説明変数のちらばりが小さくなればなるほど、$\beta_1$を正確に推定することが難しくなる、ということです。

＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝

上の分散に平方根をとったものが、最小二乗推定量$\hat{\beta}_1$の標準偏差となります。

$$SD(\hat{\beta}_1|X_1 = x_1, \ldots, X_N = x_N) = \frac{\sigma}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

69

## 誤差項の分散の推定

$$SD(\hat{\beta}_1|X_1 = x_1, \ldots, X_N = x_N) = \frac{\sigma}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

- これ、分母はデータから直で計算できますね。
- でも分子の$\sigma$は分りません、なので、このままでは最小二乗推定量$\hat{\beta}_1$の散らばりは分からない。
- 分からないなら$\sigma$を推定しちゃいましょう。

70

- $\sigma^2 = VAR(U|X) = E(U^2|X) - [E(U|X)]^2 = E(U^2|X)$
- 繰り返し期待値の法則より

$$E\left(E(U^2|X)\right) = E(U^2) \quad \rightarrow E(\sigma^2) = E(U^2) \rightarrow \sigma^2 = E(U^2)$$

- $\sigma^2 = E(U^2)$、これモーメント法（期待値のオペレーターを標本平均に変える）が使えますね。考えられる推定量の候補は

$$\frac{1}{N}\sum_{i=1}^{N} U_i^2$$

- でも$U_i$は誤差項、データとして存在しない。どうする？
- パラメーター推定後、残差$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$を計算、それを使えばいい。

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} \hat{U}_i^2$$

71

- この誤差項の分散の推定量

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} \hat{U}_i^2$$

一致性を持つが、不偏性は持たないことが知られている。

- そのために不偏性を持つ

$$\hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N} \hat{U}_i^2$$

が用いられることがある。

ノート：これはサンプルサイズがそれなりに大きければマイナーな問題です。

- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$は回帰直線の「**標準誤差**」と呼ばれます。

| Source | SS | df | MS |
|---|---|---|---|
| Model | 359065.326 | 1 | 359065.326 |
| Residual | 750369.903999998 | | .750371404 |
| Total | 1109435.23999999 | | 1.10943634 |

| | |
|---|---|
| Number of obs | = 1000000 |
| F( 1,999998) | = . |
| Prob > F | = 0.0000 |
| R-squared | = 0.3236 |
| Adj R-squared | = 0.3236 |
| Root MSE | = .86624 |

| y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| x | .5992904 | .0008663 | 691.75 | 0.000 | .5975924 | .6009884 |
| _cons | 2.503237 | .0027384 | 914.13 | 0.000 | 2.49787 | 2.508605 |

72

- やっとですが、最小二乗推定量$\hat{\beta}_1$の標準偏差に戻ってきました。

$$SD(\hat{\beta}_1|X_1 = x_1, \dots, X_N = x_N) = \frac{\sigma}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

- $\sigma$は分からないから、これはこのままでは使えない（最小二乗推定量$\hat{\beta}_1$の散らばりが分からない）それなら推定しようということで、$\hat{\sigma}$を推定しました。

- $\hat{\sigma}$を$\sigma$にぶちこんだもの、

$$\boxed{SE(\hat{\beta}_1|X_1 = x_1, \dots, X_N = x_N) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}}$$

を「$\hat{\beta}_1$**の標準誤差（standard error）**」と呼びます。これが最小二乗推定量$\hat{\beta}_1$の散らばりの指標です。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  359065.326     1  359065.326           Prob > F      =  0.0000
    Residual |  750369.903999998  .750371404           R-squared     =  0.3236
-------------+------------------------------           Adj R-squared =  0.3236
       Total |  1109435.23999999  1.10943634           Root MSE      =  .86624

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .5992904   .0008663   691.75   0.000     .5975924    .6009884
       _cons |   2.503237   .0027384   914.13   0.000      2.49787    2.508605
------------------------------------------------------------------------------
```

73