回帰分析 I

10. 多重回帰モデル

イントロダクション

- 単純線形回帰モデル：従属変数を一つの説明変数で説明。
- 最小二乗推定量が不偏性をもつためには$E(U|X)=0$が必要。だから$COV(X,U)=0$は当然成り立ってなければならないし、、、、
- 一致性を持つためにも$COV(X,U)=0$が必要。
- しかし、この仮定は成り立ちづらいかもしれない。
- 例を使って考えてみましょう。
- 講義ノート2で扱った例です。おさらいします。

- 米国の住宅金融機関の話です。
- 「住宅ローンの承認を判断する際、人種を考慮してはならない」、こういう法律があります。
- すなわち、ローンの応募者が二人いて、
  - ➢ 一人は白人、一人は黒人
  - ➢ ただし他の条件（例えば所得、職種、勤続年数など）はすべて同じ

なら、この二人には等しくローンが認められなければならないということ。

- 人種差別はだめ、ってこと。
- 次の調査報告があります。1990年代初期、、、
  - ➢ 黒人の応募者の28%が住宅ローンを承認されなかった（72%が承認された）。
  - ➢ 一方、白人の応募者で承認されなかったのは9%程度（91%が承認された）。
- この調査報告は、人種差別があったことを示している？

- 黒人の応募者の72%が住宅ローンを承認された。
- 白人の応募者の91%が住宅ローンを承認された。
- この結果を単回帰モデルの形で書いてみます。
- Yは承認されたときに1、承認されなかったときに0を取る従属変数。
  - ➢ 従属変数が離散変数なので奇妙に感じるかもしれません。
  - ➢ が、これは「線形確率モデル」と呼ばれるものでれっきとしたモデルです。
  - ➢ このモデルについてはコースの後半で取り扱います。
- X_1は白人なら1、黒人なら0を取る説明変数（白人ダミー）。
- 調査報告のデータを使って、Yを従属変数、X_1を説明変数とする単回帰モデルを最小二乗法を使って推定すると、、、
- β_0（定数項）の推定値として0.72, β_1（白人ダミーの係数）の推定値として0.19が得られます。

- 承認されるどうかに影響を与える要因は、もちろん白人か黒人か(X_1)だけではありません。
- 重要な要因の一つはもちろん所得(X_2)でしょうね。
- それでは母集団で次の関係が成り立っているとしましょう:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

- 話を単純にするために、$E(U|X_1, X_2) = 0$と仮定します。ですから$COV(X_1, U) = 0$そして$COV(X_2, U) = 0$です。
- ここで分析者が所得の影響を考慮せずに以下の単回帰モデルを推定したとしましょう。

$$Y = \beta_0 + \beta_1 X_1 + V$$

- このときの誤差項Vは、$V = \beta_2 X_2 + U$ということになりますね。
- 最小二乗推定量は一致性を持つでしょうか？

5

- 例の式を思い出しましょう。

$$plim\hat{\beta}_1 = \beta_1 + \frac{COV(X_1, V)}{VAR(X_1)}$$

- $COV(X_1, V)$の符号を考えればいいですね。
- $V = \beta_2 X_2 + U$ですから、X_1と$\beta_2 X_2 + U$の相関を考えればいいと。
- 仮定より$COV(X_1, U) = 0$なので、結局X_1と$\beta_2 X_2$の相関がどうなっているかが重要ですね。
- X_1(白人ダミー)とX_2(所得)の相関は、黒人にくらべて白人の方が所得が高い傾向にあるので、正だと考えられます。
- また、所得が高い方がローンの承認確率が高くなる傾向にあるでしょうから、$\beta_2 > 0$と考えられます。
- よってX_1と$\beta_2 X_2$は正に相関すると考えられ、$COV(X_1, V) > 0$と推測できますね。

6

- ということは、$plim\hat{\beta}_1 > \beta_1$、つまり最小二乗推定量は上に(漸近)バイアスする。
- これより、白人であるということが承認確率に与える影響を過大に評価しがちだ、と推測できます。
- 単回帰モデルを最小二乗法で推定すると、β_1(白人ダミーの係数)の推定値として0.19が得られますが、実際のβ_1(白人ダミーの係数)はこれより小さい、、、
- ゼロの可能性さえある、、、、ということになります。
- 従って、前述の調査報告は、人種差別があることのエビデンスとしてはかなり弱い、と結論づけることができます。
  - このような議論の仕方の「こつ」は後ほど説明します。
  - 少しのトレーニングでできるようになります。

7

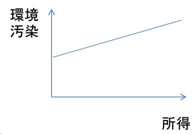
多重線形回帰モデル

- それでは、どうすればよいのでしょうか？
- 今考えている例では、$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$を推定すればいいですよね。
- 話を一般化します。
- 単回帰モデルでは、従属変数を説明する様々な要因が誤差項に含まれてしまい、それらが説明変数と相関してしまう、これが問題です。
- なら、従属変数を説明する要因をできる限りモデルに入れてしまおう、っていうのが**多重回帰モデル**です。
- そうすることによって、説明変数と誤差項が相関してしまうという問題の解決を図ります。

8

- また、多重回帰モデルは、変数間の関係の一般化に役立ちます。
- 例えば、所得（$income$）と環境汚染（$emission$）の国レベルでの関係について考えましょう。
- 単回帰モデル：

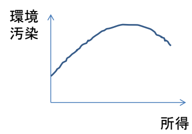
$$emission = \beta_0 + \beta_1 income + u$$



- 多重回帰モデル

$$emission = \beta_0 + \beta_1 income + \beta_2 income^2 + u$$

- 所得の2乗項をモデルに入れることで、右のような関係を考慮することが可能になります。



多重線形回帰モデルの重要な仮定

- 多重線形回帰モデル：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$$

- このモデルにおける重要な仮定：

$$E(U|X_1, \ldots, X_k) = 0$$

- この仮定が成り立つなら、「誤差項とすべての説明変数が相関しない」ということになります。すなわち、

$$COV(X_j, U) = 0, \quad j = 1, \ldots, k$$

は少なくとも成り立つ必要があります（十分ではありませんが）。

- $E(U|X_1, \ldots, X_k) = 0$は最小二乗推定量が不偏性を持つために必要な仮定です。
- 一致性のためには、これより弱い仮定である$COV(X_j, U) = 0,\ j = 1, \ldots, k$が成り立っていればいいです。
- コメント：単回帰モデルの時と同じですね、説明変数が複数あることを除いて。

最小二乗推定量

- 残差の二乗和を最小にするようにパラメータを選びます。
- 残差は$\widehat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}$で定義されます。
- 従って、以下の最小値問題を解くことになります：

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}\right)^2$$

- この最小値問題の一階の条件は、

$$\sum\nolimits_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

$$\sum\nolimits_{i=1}^{N} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum\nolimits_{i=1}^{N} X_{ki}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

- k ＋1本の連立方程式。
- 実際の計算では、$(Y_i, X_{1i}, \ldots X_{ki})$の所にデータ$(y_i, x_{1i}, \ldots x_{ki})$が入ります。
- 従って、この連立方程式における未知数は$(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$で$k+1$個。
- （ある条件の下で）一意に解けます。
- この「ある条件」については後ほど。

- モーメント法も同じ解を与えます。
- 仮定$E(U|X_1,\dots,X_k)=0$は、$E(U)=0, E(X_1U)=0, \cdots\cdots\cdots, E(X_kU)=0$を示します（講義ノート7の52ページ参照）。
- 従って、もし仮定が正しいなら、母集団において、

$E(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k)=0$
$E(X_1(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$
$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$
$E(X_k(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$

が成り立ちます。

- それでは 、サンプル$\{(Y_i,X_{1i},\dots,X_{ki}): i=1,\dots,N\}$においても、平均的にはこの関係が成り立つように$(\hat\beta_0,\hat\beta_1,\dots,\hat\beta_k)$を選びましょう。
- 期待値のオペレーター$E(\cdot)$を標本平均$\frac{1}{N}\sum_{i=1}^N(\cdot)$に変えればいいです。

13

- モーメント法も同じ解を与えます。
- 仮定$E(U|X_1,\dots,X_k)=0$は、$E(U)=0, E(X_1U)=0, \cdots\cdots\cdots, E(X_kU)=0$を示します（講義ノート7の52ページ参照）。
- 従って、もし仮定が正しいなら、母集団において、

$E(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k)=0$ $\quad \frac{1}{N}\sum_{i=1}^N\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$
$E(X_1(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$ $\quad \frac{1}{N}\sum_{i=1}^N X_{1i}\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$
$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$ $\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$
$E(X_k(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$ $\quad \frac{1}{N}\sum_{i=1}^N X_{ki}\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$

が成り立ちます。

- それでは 、サンプル$\{(Y_i,X_{1i},\dots,X_{ki}): i=1,\dots,N\}$においても、平均的にはこの関係が成り立つように$(\hat\beta_0,\hat\beta_1,\dots,\hat\beta_k)$を選びましょう。
- 期待値のオペレーター$E(\cdot)$を標本平均$\frac{1}{N}\sum_{i=1}^N(\cdot)$に変えればいいです。

14

- モーメント法も同じ解を与えます。
- 仮定$E(U|X_1,\dots,X_k)=0$は、$E(U)=0, E(X_1U)=0, \cdots\cdots\cdots, E(X_kU)=0$を示します（講義ノート7の52ページ参照）。
- 従って、もし仮定が正しいなら、母集団において、

$E(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k)=0$ $\quad \frac{1}{N}\sum_{i=1}^N\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$
$E(X_1(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$ $\quad \frac{1}{N}\sum_{i=1}^N X_{1i}\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$
$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$ $\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$
$E(X_k(Y-\beta_0-\beta_1X_1-\cdots-\beta_k X_k))=0$ $\quad \frac{1}{N}\sum_{i=1}^N X_{ki}\left(Y_i-\hat\beta_0-\hat\beta_1X_{1i}-\cdots-\hat\beta_kX_{ki}\right)=0$

が成り立ちます。

最小二乗法の一階の条件と同じ。
従って、モーメント法は最小二乗法と同じ推定値を与える。

- それでは 、サンプル$\{(Y_i,X_{1i},\dots,X_{ki}):$ り
立つように$(\hat\beta_0,\hat\beta_1,\dots,\hat\beta_k)$を選びましょ
- 期待値のオペレーター$E(\cdot)$を標本平均$\frac{1}{N}\sum_{i=1}^N(\cdot)$に変えればいいです。

15

OLS回帰直線の解釈

- 推定された式$\hat Y=\hat\beta_0+\hat\beta_1X_1+\cdots+\hat\beta_kX_k$を「OLS回帰曲線」と呼びます。
- これは母集団の回帰直線$E(Y|X)=\beta_0+\beta_1X_1+\cdots+\beta_kX_k$を推定したバージョンです。
- ここでは推定された式の解釈の仕方について説明します。
- ここでは単純化のため、説明変数を二つにしましょう。

$$\hat Y=\hat\beta_0+\hat\beta_1X_1+\hat\beta_2X_2$$

- X_1の変化をΔX_1、X_2の変化をΔX_2、そして$\hat Y$の変化を$\Delta\hat Y$と表します。このとき

$$\Delta\hat Y=\hat\beta_1\Delta X_1+\hat\beta_2\Delta X_2$$

という関係があります。ここでX_2を固定しましょう、すなわち$\Delta X_2=0$です。このとき、、、

16

- $\Delta\hat{Y} = \hat{\beta}_1 \Delta X_1$となります。

- 従って、X_1の係数の推定値$\hat{\beta}_1$は、

「X_2が一定の下で、X_1が一単位増加したときに$\hat{Y}$（Yの推計値）がどれだけ変化するか」

を表します。

- 大切なのは、X_2が一定の下で、というところです。

- 一般的には、ある説明変数の係数の推定値は、他の説明変数を一定にしたとして、変数の一単位の増加が$\hat{Y}$（Yの推計値）をどれだけ変化させるか」

を表します。

- また、$\Delta\hat{Y} = \hat{\beta}_1 \Delta X_1$は、$X_1$の係数の推定値は、他の説明変数の影響をコントロールした上で得られたものであることを示しています。

17

- 学術論文などで結果を説明するときは、

「X_1の係数の推定値は、X_1が一単位増加したときに、Yが平均すると$\hat{\beta}_1$だけ変化することを示している」

などとして、「他の説明変数が一定のとき」の部分を明示的に言わないこともあります。

- ちなみに、上の説明は、Yがレベル、X_1もレベルの時の説明の仕方です。

- Yが自然対数を取ったもの、例えば$Y = ln(A)$で、X_1がレベルなら、

「X_1の係数の推定値は、X_1が一単位増加したときに、Aが平均すると$100\hat{\beta}_1$パーセントだけ変化することを示している」

となります。

18

- また、Yが自然対数を取ったもの、例えば$Y = ln(A)$、X_1も自然対数を取ったもの、例えば$X_1 = ln(B)$なら、

「$ln(B)$の係数の推定値は、Bが1パーセント増加したとき、Aが平均すると$\hat{\beta}_1$パーセントだけ変化することを示している」

となります。

19

最小二乗推定量の不偏性

- 最小二乗推定量が不偏性を持つための仮定は以下の通り。

- 仮定1：母集団において以下の関係がある。

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$$

- 仮定2：母集団からのランダムサンプル$\{(X_{i1}, \dots, X_{ik}, Y_i): i = 1, \dots, N\}$がある。

- 仮定3：完全な多重共線性がない。

- 仮定4A：誤差項Uの条件付期待値はゼロ。すなわち、

$$E(U|X_1, \dots, X_k) = 0$$

20

最小二乗推定量の一致性

- 最小二乗推定量が一致性を持つための仮定は以下の通り。
- 仮定1：母集団において以下の関係がある。

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$$

- 仮定2：母集団からのランダムサンプル$\{(X_{i1}, \dots, X_{ik}, Y_i) : i = 1, \dots, N\}$がある。
- 仮定3：完全な多重共線性がない。
- 仮定4B：誤差項はいかなる説明変数とも相関なし。すなわち、

$$COV(X_j, U) = 0, j = 1, \dots, k$$

21

- 仮定4A、4Bは単回帰モデルの時の仮定の拡張です。
  - 説明変数が複数になったので、それに対応しています。
  - 特に目新しいことはありません。
- ここで目新しいものは仮定3ですね。
- 以下では仮定3について説明します。

「完全な多重共線性」

一つの説明変数が他の説明変数の完全な線形関数で表されるとき、その説明変数は**完全に多重共線的**である、または**完全な多重共線性**をもつ、と言います。

完全な多重共線性がある場合、最小二乗推定量を計算することはできません。

22

完全な多重共線性

- なぜ完全な多重共線性がある場合最小二乗推定量を計算することが不可能なのでしょうか？
- 完全な多重共線性がある場合の最も簡単な例を考えましょう。

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

$$X_2 = 2X_1$$

- 一つの説明変数X_2が、他の説明変数のX_1の倍になっているケースです。
  - 「完全な」線形関数ですね。
- この場合、「X_1を一定にしたとして、X_2の一単位の増加がYをどれだけ変化させるか」を考えるのはそもそもナンセンスです。
- なぜなら、$X_2 = 2X_1$なので、X_1を一定にすればX_2も一定になりますから。

23

- 完全な多重共線性がある場合に最小二乗推定量を計算するのが不可能であること、式を使って見てみましょう。
- 先ほど最小二乗法（モーメント法でも同じですが）が与えた条件を見てみましょう。
- 説明変数が2個の場合は以下の通りです。

$$\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_{i=1}^{N} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_{i=1}^{N} X_{2i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

ここでもし$X_{2i} = 2X_{1i}$なら、最後の式は

$$\sum_{i=1}^{N} 2X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

となり2番目の式と同じになります。未知数（パラメータ）は3個、式は2本、この連立方程式を解くことはできません。

24

完全な多重共線性の例

- それでは完全な多重共線性の例をいくつか見ていきましょう。
- 例1：X_1= 市に住む外国人比率　X_2=　市に住む外国人比率（%表示）
  - $X_2 = 100X_1$ですから完全な多重共線性を持ちます。
- 例2：X_1= 一日の労働時間　X_2= 一日の労働以外の時間
  - $X_1 = 24 - X_2$。一つの説明変数が他の説明変数の完全な線形関数。
  - よって完全な多重共線性を持つことになります。
- 例3：X_1= 一日の労働時間　X_2= 一日の睡眠時間　X_3= 一日の労働と睡眠以外の時間
  - $X_1 = 24 - X_2 - X_3$。一つの説明変数が他の説明変数の完全な線形関数。
  - よって完全な多重共線性を持つことになります。

- 例4: 人が住んでいる地域を3つのタイプに分類したとしましょう：「田園」、「郊外」、「都市」の3つです。

  $X_1 = 1$ 田園　　$X_2 = 1$ 郊外　　$X_3 = 1$ 都市
  $= 0$ 田園以外　　$= 0$ 郊外以外　　$= 0$ 都市以外

  - これは$X_1 + X_2 + X_3 = 1$となります。$X_1 = 1 - X_2 - X_3$とも表せますね。
  - 完全な多重共線性を持つことになります。
- 例5: これは少しトリッキーで気づきにくい例です。

  $X_1 = 1$ 学歴が高校卒業以上
  $= 0$ それ以外

ただし使っているデータセットには高卒未満の人はいなかった。

  - このときすべての人について$X_1 = 1$ となります。
  - これは定数項($X_0 = 1$)と全く同じになりますから、多重共線性が生じます。
  - データセットに高卒未満の人がいればこの問題は生じません。
  - 従って、完全な多重共線性は自分が使っているデータセットにも依存します。

完全な多重共線性の解決策

- 一つの説明変数が他の説明変数の完全な線形関数で表されるならば、そのうちのどれか一つの説明変数をモデルから抜くことで解決できます。
- どの変数を抜いても結果に本質的な違いはありません。
  - 例1、2であれば、X_1かX_2をモデルから抜けばいいです。
  - 例3、4であれば、X_1かX_2かX_3をモデルから抜けばいいです。
- 例5の場合は、X_1をモデルから抜かざるを得ません。
  - 高卒以上の人しかデータセットにいない場合、高卒以上であることの効果は推定することはできません。
- 最近の計量分析のソフトウェアの多くは、完全な多重共線性がある場合、自動的に変数を抜くことが多いです。

Stataを使って確認

- Stataは完全な多重共線性がある場合、自動的に変数を抜きます。確認しておきましょう。

- $N = 10000000$
- $Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差5
- $X_2 = 0.5{*}X_1$ （完全な多重共線性）
- $U\sim$ 正規分布　平均0 標準偏差50

これは何で7？

```
note: x2 omitted because of collinearity
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =       .
       Model |  1.2242e+09     1  1.2242e+09           Prob > F      =  0.0000
    Residual |  2.4994e+09999998  2499.42288           R-squared     =  0.3288
-------------+------------------------------           Adj R-squared =  0.3288
       Total |  3.7236e+09999999   3723.6127           Root MSE      =  49.994

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   6.999556   .0100015   699.85   0.000     6.979954    7.019159
          x2 |  (omitted)
       _cons |   4.930779   .0942868    52.30   0.000      4.74598    5.115578
------------------------------------------------------------------------------
```

- $Y = 5 + 2^*X_1 + 10^*X_2 + U$
- $X_2 = 0.5^*X_1$　これを上の式に代入。
- $Y = 5 + 2 * X_1 + 10 * (0.5X_1) + U$
  $= 5 + 2 * X_1 + 5 * X_1 + U$
  $= 5 + 7 * X_1 + U$
- これが推定すべきモデル。なので推定結果はおかしくないですね。

29

不完全な多重共線性

- 「完全な多重共線性」と「不完全な多重共線性」、名前は似ていますが概念的には全く異なるものです。
- 「**不完全な多重共線性**」とは、2つかそれ以上の説明変数が高い相関を示す場合を言います。
- すなわち、2つかそれ以上の説明変数間に完全ではない線形関係があり、相関が高いような状況です。
- 「不完全な多重共線性」があったとしても、先ほど挙げた仮定のいずれにも反しません。
- 従って、先ほどの仮定1－4が満たされていれば、「不完全な多重共線性」があったとしても、最小二乗推定量は不偏性・一致性を持つことになります。
- 「不完全な多重共線性」があるときに何が起こるかについては、後ほど説明します。

30

欠落変数バイアス（重要）

- 大切な変数がモデルの中から抜け落ちてしまったことから生じる推定量のバイアスを「**欠落変数バイアス**（omitted variable bias）」といいます。
  - 「除外変数バイアス」や
  - 「過小定式化により生じるバイアス」などともいいます。
- ここで導出するルールを覚えると、推定量がバイアスしていると考えられる場合、その方向を簡単に推測できます。
- ここでは話を簡単にするために、以下の説明変数が2つのモデルを考えましょう。

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U,$$
$$E(U) = 0, COV(X_1, U) = 0, COV(X_2, 0) = 0$$

- このモデルは、誤差項と説明変数が相関していませんから、最小二乗推定量は一致性を持ちます（漸近バイアス無し）。

31

- ここで以下の2つのケースを考えましょう。
- ケース1：分析者が、X_2がYに与える可能性を思いつかなかった。
  - これは残念なケースです。
  - こういうことが起きないように先行研究をよく見てみましょう。
- ケース2：X_2がデータとして存在しない。
  - これは実証分析でよく起こります。
  - 例えば、賃金を説明するモデルを考えます。
  - 「生まれ持った能力」は賃金に影響を与える可能性がありますね。
  - しかし、そのような変数はデータとして存在しないのが普通です。
  - 企業の何らかのパフォーマンスを説明するモデルを考えます。
  - 「社風」や「社長の性格」は経営判断に影響を与えるかもしれません。
  - 従って企業のパフォーマンスにもです。
  - しかし、そのような変数はデータとして存在しないのが普通です。

32

- 結果として、分析者がY とX_1だけからなる単純線形回帰モデルを最小二乗法を使い推定するとしましょう。
- このときの最小二乗推定量を$(\tilde{\beta}_0, \tilde{\beta}_1)$とします。
- 設定のまとめ：

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$
$E(U) = 0, COV(X_1, U) = 0, COV(X_2, 0) = 0$
最小二乗推定量$(\hat{\beta}_0, \hat{\beta}_1)$はバイアス無し

ただし分析者は

$$Y = \beta_0 + \beta_1 X_1 + V$$

を最小二乗法で推定：最小二乗推定量$(\tilde{\beta}_0, \tilde{\beta}_1)$

Question

このとき最小二乗推定量$(\tilde{\beta}_0, \tilde{\beta}_1)$は一致性を持つ？

- Answer: 持つ。<u>ただし</u>

1. $\beta_2 = 0$ か
2. $COV(X_1, X_2) = 0$

が成り立っている時<u>だけ</u>。

- 逆に言うと、

$\beta_2 \neq 0$かつ$COV(X_1, X_2) \neq 0$ならば、最小二乗推定量は一致性を持たない

コメント：$\beta_2 = 0$ であるならX_2はYに影響を与えないのであるから、そもそもモデルにX_2を入れる必要はない。入れなくてもバイアスが生じないのは当たり前っていえば当たり前。

証明：$Y = \beta_0 + \beta_1 X_1 + V$の最小二乗推定量$\tilde{\beta}_1$は（講義ノート8参照）、

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)Y_i}{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

分子を考えましょう。

$$\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)Y_i = \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{U}_i)$$

$$= \hat{\beta}_0 \sum_{i=1}^{N}(X_{i1} - \bar{X}_1) + \hat{\beta}_1 \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i1} + \hat{\beta}_2 \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i2} + \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)\hat{U}_i$$

第一項はゼロ、$\sum_{i=1}^{N}(X_{i1} - \bar{X}_1) = 0$だから（講義ノート3参照）。

第四項もゼロ。これは、標本のモーメント条件$\sum_{i=1}^{N} \hat{U}_i = 0$と$\sum_{i=1}^{N} X_{i1}\hat{U}_i = 0$を使えば簡単に示せる。

第二項に$\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i1} = \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2$（講義ノート3参照）、

第三項に$\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i2} = \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)$（講義ノート3参照）を使うと、$\tilde{\beta}_1$は

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

となることが示せる。これは

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

と書き直すことができる。両辺に*plim*をとると

$$plim\tilde{\beta}_1 = plim\left(\hat{\beta}_1 + \hat{\beta}_2 \frac{\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}\right)$$

$$plim\tilde{\beta}_1 = plim\hat{\beta}_1 + plim\hat{\beta}_2 \frac{plim\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{plim\frac{1}{N}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

$$plim\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{COV(X_1, X_2)}{VAR(X_1)}$$

従って、$plim\tilde{\beta}_1 = \beta_1$と成るためには、$\beta_2 = 0$か$COV(X_1, X_2) = 0$が必要。

$$plim\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{COV(X_1, X_2)}{VAR(X_1)}$$

- この式から以下の**超重要ルール**（これ暗記）が導かれる：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$
$$E(U) = 0, COV(X_1, U) = 0, COV(X_2, 0) = 0$$

ただし分析者は

$$Y = \beta_0 + \beta_1 X_1 + V$$

を最小二乗法で推定：そのときの推定量$\tilde{\beta}_1$の（漸近）バイアスのパターンは以下の通り。

	$COV(X_1, X_2) > 0$	$COV(X_1, X_2) < 0$
$\beta_2 > 0$	上にバイアス	下にバイアス
$\beta_2 < 0$	下にバイアス	上にバイアス

37

- このテーブルはとても重要なテーブルです。
  - このコースで学んだことは将来ほとんど忘れてしまうと思います。
  - でもこのテーブルだけは覚えていた方が良いです。
  - なぜなら、このテーブルをうまく使うと、他の人による統計を使った議論の妥当性を瞬時に判断できる場合があるからです。
  - 特に欠落変数があると考えられるような場合です。
  - 基本的には同じパターンが不偏性についても成り立ちます。
- 以下ではこのテーブルの使い方について説明します。

38

	$COV(X_1, X_2) > 0$	$COV(X_1, X_2) < 0$
$\beta_2 > 0$	上にバイアス	下にバイアス
$\beta_2 < 0$	下にバイアス	上にバイアス

- 最小二乗法による結果があるとします。
- ステップ0：分析の興味の対象である説明変数をX_1とする。
- ステップ1：モデルの中に入っていないけれども、従属変数に影響を与えるだろうと思われる変数（X_2）を考える。
- ステップ2：その変数が従属変数へ与える影響（β_2）はプラス（$\beta_2 > 0$）なのか、マイナス（$\beta_2 < 0$）なのか考える。
- ステップ3：X_1とX_2の相関（$COV(X_1, X_2)$）はプラス（$COV(X_1, X_2) > 0$）なのか、マイナスなのか（$COV(X_1, X_2) < 0$）を考える。
- ステップ4：テーブルを使いバイアスの方向を判断する。

39

テーブルの使い方：例1

- Yはローンが承認されたときに1、承認されなかったときに0を取る従属変数。
- X_1は白人なら1、黒人なら0を取る説明変数（白人ダミー）。
- 推定モデル：$Y = \beta_0 + \beta_1 X_1 + V$ 最小二乗法を使って推定
- ステップ0：分析の興味の対象である説明変数はX_1。
- ステップ1：モデルの中に入っていないけれども、従属変数に影響を与えるだろうと思われる変数（X_2）は所得。
- ステップ2：その変数が従属変数へ与える影響（β_2）はプラス（$\beta_2 > 0$）。
- ステップ3：X_1とX_2の相関（$COV(X_1, X_2)$）はプラス（$COV(X_1, X_2) > 0$）。
- ステップ4：テーブルより最小二乗推定量$\tilde{\beta}_1$は<u>上に</u>バイアスする。
- 従って最小二乗法によるβ_1の推定値は、白人であるということがローンの承認確率に与える影響（β_1）を<u>過大に評価</u>している可能性が高い、と判断する。

40

- このルールは、今の例のように、二つのグループの平均の差を使って因果関係の存在を議論している場合に特に使えます。
- 例：読書している学生は、していない学生より、テストのスコアが平均すると10点良かった。従って、読書は成績の向上に役立つ。
- この結果は、以下の単回帰モデル

Y:テストのスコア
X_1: ＝ 1（読書している） ＝0（読書していない）
$Y = \beta_0 + \beta_1 X_1 + V$

を最小二乗法で推定、β_1の推定値が10であった、ということと同じと考えられます。

41

テーブルの使い方：例2

- Y：共通テストにおける生徒の平均点
- X_1: 先生一人に対しての生徒数
- 推定モデル：$Y = \beta_0 + \beta_1 X_1 + V$　最小二乗法を使って推定

- ステップ0：分析の興味の対象である説明変数はX_1。
- ステップ1：モデルの中に入っていないけれども、従属変数に影響を与えるだろうと思われる変数（X_2）は生徒の親の所得。
- ステップ2：その変数が従属変数へ与える影響（β_2）はプラス（$\beta_2 > 0$）。
- ステップ3：X_1とX_2の相関（$COV(X_1, X_2)$）はマイナス（$COV(X_1, X_2) < 0$）。
- ステップ4：テーブルより最小二乗推定量$\tilde{\beta}_1$は下にバイアスする。
- 従って最小二乗法によるβ_1の推定値は、クラスサイズがテストのスコアに与える（あるとすれば負の）影響を過剰に評価している可能性が高い、と判断する。

42

シミュレーション1：$COV(X_1, X_2) = 0$

- 体感してみましょう。
- $N = 10000000$
- $Y = 5 + 2^* X_1 + 10^* X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- ここではX_1とX_2を独立に発生させているので$COV(X_1, X_2) = 0$です。
- U~ 正規分布　平均0 標準偏差5
- それでは説明変数がX_1だけのモデルを最小二乗法で推定します。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) = 2393.06
       Model |    61364280     1    61364280           Prob > F      =  0.0000
    Residual |  2.5643e+10999998   25642.612           R-squared     =  0.0024
-------------+------------------------------           Adj R-squared =  0.0024
       Total |  2.5704e+10999999  25703.9507           Root MSE      =  160.13

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.958907    .040044    48.92   0.000     1.880422    2.037391
       _cons |   25.28178   .3579414    70.63   0.000     24.58022    25.98333
------------------------------------------------------------------------------
```

- Q：定数項は25と推定されました。なぜ？

43

シミュレーション2：$COV(X_1, X_2) = 16, VAR(X_1) = 16$

- 体感してみましょう。
- $N = 10000000$

$$plim\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{COV(X_1, X_2)}{VAR(X_1)}$$

- $Y = 5 + 2^* X_1 + 10^* X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- ただし$CORR(X_1, X_2) = 0.25$ なので$COV(X_1, X_2) = 16$
- U~ 正規分布　平均0 標準偏差5
- それでは説明変数がX_1だけのモデルを最小二乗法で推定します。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =96398.89
       Model |  2.3086e+09     1  2.3086e+09           Prob > F      =  0.0000
    Residual |  2.3948e+10999998   23948.268           R-squared     =  0.0879
-------------+------------------------------           Adj R-squared =  0.0879
       Total |  2.6257e+10999999  26256.8328           Root MSE      =  154.75

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   12.00897   .0386785   310.48   0.000     11.93316    12.08478
       _cons |  -54.81944   .3461456  -158.37   0.000    -55.49787     -54.141
------------------------------------------------------------------------------
```

- X_1の効果の推定値は本物の6倍！に（上にバイアス）。

44

シミュレーション3 : $COV(X_1, X_2) = -16, VAR(X_1) = 16$

- 体感してみましょう。
- $N = 10000000$

$$plim\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{COV(X_1, X_2)}{VAR(X_1)}$$

- $Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差4
- $X_2 \sim$ 正規分布　平均2　標準偏差16
- ただし$CORR(X_1, X_2) = -0.25$ なので$COV(X_1, X_2) = -16$
- $U \sim$ 正規分布　平均0 標準偏差5
- それでは説明変数がX_1だけのモデルを最小二乗法で推定します。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  1,999998) =42684.04
       Model |  1.0222e+09     1  1.0222e+09           Prob > F      =  0.0000
    Residual |  2.3948e+10999998   23948.268           R-squared     =  0.0409
-------------+------------------------------           Adj R-squared =  0.0409
       Total |  2.4970e+10999999  24970.4539           Root MSE      =  154.75

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -7.991027   .0386785  -206.60   0.000    -8.066836   -7.915219
       _cons |   105.1806   .3461456   303.86   0.000     104.5021     105.859
------------------------------------------------------------------------------
```

- X_1の効果の推定値はかなり大きいマイナスに（下にバイアス）。

テーブル使用上の注意点

1. 説明変数が多数ある時は、欠落変数バイアスの方向はここで説明したのほど単純ではありません。

   - そのような場合でも、2変数の場合のテーブルを使って、バイアスの方向を考えるのが一般的です。

2. このテーブルは、欠落変数バイアスに特化したものです。

   - バイアスは欠落変数以外の理由でも起こります。
   - $plim\hat{\beta}_1 = \beta_1 + \frac{COV(X_1, V)}{VAR(X_1)}$は欠落変数が原因のバイアスに加えて、それ以外の理由で生じるバイアスの時にも使えます。

3. バイアスが生じている可能性があるとき、得られた結果が全く使えないかというと、必ずしもそうではありません。

- 次の例を考えましょう。

- 自分の仮説が「XはYを増やす」だとします。
- 最小二乗法を使い得られたXの係数は正（例えば0.5）でした。
- しかしバイアスが生じていそうです。
- ただしバイアスの方向は下である可能性が高いとします。
  - 推定された係数の値は、真の値より「下に」ある可能性が高いということ。
- このとき真の値は0.5より大きい可能性が高いわけですから（例えば0.8）、推定結果は「XはYを増やす」という仮説と整合的であると一応主張できます。
  - しかし、仮説をサポートするエビデンスとしては弱いものとなります。
  - また、効果の「大きさ」に関しては何も言えません。
  - またバイアスの理屈を知らない人に言っても通じないかもしれません。

4. 欠落変数バイアスが—、、、といって、なんでもかんでも他の人の分析結果を批判するのはあまり健全ではありません。

- 例えば、ある分析者の興味対象が、X_1とYの関係だったとします。
- この分析者は、モデルの中に説明変数として$X_1, \ldots, X_{10}$を入れて最小二乗法で推定していました。
- その分析結果に対して、
  - これこれこういう変数がモデルから落ちている。
  - よって、X_{10}の係数の最小二乗推定量は上にバイアスしている可能性が高い。
  - X_{10}の係数の推定値は過大に評価されているのでは？

という批判をするのは本質的ではないということです、X_{10}は分析者にとって興味の対象ど真ん中というわけではないのですから。

モデルに不必要な変数を入れると？

- モデルから重要な変数を落とすと最小二乗推定量はバイアスすることが分かりました。
- それでは、不必要な変数を入れた場合はどうなるでしょうか？
- 分析者が

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U, E(U|X_1, X_2, X_3) = 0$$

を推定したとしましょう。ただしX_3はYとは関係ありません。すなわち$\beta_3 = 0$です。

- このとき最小二乗推定量はバイアスする？

49

モデルに不必要な変数を入れると？

- 答えは、最小二乗推定量は<u>バイアスしない</u>です。
- 不偏性・一致性ともに持ちます。
- 仮定1－4Aさえ成り立っていれば、β_jがいかなる値であっても最小二乗推定量は不偏性を持ちます。すなわち$E(\hat{\beta}_3) = 0$です。
  - ➢ 仮定1－4Aにおいて、β_jの値は特定されていません。
  - ➢ 従って、$\beta_j = 0$の場合も含みます。
- 一致性についても同様です。
- シミュレーションしてみましょう（一致性だけやります）。

50

- $N = 10000000$
- $Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- X_3 ~ 正規分布　平均10　標準偏差16
- U~ 正規分布　平均0 標準偏差5
- それでは説明変数がX_1, X_2, X_3からなるモデルを最小二乗法で推定します。

Source	SS	df	MS
Model	2.5679e+10	3	8.5596e+09
Residual	24994167.1	999996	24.9942671
Total	2.5704e+10	999999	25703.9507

Number of obs	= 1000000
F(3,999996)	= .
Prob > F	= 0.0000
R-squared	= 0.9990
Adj R-squared	= 0.9990
Root MSE	= 4.9994

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	1.999945	.0012502	1599.71	0.000	1.997494	2.002395
x2	9.999948	.0003124	3.2e+04	0.000	9.999336	10.00056
x3	-.0002075	.0003124	-0.66	0.507	-.0008198	.0004048
_cons	4.995345	.0116203	429.88	0.000	4.97257	5.018121

51

- いまは説明変数間に相関が無いケースでした。説明変数間に相関があるケースはどうでしょうか？
- $N = 10000000$
- $Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- X_3 ~ 正規分布　平均10　標準偏差16
- ただしX_2とX_3は相関している。相関係数0.5
- U~ 正規分布　平均0 標準偏差5
- それでは説明変数がX_1, X_2, X_3からなるモデルを最小二乗法で推定します。

Source	SS	df	MS
Model	2.5679e+10	3	8.5596e+09
Residual	24994167.1	999996	24.9942671
Total	2.5704e+10	999999	25703.9507

Number of obs	= 1000000
F(3,999996)	= .
Prob > F	= 0.0000
R-squared	= 0.9990
Adj R-squared	= 0.9990
Root MSE	= 4.9994

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	1.999945	.0012502	1599.71	0.000	1.997494	2.002395
x2	10.00007	.0003605	2.8e+04	0.000	9.999361	10.00077
x3	-.0002396	.0003607	-0.66	0.507	-.0009466	.0004674
_cons	4.995426	.0116539	428.65	0.000	4.972585	5.018268

52

モデルに不必要な変数を入れると？

- それでは、不必要な変数をモデルに入れることに害はないのでしょうか？
- 答えは、害はある、です。
- 特殊なケースを除いて、最小二乗推定量の分散が大きくなります。
- 理由はのちほど。

最小二乗推定量の分散

- ここでは最小二乗推定量の分散について考えていきます。
- 単回帰モデルと同様、誤差項の分散に関して、均一分散（講義ノート9参照）の仮定をおきます。
- 仮定5：説明変数がいかなる値でも、誤差項の分散は一定。すなわち、

$$VAR(U|X_1, \ldots, X_k) = \sigma^2$$

- このとき仮定1ー5の下で以下が成り立ちます（証明無しで結果だけ）。

$$VAR(\hat{\beta}_j|X_1 = x_1, \ldots, X_k = x_k) = \frac{\sigma^2}{(1 - R_j)\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}$$

ただしR_jは従属変数がX_jで説明変数が$1, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$である回帰モデルの決定係数。

$$VAR(\hat{\beta}_j|X_1 = x_1, \ldots, X_k = x_k) = \frac{\sigma^2}{(1 - R_j)\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}$$

- 単回帰モデルのときの最小二乗推定量の分散に基本的には似ています。
- ここでは、まずこの式が示すことを考えていき、それを体感するためにシミュレーションすることにします。
- 大切なことは3つ。

1. $\sigma^2 \uparrow$ → $\uparrow VAR(\beta_j)$
2. X_jの散らばり ↑ → $\downarrow VAR(\beta_j)$
3. R_j ↑ → $\uparrow VAR(\beta_j)$

とりあえずシミュレーションすることにします。

それにより、そもそも$VAR(\hat{\beta}_j|X_1 = x_1, \ldots, X_k = x_k)$が何なのかが分かります。

- N人いるとします。
- それぞれの人に、乱数を発生させ、X_1, X_2 を作ってあげます。
  - ➢ $x_{i1}, \ldots, x_{N1}$； $x_{i2}, \ldots, x_{N2}$
- これは固定します。

- それぞれの人に、乱数を発生させ、Uを作ってあげます。
  - ➢ $u_1, \ldots, u_N$
- $y_i = \beta_0 + \beta_1 x_{i1} + + \beta_2 x_{i2} + u_i$をそれぞれの人に計算します。（$\beta_0, \beta_1, \beta_2$ ）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_{i1}, x_{i2}）ができます。
- N人分の（y_i, x_{i1}, x_{i2}）をデータとして使い、真の値（$\beta_0, \beta_1, \beta_2$ ）を知らない体で、モデル$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$を最小二乗法で推定します。
- 最小二乗推定値（$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$）が得られます。ここでは$\hat{\beta}_2$にフォーカスします。
- またそれぞれの人に乱数を発生させ、Uを作ってあげます、、、、
- これをM回繰り返します → $\hat{\beta}_2^{(1)}, \hat{\beta}_2^{(2)}, \ldots, \hat{\beta}_2^{(M)}$が得られます。

- この$\hat{\beta}_2^{(1)}, \hat{\beta}_2^{(2)}, \ldots, \hat{\beta}_2^{(M)}$をヒストグラムにします。
- （Mが十分に大きければ）これが$X_1 = x_1, X_2 = x_2$を条件にした時の、最小二乗推定量$\hat{\beta}_2$の分布です。
- $VAR(\hat{\beta}_j | X_1 = x_1, X_2 = x_2)$はこの分布の散らばりということになります。
- $E(\hat{\beta}_2 | X_1 = x_1, X_2 = x_2) = \beta_2$だとすると、最小二乗推定量は平均的には真の値になるってこと。でも、これはあくまで平均的にはって話。
- このとき$VAR(\hat{\beta}_j | X_1 = x_1, X_2 = x_2)$が小さければ、推定値は真の値の周りに出やすいってことだし、、、
- $VAR(\hat{\beta}_j | X_1 = x_1, X_2 = x_2)$が大きいなら、推定値が真の値から離れて出ることが珍しくなくなります。

シミュレーション

- $\sigma^2 \uparrow \rightarrow \uparrow VAR(\hat{\beta}_j | X_1 = x_1, X_2 = x_2)$を体感してみましょう。

実験1
$N = 200$
$Y = 5 + 2^* X_1 + 10^* X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U ~ 正規分布　平均0 標準偏差0.01
最小二乗法で推定。10000回繰り返す。

実験2
$N = 200$
$Y = 5 + 2^* X_1 + 10^* X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U ~ 正規分布　平均0 標準偏差1
最小二乗法で推定。10000回繰り返す。

実験3
$N = 200$
$Y = 5 + 2^* X_1 + 10^* X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U ~ 正規分布　平均0 標準偏差100
最小二乗法で推定。10000回繰り返す。

実験4
$N = 200$
$Y = 5 + 2^* X_1 + 10^* X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U ~ 正規分布　平均0 標準偏差1000
最小二乗法で推定。10000回繰り返す。

$Y = 5 + 2^* X_1 + 10^* X_2 + U$, U ~ 正規分布　平均0 標準偏差σ

最小二乗推定量$\hat{\beta}_2$の分布

- 実験1（σ=0.01）

Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	10000	10	.0000465	9.999817	10.00062

- 実験2（σ=1）

Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	10000	9.999979	.0046507	9.981714	10.0625

- 実験3（σ=100）

Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	10000	9.997938	.4650672	8.171398	16.25

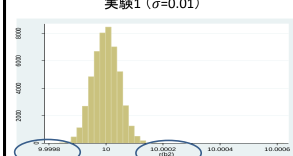
- 実験4（σ=1000）

Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	10000	9.979378	4.650672	-8.286018	72.5

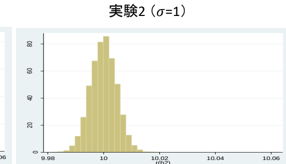
$Y = 5 + 2^* X_1 + 10^* X_2 + U$, U ~ 正規分布　平均0 標準偏差σ

最小二乗推定量$\hat{\beta}_2$の分布

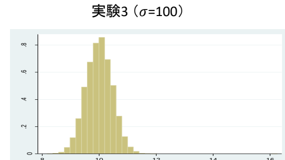
実験1（σ=0.01）



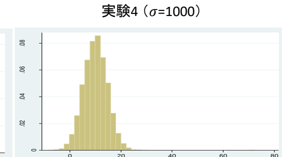
実験2（σ=1）



実験3（σ=100）

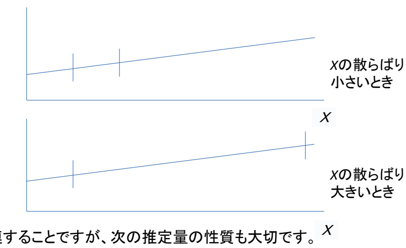


実験4（σ=1000）



- すべてのケースで最小二乗推定量$\hat{\beta}_2$の分布の平均はほぼ10で当たり。
- これは説明変数と誤差項を独立に発生させているため。
  - $E(U|X) = 0$が成り立つ。
  - 最小二乗推定量は不偏性を持つ。
- ただし最小二乗推定量$\hat{\beta}_2$のちらばりは、誤差項の分散に依存している。
- 誤差項の分散が小さい（σ=0.01）ときは、外れたたとしてもほぼ10の周り。
  - 一番大きいときでも10.0006。
  - 一番小さいときでも9.9998。
- 誤差項の分散が大きい（σ=1000）ときは、かなりぶれる。
  - 一番大きいときは72.5
  - 一番小さいときは-8（符号が逆）

- X_jの散らばり ↑→↓$VAR(\hat{\beta}_j)$
- データにおいて説明変数の散らばりが大きければ大きいほど、最小二乗推定量の精度が高くなる、ということです。理由は以下の絵を使って授業で。



- これに関連することですが、次の推定量の性質も大切です。

$$VAR(\hat{\beta}_j|X_1, \dots, X_k) = \frac{\sigma^2}{(1-R_j)\sum_{i=1}^{N}(X_{ij}-\bar{X}_j)^2}$$

- $\sum_{i=1}^{N}(X_{ij}-\bar{X}_j)^2$が大きくなれば推定量の分散は小さくなる。
- ここで大切なのはNで割っていないこと。
- 従って、$\sum_{i=1}^{N}(X_{ij}-\bar{X}_j)^2$は$N$（サンプルサイズ）が大きくなれば自動的に大きくなる。
- ってことは、次のことが言えます。

サンプルサイズが大きくなれば、最小二乗推定量の散らばりは小さくなっていく。

- サンプルサイズが大きくなるにつれ推定量の「精度」が高まっていく、ということですね。

シミュレーション

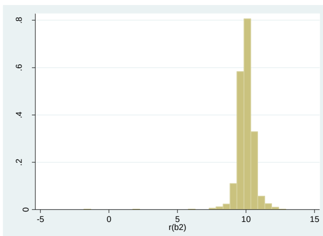
- サンプルサイズ(N)↑→↓$VAR(\hat{\beta}_j)$を体感してみましょう。

実験1
$N = 5$
$Y = 5 + 2^*X_1 + 10^*X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U~ 正規分布　平均0 標準偏差10
最小二乗法で推定。1000回繰り返す。

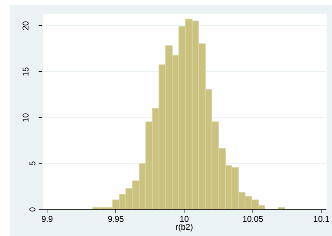
実験2
$N = 1000$
$Y = 5 + 2^*X_1 + 10^*X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U~ 正規分布　平均0 標準偏差10
最小二乗法で推定。1000回繰り返す。

実験3
$N = 50000$
$Y = 5 + 2^*X_1 + 10^*X_2 + U$
X_1 ~ 正規分布　平均8 標準偏差4
X_2 ~ 正規分布　平均2　標準偏差16
U~ 正規分布　平均0 標準偏差10
最小二乗法で推定。1000回繰り返す。

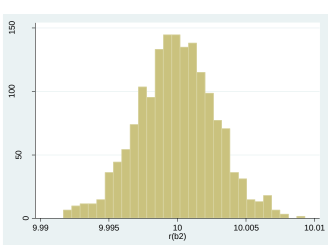
$N = 5$



$N = 1000$



$N = 50000$



$$VAR(\hat{\beta}_j | X_1, \ldots, X_k) = \frac{\sigma^2}{(1 - R_j)\sum_{i=1}^{N}(X_{ij} - \bar{X}_j)^2}$$

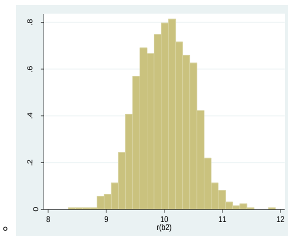
- $R_j \uparrow \rightarrow \uparrow VAR(\hat{\beta}_j)$
- ただしR_jは従属変数がX_jで説明変数が$1, X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$である回帰モデルの決定係数。
- これはどういうことかというと、説明変数間の相関の話です。
- 単純化のために説明変数を二つにして考えます。
- 従属変数がX_2、説明変数が1とX_1である回帰モデル（別にこのモデルには興味はないです）の決定係数が高いとは、X_1とX_2の相関が高いことを意味します。
- 上の式が示しているのは、X_1とX_2の相関が高いとき最小二乗推定量$\hat{\beta}_2$の散らばりが大きくなる（すなわち精度が低くなる）ことを示しています。

- これはどういうことでしょうか？
- 直感的には次のように考えればいいです。
- X_1が動くとYも動く。ここからX_1のYに与える影響を推測しようとします。
- しかしX_1とX_2の動きがとても似ている場合（正の相関大きい）、Yの変化がX_1によってもたらされたものなのか、X_2によってもたらされたものなのか判別することが難しくなりますよね。
- これにより推定量の精度が低くなる、のはなんとなく想像できると思います。
- X_1とX_2が反対に動く傾向がある場合（負の相関大きい）でも、同じことが言えます。

シミュレーション

- 説明変数は二つ、X_1とX_2
- $N = 200$
- モデル：$Y = 5 + 2*X_1 + 10*X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- ただしX_1とX_2は独立（相関無し）
- U~ 正規分布　平均0 標準偏差100
- 最小二乗推定量$\hat{\beta}_2$の分布（$M = 1000$）。

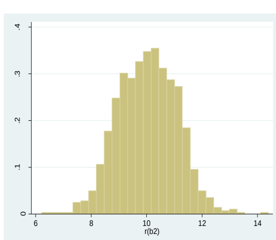
Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	1000	10.01828	.4611893	8.351357	11.9159



シミュレーション

- 説明変数は二つ、X_1とX_2
- $N = 200$
- モデル：$Y = 5 + 2*X_1 + 10*X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- ただしX_1とX_2は相関あり。
- 相関係数0.9
- U~ 正規分布　平均0 標準偏差100
- 最小二乗推定量$\hat{\beta}_2$の分布（$M = 1000$）。

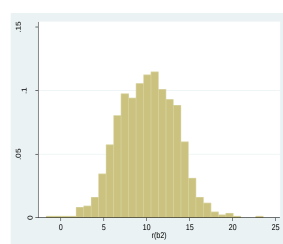
Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	1000	10.04193	1.058041	6.217753	14.39537



シミュレーション

- 説明変数は二つ、X_1とX_2
- $N = 200$
- モデル：$Y = 5 + 2*X_1 + 10*X_2 + U$
- X_1 ~ 正規分布　平均8 標準偏差4
- X_2 ~ 正規分布　平均2　標準偏差16
- ただしX_1とX_2は相関あり。
- 相関係数0.99!
- U~ 正規分布　平均0 標準偏差100
- 最小二乗推定量$\hat{\beta}_2$の分布（$M = 1000$）。

Variable	Obs	Mean	Std. Dev.	Min	Max
b2f	1000	10.12956	3.269285	-1.686924	23.58142



- 30ページで触れた「**不完全な多重共線性**」とは、2つかそれ以上の説明変数間に完全ではない線形関係があり、相関が高いような状況です。
- 今見たように、不完全な多重共線性がある場合は最小二乗推定量の分散が大きくなります。
- 従って、不完全な多重共線性があることは分析者にとって望ましことではありません、推定量の精度が落ちるわけですから。
- ただし、「不完全な多重共線性」があったとしても、<u>先ほど挙げた仮定のいずれにも反しません</u>。
- 従って、「不完全な多重共線性」があったとしても、最小二乗推定量は不偏性・一致性を持つことになります。
  - ➢ シミュレーションで不偏性が成り立つのは見ました。
  - ➢ 念のため一致性も見ておきましょう。

シミュレーション

- 説明変数は二つ、X_1とX_2
- $N = 1000000$（一致性のシミュレーションです）
- モデル: $Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差4
- $X_2 \sim$ 正規分布　平均2　標準偏差16
- ただしX_1とX_2は相関あり。相関係数0.99!
- $U \sim$ 正規分布　平均0 標準偏差100
- 最小二乗法による推定結果

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  2,999997) =       .
       Model |  2.8130e+10     2  1.4065e+10           Prob > F      =  0.0000
    Residual |  1.0006e+10999997  10006.1558           R-squared     =  0.7376
-------------+------------------------------           Adj R-squared =  0.7376
       Total |  3.8136e+10999999  38135.7603           Root MSE      =  100.03

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.035064   .1772203    11.48   0.000     1.687718    2.382409
          x2 |   9.978366   .0443091   225.20   0.000     9.891522    10.06521
       _cons |   4.879367   1.333933     3.66   0.000     2.264904     7.49383
------------------------------------------------------------------------------
```


「**不完全な多重共線性**」に関する注意

- 多重共線性を英語ではmulticollinearity（マルチコリニアリティ、略してマルチコ）と言います。
- 通常「マルチコがある」と誰かが言った場合、それは不完全な多重共線性のことを意味します。
- 「マルチコ」は最小二乗推定量の分散を大きくする（推定量の精度が下がる）ことが知られているため、「マルチコがー、マルチコがー、、、」と言う人をよく見かけます。
  - ➢ 多分マルチコって、言いたいだけです。
- たしかに「マルチコ」は最小二乗推定量の精度を下げます。しかし、、、

- 推定量の精度に影響を与えるのは「マルチコ」だけではありません。
- 例えば、先ほど見たように、サンプルサイズ（N）が大きければ推定量の精度は上がります。
- 従って、仮に「マルチコ」があったとしても、十分に大きなデータがあれば、推定量の精度はそれほど悪くないかもしれません。
- 一方、「マルチコ」があって、かつサンプルサイズ（N）が小さいなら、推定量の精度は良くないでしょう。
- 残念なことに、どの程度の相関が説明変数間にあると問題になるのかについて明確な定義はありません。
  - ➢ それはそうです、なぜならサンプルサイズとの兼ね合いになるからです。
  - ➢ さっきみましたよね、相関係数が0.99！でも、サンプルサイズが大きいならほぼドンピシャで当ててました（一致性ありますから）。

- 人によっては、説明変数間の相関が高いときに、その中の一つを分析から落とすことをアドバイスしてきます。
- 残念ながら、このアドバイスは正しくないかもしれません。
- 欠落変数バイアスが生じるかもしれないからです。
- 見てみましょう。

77

- $N = 1000000$（一致性のシミュレーションです）
- モデル：$Y = 5 + 2{*}X_1 + 10{*}X_2 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差4
- $X_2 \sim$ 正規分布　平均2　標準偏差16
- ただしX_1とX_2は相関あり。相関係数0.99!
- $U \sim$ 正規分布　平均0 標準偏差100
- $Y = 5 + 2{*}X_1 + V$を最小二乗法により推定

Source	SS	df	MS
Model	2.7776e+10	1	2.7776e+10
Residual	1.0526e+10999998		10525.6486
Total	3.8302e+10999999		38301.6699

Number of obs	= 1000000
F(1,999998)	= .
Prob > F	= 0.0000
R-squared	= 0.7252
Adj R-squared	= 0.7252
Root MSE	= 102.59

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	41.6193	.0256203	1624.47	0.000	41.56908	41.66951
_cons	-292.0834	.2292232	-1274.23	0.000	-292.5326	-291.6341

78

- じゃあ、説明変数間の相関が高いときは、どうしたらいいのよ？って疑問を持たれるかもしれません。
- サンプルサイズを大きくしなよ、は一つの答えです。
  - これは残念ながらできないことが多いです、現実的には。
- 一つ知っておくといいこと、があります。
- それは、

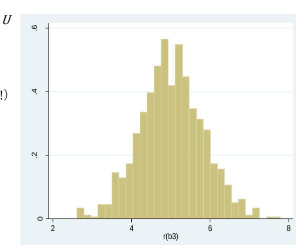
いくつかの説明変数間に「マルチコ」の問題があったとしても、それは他の変数の最小二乗推定量の分散には直接的には影響しない

ということです。

- 例えば、$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$というモデルを推定するとき、$X_1$と$X_2$の相関が高かったとしても、それは最小二乗推定量$\hat{\beta}_3$の分散には影響を与えないということです。

79

- 見てみましょう。
- $N = 200$
- モデル：$Y = 5 + 2{*}X_1 + 10{*}X_2 + 5{*}X_3 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差4
- $X_2 \sim$ 正規分布　平均2　標準偏差16
- $X_3 \sim$ 正規分布　平均5　標準偏差9
- ただしX_1とX_2は相関あり（相関係数0.99!）
- $U \sim$ 正規分布　平均0 標準偏差100
- 最小二乗推定量$\hat{\beta}_3$の分布（$M = 1000$）。

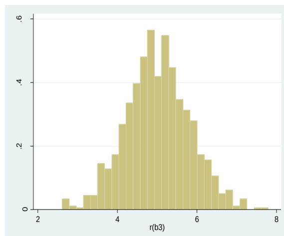


Variable	Obs	Mean	Std. Dev.	Min	Max
b3f	1000	5.021486	.8021252	2.610176	7.79676

80

- 見てみましょう。

- $N = 200$
- モデル: $Y = 5 + 2^{*}X_1 + 10^{*}X_2 + 5^{*}X_3 + U$
- $X_1 \sim$ 正規分布　平均8 標準偏差4
- $X_2 \sim$ 正規分布　平均2　標準偏差16
- $X_3 \sim$ 正規分布　平均5　標準偏差9
- ただしX_1とX_2は相関なし
- $U \sim$ 正規分布　平均0 標準偏差100
- 最小二乗推定量$\hat{\beta}_3$の分布（$M = 1000$）。



Variable	Obs	Mean	Std. Dev.	Min	Max
b3f	1000	5.021486	.8021252	2.610176	7.79676

- ここから以下のことが言えます。

1. 自分の興味の対象である説明変数に「マルチコ」が起きていないなら、特に問題はない。

2. 興味の対象ではない説明変数（こういう変数を**コントロール変数**と呼びます）間のマルチコはあまり気にする必要が無い。

   - なぜなら、今みたように、自分の興味の対象である説明変数の最小二乗推定量の分散には影響しないから。
   - 分析から抜いたりすると欠落変数バイアスが起きるので、抜かないほうが良い。

3. しかし、興味の対象である説明変数に「マルチコ」が起きてるときは、気にする必要があるかもしれません。

   - サンプル・サイズが十分に大きければ問題はないかもしれないが、、、
   - サンプル・サイズが小さいなら、精度の高い推定量にはならないので、リサーチク・エスチョン自体を修正する必要があるかもしれません。

モデルに不必要な変数を入れると？

- 不必要な変数をモデルに入れても不偏性・一致性には関係ないことは説明しました。
- ただし、特殊なケースを除いて、最小二乗推定量の分散が大きくなります。
- これは以下の式よりわかりますね。

$$VAR(\hat{\beta}_j | X_1, \dots, X_k) = \frac{\sigma^2}{(1 - R_j)\sum_{i=1}^{N}(X_{ij} - \bar{X}_j)^2}$$

ただしR_jは従属変数がX_jで説明変数が$1, X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k$である回帰モデルの決定係数。

- X_kがYには全く影響を与えない変数（不必要な変数）だとしましょう。
- 一般的には、変数間に何らかの相関はありますから、X_kがモデルに入っていると、入っていないときにくらべてR_jは大きくなります。
- したがって、$VAR(\hat{\beta}_j | X_1, \dots, X_k)$は$X_k$が入ることで大きくなります。

モデルに不必要な変数を入れる？

- まだ「統計的検定」については、なにもやっていませんが、、、、
- 検定の結果、X_kの係数がゼロであることを否定するだけの統計的根拠がない、ことが分かったとしましょう。
- ただし、このX_kはコントロール変数（興味の対象ではないが、Yに影響を与えるかもしれない変数）とします
- ことときX_kをモデルから抜いたほうが良いでしょうか？
- この答えは、「分析者の目的による」です。
- 目的が因果効果の推定であり、分析結果を人に見せるときは、抜かない方が良いです。
- 理由は、その説明変数をモデルに入れていないということで、欠落変数バイアスを疑われてしまうことがあるからです。

- 皆さんの（懸賞論文における）分析の目的は、ほとんどの場合「因果効果の推定」だと思います。
- 従って、不必要な説明変数でもモデルから抜かないほうがいいでしょう。
  - 「この変数はYと関係がない」というのも立派な結果ですし。
- 一方、目的がYの値の<u>予測</u>であるならば、<u>抜いた方が良い</u>です。
- Yをどれだけうまく「予測」できるかが目的ならば、必要のない変数をモデルに入れて推定量の精度を落としてしまうのは悪かもしれません。
- 「機械学習」の有力なツールの一つは回帰分析ですが、その目的は予測であることが多いです。
- そこでは、いかに変数を抜いていくかについて、いろいろな方法が考えられています、このコースの範囲外ですが。

回帰直線の「**標準誤差**」

$$VAR\left(\hat{\beta}_j \middle| X_1 = x_1, \dots, X_k = x_k\right) = \frac{\sigma^2}{(1 - R_j)\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}$$

- この式は誤差の分散σ^2を含んでいます。
- このままでは、絵に描いた餅です。
- 最小二乗推定量$\hat{\beta}_j$散らばりを知るためには、分散σ^2を推定する必要がありますね。
- 以下のように推定します。

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

で求められます。ここで$\hat{U}_i$は残差です。

- これに平方根を取ったもの、すなわち$\hat{\sigma}$を回帰直線の「**標準誤差**」と言います。
  - Root Mean Squared Errorともよばれます。Stataはこの言葉を使っています。

- $N - k - 1$で割っているのが奇妙ですね。
  - ちなみに単回帰モデルの時はk =1だったので、$N - 2$で割っていました。
- 理由ですが、その前に、、、
- 実は分散の推定にも、不偏性や一致性の概念があります。
- 標本分散、$N - 1$で割りますよね。あれは不偏性のためです。
- 例えば、Xっていう確率変数があって、母集団においてその分散がσ^2だとします。
- ランダムサンプル$\{X_1, \dots, X_N\}$があったとして、標本分散（推定量）は

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

です。

- このとき$E(\hat{\sigma}^2) = \sigma^2$となることが知られています。

- ってことは、Nで割った推定量 $\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$ だと、その期待値はσ^2と等しくはならないですね。
- 従ってNで割った上の推定量は不偏性を持ちません。
- そのため$N - 1$で割る推定量が使われています。
  - この$N - 1$で割ることを「自由度の修正」と言います。
- ただしどちらの推定量も一致性を持つことが知られています。
- Nが大きくなれば、$N - 1$で割ってもNで割っても大差ないですね。
- 同じような話が誤差項の分散の推定でもあります。
- $N - k - 1$で割る推定量（86ページ）には不偏性があります。
  - ここでは$N - k - 1$で割ることが「自由度の修正」です。
- ただしNが大きければ、Nで割っても問題ありません。

修正済み決定係数

- ようやくこの講義ノート、最後のトピックです。

```
      Source |       SS       df       MS              Number of obs = 1000000
-------------+------------------------------           F(  2,999997) =       .
       Model |  2.8130e+10     2  1.4065e+10           Prob > F      =  0.0000
    Residual |  1.0006e+10999997  10006.1558           R-squared     =  0.7376
-------------+------------------------------           Adj R-squared =  0.7376
       Total |  3.8136e+10999999  38135.7603           Root MSE      =  100.03

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.035064   .1772203    11.48   0.000     1.687718    2.382409
          x2 |   9.978366   .0443091   225.20   0.000     9.891522    10.06521
       _cons |   4.879367   1.333933     3.66   0.000     2.264904     7.49383
------------------------------------------------------------------------------
```

- これです、**修正済み決定係数**と言います。
- 何を修正したんだ、っていう話ですが、これは説明変数の数です。
- 決定係数は、新しい変数がモデルに追加されると大きくなります（小さくなることはありません）。
- なので、決定係数が大きくなったとしても、変数の追加によって本当にモデルの当てはまりが改善したかはっきりしないという問題があります。

- 決定係数R^2は

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

ただし

- $SST = \sum_{i=1}^{N}(y_i - \bar{y})^2$（データの$y$の散らばり）
- $SSE = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$（モデルが予測した$\hat{y}$の散らばり）
- $SSR = \sum_{i=1}^{N}\hat{u}_i^2$（残差の散らばり）

で、「yの散らばりの中で、説明変数により説明（予測）される割合」でした。

- 修正済み決定係数$\bar{R}^2$は

$$\bar{R}^2 = 1 - \frac{N-1}{N-k-1}\frac{SSR}{SST}$$

です。「修正」している点は、残差の二乗和と全体の二乗和の比率に係数$\frac{N-1}{N-k-1}$がかけられている点です。

$$\bar{R}^2 = 1 - \frac{N-1}{N-k-1}\frac{SSR}{SST}$$

- 修正済み決定係数について知っておくべきことは以下の3点です。

1. $\frac{N-1}{N-k-1}$は常に1より大きくなるので、$\bar{R}^2$は常にR^2より常に小さくなります。
   - Q: でも89ページでは二つの値は同じ。なぜでしょう？
2. $\bar{R}^2$は説明変数を追加することで、2つの相反する影響を受けます。
   - SSR（残差の二乗和）が減少する → $\bar{R}^2$が大きくなる
   - kが大きくなる → $\frac{N-1}{N-k-1}$が大きくなる → $\bar{R}^2$が小さくなる

従って、最終的に$\bar{R}^2$が大きくなるか小さくなるかは、これら二つの効果のどちらが強いかに依存します。

3. $\bar{R}^2$は負になることがあります。これは説明変数を加えても残差の二乗和の減少がほとんどなく、$\frac{N-1}{N-k-1}$が大きくなることの効果を相殺できない場合に起こります。