

第 15 回 : まとめ

北村 友宏

2021 年 1 月 22 日

本日の内容

1. ミクロデータとは
2. 線形回帰モデル
3. 係数の仮説検定
4. ダミー変数を含む回帰
5. 回帰係数の均一性の検定
6. 変数の対数変換
7. 誤差項の不均一分散への対処
8. レポート・論文用の表の作成

ミクロ・集計データ

- ▶ 個人，家計，事業所，企業などの観測単位からなるデータを**ミクロデータ (micro data)** という。
 - ▶ 個人の所得，消費・個別市場での財の価格・数量，個人の成績
- ▶ ミクロデータを市町村，都道府県，国などの単位で合計または平均したデータを**集計データ (aggregate data)** という。
 - ▶ 個人の所得，消費の各都道府県における平均

この授業では「ミクロデータを都道府県別や市町村別に集計したもの」を「広義のミクロデータ」とする。

⇒ タームペーパーでは「広義のミクロデータ」を用いてよい。

線形回帰モデル

- ▶ 単回帰モデル :

- ▶ 定数項以外の説明変数が1つである回帰モデル

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

- ▶ 重回帰モデル :

- ▶ 定数項以外に説明変数が複数ある回帰モデル

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i.$$

係数の仮説検定

係数の仮説検定における帰無仮説と対立仮説は、

- ▶ 帰無仮説：「その係数は 0」
- ▶ 対立仮説：「その係数は 0 と異なる」

- ▶ p 値が 0.1 を超える・アスタリスクなし:
 - ▶ 有意水準 10%でも「その係数は 0」の帰無仮説を採択.
- ▶ p 値が 0.1 以下 (未満)・* :
 - ▶ 有意水準 10%で「その係数は 0」の帰無仮説を棄却.
- ▶ p 値が 0.05 以下 (未満)・** :
 - ▶ 有意水準 5%で「その係数は 0」の帰無仮説を棄却 (有意水準 10%でも棄却できる).
- ▶ p 値が 0.01 以下 (未満)・*** :
 - ▶ 有意水準 1%で「その係数は 0」の帰無仮説を棄却 (有意水準 10%や 5%でも棄却できる).

※検定統計量が連続型の確率分布 (正規分布, t 分布, カイ二乗分布, F 分布など) に従う場合, 「以上」と「超える」, 「以下」と「未満」は区別しなくて良い.

▶ 帰無仮説が棄却された場合

➡ 対立仮説を採択し、「その係数は統計的に有意に0と異なる」と判断。

- ▶ 「その説明変数は被説明変数と統計的に有意に相関している」と解釈。
- ▶ 定数項の検定の場合は「定数項は統計的に有意に0と異なる」と解釈。

▶ 帰無仮説が採択された場合

➡ 「その係数は0と異なるとは言えない」と判断。

- ▶ 「その説明変数は被説明変数と相関しているとは言えない」と解釈。
- ▶ 定数項の検定の場合は「定数項は統計的に有意に0と異なるとは言えない」と解釈。

何を帰無仮説としているのか、そして棄却・採択の意味を考えること！

ダミー変数を含む回帰

d_i をダミー変数として，説明変数にダミー変数を含む線形回帰モデル

$$y_i = \beta_0 + \beta_X x_i + \beta_D d_i + u_i,$$

においては，

- ▶ 除外したダミー変数が表すものを基準として，ダミー変数の（偏）回帰係数は**基準と比較して**どの程度，被説明変数に対する影響度合いが異なるか，という解釈。
 - ▶ e.g., 「ワンルームダミー+その他ダミー=1」となっているとき，「その他ダミー」を除外して「ワンルームダミー」を説明変数に用いた場合，「ワンルームダミー」の係数は，「ワンルームのマンションはそれ以外の種類のマンションと比べて被説明変数がどの程度異なる傾向があるか」を表す。

回帰係数の均一性の検定

d_i をダミー変数として,

$$y_i = \beta_0 + \beta_X x_i + \beta_D d_i + \beta_{DX} d_i x_i + u_i,$$

を推定し,

- ▶ 帰無仮説 : $\beta_D = 0$ and $\beta_{DX} = 0$
- ▶ 対立仮説 : $\beta_D \neq 0$ or $\beta_{DX} \neq 0$

を検定すれば, $d_i = 0$ のグループと $d_i = 1$ のグループで回帰係数が異なるかどうかを検定できる.

- ▶ 帰無仮説棄却 \Rightarrow 両グループで回帰係数が統計的に有意に異なる.
- ▶ 帰無仮説採択 \Rightarrow 両グループの回帰係数には統計的に有意な差がない.

変数の対数変換

対数変換していないものを**レベル (level)** という。
説明変数や被説明変数がレベルなのか対数値なのかによって、説明変数の回帰係数の解釈が異なる。

- ▶ **Level-Level Model:**

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- ▶ β_1 の解釈 :

- x_i が 1 **単位** 増加すると y_i は β_1 **単位** 増加する。

- ▶ **Log-Log Model:**

$$\ln y_i = \beta_0 + \beta_1 \ln x_i + u_i$$

- ▶ β_1 の解釈 :

- x_i が 1% 増加すると y_i は β_1 % 増加する。

- (「 $100 \times \beta_1$ %」 にしない!)

- ▶ β_1 は y_i の x_i に対する**弾力性 (elasticity)** .

▶ Log-Level Model:

$$\ln y_i = \beta_0 + \beta_1 x_i + u_i$$

▶ β_1 の解釈:

x_i が 1 単位増加すると y_i は $100 \times \beta_1\%$ 増加する.

▶ β_1 は y_i の x_i に対する半弾力性 (semi-elasticity) .

▶ Level-Log Model:

$$y_i = \beta_0 + \beta_1 \ln x_i + u_i$$

▶ β_1 の解釈:

x_i が 1% 増加すると y_i は $\frac{1}{100} \times \beta_1$ 単位増加する.

▶ あまり使われない.

誤差項の不均一分散

- ▶ $V(u_i | x_i)$ が一定でないことを（条件付き）不均一分散（heteroskedasticity）という。
- ▶ 不均一分散があっても、頑健標準誤差（robust standard error）を計算すれば、標準誤差をより厳密に求め、仮説検定をより厳密に行うことができる。
- ▶ 実行可能な一般化最小二乗法（Feasible Generalized Least Squares, FGLS）を用いて、モデル推定の段階で誤差項の不均一分散を取り除く方法もある。

「OLS + 頑健標準誤差」と「FGLS」

- ▶ 通常の最小二乗法（OLS）で推定し，頑健標準誤差を計算する方法
 - ▶ 誤差項に不均一分散があるという前提で推定を行い，標準誤差を修正している。
 - ▶ 実証分析では，こちらがよく使われる。
- ▶ 実行可能な一般化最小二乗法（FGLS）で推定する方法
 - ▶ 誤差項の不均一分散を取り除いた状態にして推定を行っている。
 - ▶ 実証分析ではあまり使われないが，観測値数が十分に大きいとき OLS 推定量よりも分散の小さな推定量が得られるという長所がある。

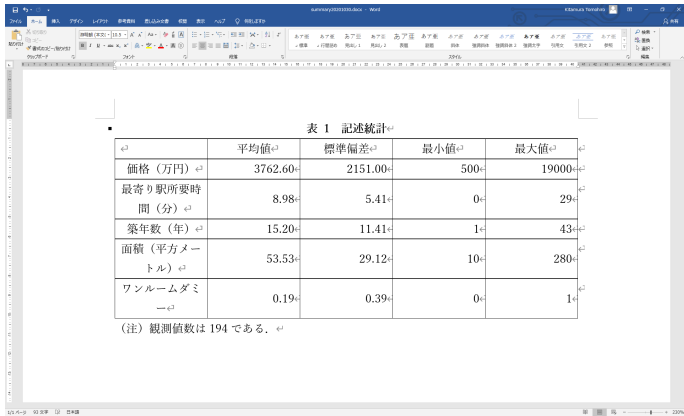
注：元のモデルの誤差項の条件付き分散が均一でない限り，OLS 推定量と FGLS 推定量は異なった値になる。

レポート・論文用の記述統計表の作成

見やすく，理解しやすい表を載せるには，

- ▶ 表番号と表のタイトルをつける。
- ▶ 変数名は統計解析ソフトでの変数名そのままではなく，分かりやすいように書き直す。
- ▶ 単位がある変数は単位を明記する。
- ▶ 小数の数値はあまり細かく表示せず，小数第1～4位程度まで示せば十分。縦方向に見たとき，小数点の位置が揃うようにする。
- ▶ 観測値数が全変数について同じ場合，表の下に「注（Note）」として「観測値数は***である」などと注記してもよい。

例えば、以下のような表を載せればよい。



The image shows a Microsoft Word document with a table titled "表 1 記述統計". The table has five columns: "平均値", "標準偏差", "最小値", and "最大値". The rows represent different housing attributes: "価格 (万円)", "最寄り駅所要時間 (分)", "築年数 (年)", "面積 (平方メートル)", and "ワンルームダミー". Below the table, a note states "(注) 観測値数は 194 である。".

	平均値	標準偏差	最小値	最大値
価格 (万円)	3762.60	2151.00	500	19000
最寄り駅所要時間 (分)	8.98	5.41	0	29
築年数 (年)	15.20	11.41	1	43
面積 (平方メートル)	53.53	29.12	10	280
ワンルームダミー	0.19	0.39	0	1

(注) 観測値数は 194 である。

レポート・論文用のモデル推定結果表の作成

見やすく，理解しやすい表を載せるには，

- ▶ 表番号と表のタイトルをつける。
- ▶ 最低限，以下の情報を載せる（線形回帰モデルの場合）。
 - ▶ 係数推定値
 - ▶ t 値または標準誤差または p 値のどれか（学術論文では標準誤差を載せる場合が多い。）
 - ▶ 決定係数
 - ▶ 観測値数
- ▶ 有意性を示すアスタリスクを付けた場合は，表の下に「注（Note）」として「表中の***，**，* はそれぞれ有意水準 1%，5%，10%で統計的に有意であることを表す」などと注記する。

- ▶ 仮説検定に用いた標準誤差の種類や、頑健なら何に対して頑健なのかを、表中に明記するか、表の下に「注 (Note)」として「不均一分散に対して頑健な標準誤差を用いている」などと注記する.
- ▶ 観測値数は、表の下に「注 (Note)」として「観測値数は〇〇〇である」などと注記してもよい
 - ▶ e.g., 1つの表に複数のモデルを載せ、観測値数が全モデルについて同じ場合など.
- ▶ 変数名は統計解析ソフトでの変数名そのままではなく、分かりやすいように書き直す.
- ▶ 小数の数値はあまり細かく表示せず、小数第2~4位程度まで示せば十分. 特に係数推定値, 標準誤差, t 値, p 値は縦方向に見たとき、可能な限り小数点の位置が揃うようにする.
 - ▶ t 値は小数第2位まで, p 値, 決定係数は小数第3位まで示せば十分.

例えば、以下のような表を載せればよい。

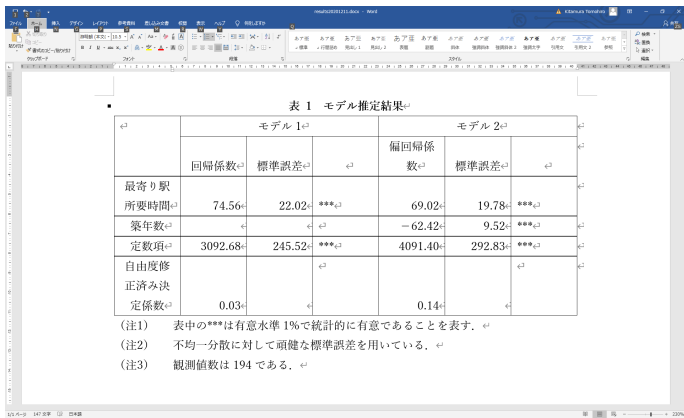


表 1 モデル推定結果

	モデル 1			モデル 2		
	回帰係数	標準誤差		偏回帰係数	標準誤差	
最寄り駅 所要時間	74.56	22.02	***	69.02	19.78	***
築年数				-62.42	9.52	***
定数項	3092.68	245.52	***	4091.40	292.83	***
自由度修 正係数	0.03			0.14		

(注1) 表中の***は有意水準 1% で統計的に有意であることを表す。
(注2) 不均一分散に対して頑健な標準誤差を用いている。
(注3) 観測値数は 194 である。