講義内容の詳細なまとめ：最小二乗推定量の性質と仮説検定

この講義では、線形回帰モデルにおける最小二乗法（Ordinary Least Squares, OLS）で得られる推定量（$\hat{\beta}$）の統計的性質を深く掘り下げ、それを用いて仮説検定や区間推定を行うための理論的背景と実践的な手法について詳細に解説しています。特に、分散不均一性（Heteroskedasticity）を考慮することの重要性が一貫して強調されています。

1. 最小二乗法（OLS）の基本と望ましい性質

- 前提: 我々が知りたいのは、$Y = \beta_0 + \beta_1 X_1 + \dots + u$ という「真の関係」だが、これは観測不可能である。我々はデータを用いてモデルを当てはめ、真のパラメータ β を推定する。
- **OLS**の原理: OLSは、観測されたデータ点と回帰直線との縦方向の距離（残差）の二乗和が最小になるように回帰係数 $\hat{\beta}$ を決定する手法である。
- 重要な仮定: この推定方法が妥当であるためには、いくつかの仮定が必要となる。中でも特に重要なのが「誤差項 **u** の条件付き期待値が**0**である」という仮定 ($E(u|X) = 0$) である。
    - 社会科学のデータでは、誤差項 u には単なる測定誤差だけでなく、モデルに含まれていない（X以外の）全ての Y に影響を与える要因が含まれるため、この仮定は非常に強い意味を持つ。
- **OLS**推定量の良い性質: 上記の仮定などが満たされる時、OLS推定量 $\hat{\beta}$ は以下の望ましい性質を持つことが証明される。
    1. 不偏性 **(Unbiasedness)**: 推定量の期待値が、真のパラメータの値と一致する ($E(\hat{\beta}) = \beta$)。
    2. 一致性 **(Consistency)**: サンプルサイズ (n) を無限に大きくすると、推定量 $\hat{\beta}$ は真のパラメータ β に確率的に収束する。
    - これらの性質があるからこそ、OLSは「良いやり方」として保証される。もし一致性がなければ、どれだけ多くのデータを集めても真の値を知ることができないということになる。

2. OLS推定量の漸近分布と分散の推定

- 漸近正規性: OLS推定量 $\hat{\beta}$ は、サンプルサイズ n が大きい時、その分布が正規分布に近似できるという性質（漸近正規性）を持つ。具体的には、$\sqrt{n}(\hat{\beta} - \beta)$ という量が、平均0の正規分布に従う。
- 分散の推定:
    - この漸近的な正規分布の分散（分散共分散行列）V は未知であるため、データから推定する必要がある。
    - 分散の計算式には観測不可能な誤差項 u が含まれるが、これを観測可能な残差 $\hat{u}$ で置き換えることで、分散の推定量 $\hat{V}$ を計算する。
    - 一致性により、サンプルサイズ n が大きければ $\hat{\beta}$ は β に近づき、結果として残差 $\hat{u}$ も誤差項 u に近づくため、この置き換えは正当化される。

3. 分散不均一 vs. 分散均一

講義全体を通じて、この2つの概念の区別が極めて重要であるとされている。

- 分散不均一 **(Heteroskedasticity)**:
    - 定義: 誤差項の分散が、説明変数 X の値によって異なる状況を指す ($Var(u|X)$ が X の関数となる）。例えば、「X が大きいほどデータのばらつきが大きくなる」といったケース

    が該当する。
  - 特徴: 実際の経済データや社会科学のデータでは、分散不均一である方が圧倒的に多いと考えられる。
- 分散均一 **(Homoskedasticity)**:
  - 定義: 誤差項の分散が、説明変数 X の値によらず一定である状況 ($Var(u|X) = \sigma^2$ (定数))。X の値がどこであっても、データのばらつき方が同じである。
  - 特徴: 非常に強い仮定であり、現実のデータでは稀。
- 頑健な標準誤差 **(Robust Standard Errors)**:
  - 講義で紹介された分散の推定式 $\hat{V}$ は、分散が不均一であっても均一であっても正しく機能するため、「分散不均一に対して頑健（ロバスト）」であると言われる。
  - 古い教科書や一部のソフトウェアのデフォルト設定では、分散均一を仮定したよりシンプルな分散の計算式が使われている場合があり、注意が必要。分散不均一が一般的な現代の計量経済学では、頑健な推定量を用いることが推奨される。

4. 仮説検定①：**t**検定（単一の係数に対する検定）

- 目的: ある特定の係数 β_m が、特定の値（例えば0）と等しいか否か（$H_0: \beta_m = 0$）を統計的に判断する。
- 手順:
  1. OLSで係数 $\hat{\beta}_m$ を推定する。
  2. $\hat{\beta}_m$ の標準誤差 (Standard Error) を計算する。これは分散の推定量 $\hat{V}$ の対角要素の平方根である。
  3. **t**統計量を計算する: $t = \frac{\hat{\beta}_m - (\text{仮説の値})}{\text{標準誤差}}$。
- **t**検定のメカニズム:
  - もし帰無仮説が正しければ: t 統計量は（漸近的に）標準正規分布（平均0、分散1）に従う。その値は、概ね-2から+2の間に収まる確率が高い。
  - もし帰無仮説が間違っていれば: 分子（$\hat{\beta}_m - 0$）は0でない値に収束し、分母（標準誤差）はサンプルサイズ n の増大と共に0に近づく。結果として、t 統計量の絶対値はプラスまたはマイナスの無限大に発散する。
- 判定: 計算された t 統計量の絶対値が、有意水準5%の場合の閾値である1.96などを超えていれば、「偶然では起こりにくい」と判断し、帰無仮説を棄却する。

5. 信頼区間

- 目的: 真のパラメータ β_m が95%などの確率で含まれると期待される区間を構成する。
- 計算式 **(95%**信頼区間**)**: $[\hat{\beta}_m - 1.96 \times SE, \quad \hat{\beta}_m + 1.96 \times SE]$。
- 解釈上の注意点:
  - これは「真の値 β_m が確率的に変動し、95%の確率でこの区間に入る」という意味ではない。
  - 真の値 β_m は未知の定数である。正しくは、「区間自体がデータから計算される確率的なものであり、同様の実験を100回繰り返した場合、そのうち95回はこの区間が真の値を含むだろう」と解釈する。

6. 仮説検定②：**F**検定（複数の係数に対する複合仮説の検定）

- 目的: 「$\beta_1=0$ かつ $\beta_2=0$」のように、複数の係数に関する制約を同時に検定する。
- 一般形: 複数の線形制約は、$R\beta - r = 0$ という行列の形で一般的に表現できる。

- **t**検定の繰り返しはなぜダメか？:
  - 個別の係数についてt検定を繰り返し行い、一つでも棄却されたら複合仮説全体を棄却する、という手続きは誤りである。
  - 個々の検定の有意水準を5%に設定しても、検定全体で誤りを犯す確率（ファミリーワイズエラー率）が5%にはならず、コントロールできなくなるためである。
- **Wald**（ウォルド）検定:
  - この複合仮説を検定するために、Wald統計量（W統計量）が用いられる。
  - これは、R\hat{\beta}-r がどれだけ0から離れているかを、係数間の相関を考慮した「距離」（二次形式）で測るものである。
  - メカニズム: 帰無仮説が正しい場合、W統計量は自由度q（制約の数）のカイ二乗分布に漸近的に従う。一方、帰無仮説が間違っている場合、W統計量は無限大に発散する。
- **F**検定:
  - 実務上は、W統計量をその自由度 q で割った**F**統計量 (F = W/q) がよく使われる。
  - 利点: F統計量は、自由度 q が変わっても棄却の目安となる値が大きく変動せず、多くの場合「およそ4」を超えるかどうかで大まかな判断ができるため、直感的に解釈しやすい。
- 注意点:
  - 制約の数 q は、仮説を表現するための「最小の数」でなければならない。冗長な制約を含めると検定が正しく行えない。
  - 多くの統計ソフトが出力するF値は、「切片項以外の全ての係数が同時に**0**である」という仮説（つまり、モデル全体に意味があるか）を検定している。

7. モデルの当てはまりの良さ：決定係数 **R^2**

- 定義: 被説明変数 Y の全変動のうち、モデル（説明変数 X）によって説明できた部分の割合を示す指標。
- 性質: 0から1の値をとり、1に近いほどモデルの当てはまりが良いことを示す。
- 問題点: 説明変数を追加すると、R^2 は必ず増加（または非減少）する。たとえ無意味な変数を加えても値が上昇するため、モデルの良し悪しの判断には使えない。
- 自由度調整済み決定係数 **(Adjusted R^2)**:
  - 上記の問題を緩和するため、説明変数の数でペナルティを課した指標。
  - 安易に変数を増やすと、この値は逆に低下することがある。通常の R^2 よりも低い値をとり、場合によっては負になることもある。
- 解釈: R^2 はあくまで当てはまりの良さの一つの側面に過ぎず、この値が高いからといってそのモデルが因果関係を正しく捉えているとは限らない点に注意が必要である。