

第 4 章: 重回歸分析: 推論

Jeffrey Wooldridge (2016).

Introductory Econometrics: A Modern Approach
Seventh Edition. Cengage Learning.

2026-03-15

準備

必要なパッケージの読み込み

▶ wooldridge パッケージの読み込み

```
library(wooldridge)
```

4-1 古典的線形回帰モデルの仮定

誤差項の正規性の仮定 (MLR.6)

- ▶ これまでのガウス・マルコフの仮定 (MLR.1 ~ MLR.5) に加えて、推論 (検定) を行うためには、誤差項 u の分布に関する仮定が必要となる。

仮定 MLR.6 (正規性) 誤差項 u は、説明変数 x_1, x_2, \dots, x_k とは独立であり、平均 0、分散 σ^2 の正規分布に従う。

$$u \sim \text{Normal}(0, \sigma^2)$$

- ▶ この仮定の下で、OLS 推定量 $\hat{\beta}_j$ もまた正規分布に従う。

$$\hat{\beta}_j \sim \text{Normal}(\beta_j, \text{Var}(\hat{\beta}_j))$$

4-2 t 検定: 単一の母集団パラメータに関する 検定

t 統計量 (1)

- ▶ 特定のパラメータ β_j に関する仮説検定を行うために、標準化を行う。

$$\frac{\hat{\beta}_j - \beta_j}{\text{sd}(\hat{\beta}_j)} \sim \text{Normal}(0, 1)$$

t 統計量 (2)

- ▶ しかし、標準偏差 $\text{sd}(\hat{\beta}_j)$ は未知の σ に依存するため、推定値である標準誤差 $\text{se}(\hat{\beta}_j)$ で置き換える。これにより、分布は t 分布となる。

$$t = \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \sim t_{n-k-1}$$

- ▶ n : サンプルサイズ、 k : 説明変数の数、 $n - k - 1$: 自由度 (df)

帰無仮説と対立仮説

- ▶ 最も一般的な検定は、変数が従属変数に影響を与えないという帰無仮説である。

$$H_0 : \beta_j = 0$$

- ▶ 対立仮説（両側検定の場合）:

$$H_1 : \beta_j \neq 0$$

- ▶ この H_0 の下での t 統計量（t 値）は以下のようなになる。

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}$$

R による実装例 (meap93)

- ▶ ミシガン州の学校データ (meap93) を用いて、数学の合格率 (math10) を、支出 (expend) と 昼食補助受給率 (lnchprg) で説明するモデル。

```
# meap93 データセットを読み込む
data(meap93)
# 支出を対数変換してモデル化
model_meap <- lm(math10 ~ log(expend) + lnchprg,
                  data = meap93)
```

R による t 検定の結果表示

```
# 結果の要約 (t 値を含む)
```

```
summary(model_meap)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-20.3608165	25.07287437	-0.8120655	4.172311e-01
## log(expend)	6.2296983	2.97263426	2.0956827	3.673061e-02
## lnchprg	-0.3045853	0.03535742	-8.6144670	1.587327e-16

t 検定の結果解釈

- ▶ $\log(\text{expend})$ の推定係数は 6.23、標準誤差は 2.97。
- ▶ t 値は 2.1。
- ▶ 自由度 405 の t 分布において、この値は非常に稀（棄却域に入る）。
- ▶ したがって、 $H_0 : \beta_{\log(\text{expend})} = 0$ は棄却され、支出と数学の合格率には統計的に有意な正の関連があると結論付けられる（因果解釈には追加の識別仮定が必要）。

4-3 p 値と信頼区間

p 値 (p-value)

- ▶ **p 値**とは、帰無仮説が正しいとした場合に、観測された t 値 (の絶対値) 以上の値が得られる確率のこと。
- ▶ p 値が有意水準 (例: $\alpha = 0.05$) より小さければ、 H_0 を棄却する。
 - ▶ $p < 0.05$: 5%水準で統計的に有意
 - ▶ $p < 0.01$: 1%水準で統計的に有意
- ▶ 先ほどの例では、 $\log(\text{expend})$ の p 値 ($\Pr(>|t|)$) は 0.0367 であり、非常に小さい。

信頼区間 (Confidence Interval)

- ▶ パラメータ β_j の真の値が含まれると期待される範囲。
- ▶ 95%信頼区間の構成:

$$\hat{\beta}_j \pm c \cdot \text{se}(\hat{\beta}_j)$$

- ▶ ここで c は t_{n-k-1} 分布の 97.5 パーセンタイル点 (自由度が大きければ約 1.96)。

Rによる信頼区間の計算例

```
# 95%信頼区間の計算  
confint(model_meap)
```

```
##                2.5 %      97.5 %  
## (Intercept) -69.6500430 28.9284100  
## log(expend)   0.3859788 12.0734177  
## lnchprg      -0.3740923 -0.2350783
```

4-4 F 検定: 複数の線形制約の検定

複数の仮説の同時検定

- ▶ 複数の変数が**同時に**意味を持たないかどうかを検定したい場合がある。
- ▶ 例: 野球選手の年俸モデルにおいて、成績に関する変数群 (打率、ホームラン数、打点) がすべて無関係か?

$$H_0 : \beta_{\text{batavg}} = 0, \beta_{\text{hruns}} = 0, \beta_{\text{rbis}} = 0$$

- ▶ 対立仮説 H_1 : 少なくとも一つの β は 0 ではない。
- ▶ この場合、個々の t 検定を見るだけでは不十分であり、**F 検定**を用いる必要がある。

F 統計量

- ▶ **制約なしモデル** (Unrestricted Model: UR) と、制約 (H_0) を課した**制約付きモデル** (Restricted Model: R) の SSR (残差二乗和) を比較する。

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n - k - 1)}$$

- ▶ q : 制約の数 (H_0 で 0 とするパラメータの数)
- ▶ $n - k - 1$: 制約なしモデルの自由度
- ▶ H_0 が正しい場合、SSR はあまり増加しないはずである。F 値が大きければ H_0 を棄却する。

R による F 検定の実装例 (mlb1)

▶ メジャーリーグの選手データ (mlb1)。

```
# mlb1 データセットを読み込む
data(mlb1)
# 制約なしモデル
model_ur <- lm(log(salary) ~ years + gamesyr + bavg +
               hrunsyr + rbisyr, data = mlb1)
# 制約付きモデル (成績変数をすべて除外)
model_r <- lm(log(salary) ~ years + gamesyr, data = mlb1)
```

anova 関数による比較

```
# F 検定の実行
```

```
anova(model_r, model_ur)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: log(salary) ~ years + gamesyr
```

```
## Model 2: log(salary) ~ years + gamesyr + bavg + hrunsyr + rbisyr
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1      350 198.31
```

```
## 2      347 183.19  3    15.125 9.5503 4.474e-06 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- ▶ 出力結果の F 値と Pr(>F) を確認する。
- ▶ p 値が非常に小さければ、成績変数はグループとして統計的に有意である。

モデル全体の有意性検定

- ▶ 通常の `summary()` 出力の最後にある F-statistic は、すべての独立変数（切片以外）が 0 であるという帰無仮説に対する検定結果である。

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

- ▶ これにより、モデル全体として説明力があるかどうかを判断できる。

まとめ

まとめ

- ▶ **正規性の仮定 (MLR.6)** の下で、t 検定や F 検定が可能になる。
- ▶ **t 検定:** 個々の変数の統計的有意性を判断する。t 値と p 値を確認する。
- ▶ **信頼区間:** パラメータの真の値が含まれる範囲を推定する。
- ▶ **F 検定:** 複数の変数（または線形制約）がグループとして有意かどうかを判断する。
- ▶ 統計的有意性（p 値が小さい）と、経済的（実質的）重要性は区別する必要がある。サンプルサイズが大きいと、実質的に意味のない小さな効果でも統計的に有意になることがある。