回帰分析(F 検定)†

担当: 長倉　大輔
(ながくらだいすけ)

† この資料は私のゼミおよび講義において使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、存在する可能性は常にあります。

回帰分析(F 検定)

- F 検定

重回帰モデル:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i\ , \ \ (i = 1, \ldots, n)$$

$E(\varepsilon_i) = 0$, $\mathrm{Var}\ (\varepsilon_i) = \sigma^2$, において複数の係数についての仮説を検定したい時には ***F* 検定**と呼ばれる検定がよく用いられる。

回帰分析(F 検定)

- F 検定の帰無仮説

複数の係数についての仮説とは、例えば

(a) $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$,

や

(b) $H_0: \beta_2 = \beta_3$, (c) $H_0: 2\beta_2 + 3\beta_3 = \beta_4$

などである。このような仮説の検定は1つの係数に関する 検定である t 検定ではできない。

回帰分析(F 検定)

- F 検定の帰無仮説の行列表示

先ほどの重回帰モデルを行列を用いて:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

と表現しよう。この時、F 検定が検定できる仮説は

$$H_0: \ \mathbf{C\beta} = \mathbf{r}$$

とかける。ここで $\mathbf{C}$ は $p \times k$ 行列、$\mathbf{r}$ は $p \times 1$ ベクトルである。

回帰分析(F 検定)

例えば、先ほどの例では

(a) では $\mathbf{C} = \mathbf{I}_k$, $\mathbf{r} = \mathbf{0}$, (この時 $p = k$)

(b) では $\mathbf{C} = [0, 1, -1, \ldots, 0]$, $\mathbf{r} = 0$, (この時 $p = 1$)

(c) では $\mathbf{C} = [0, 2, 3, -1, \ldots, 0]$, $\mathbf{r} = 0$, (この時 $p = 1$)

とすればよい。

回帰分析(F 検定)

- F 検定の直観的説明

係数 $\boldsymbol{\beta}$ は直接観測できないのでその最小二乗推定量 $\hat{\boldsymbol{\beta}}$ で置き換えるとすると、先ほどの仮説が成り立っているとすれば、

$$\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{r}$$

は 0 (スカラーもしくはベクトル) に近い値をとるであろう。

F 検定とは、この値が**十分に 0 に近いかどうか** (近ければ仮説を採択、逆に離れていれば仮説を棄却) を統計的に検証する方法である。

回帰分析(F 検定)

- F 検定統計量の定義

F検定統計量 F は具体的には以下のように定義される。

(F検定統計量)

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{r})^{\mathrm{T}}[s^2\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}}]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{r})}{p}$$

ここで s^2 はσ^2のOLS推定量である。

回帰分析(F 検定)

- F 検定統計量について

実は $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ は帰無仮説のもとで、漸近的に(もしくは誤差項に正規分布を仮定したときに)

$$\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r} \sim N(0, \sigma^2\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}})$$

となる事を示すことができる。つまり F 検定統計量とは $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ の分散共分散行列の逆行列に左右から $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ をかけて(左から掛けられているものは転置を取っていることに注意)、σ^2 をそのOLS推定量 s^2 で置き換えたものを制約の数 p で割ったものとして定義されている。

回帰分析(F 検定)

- F 検定の分布

F 検定統計量は漸近的に自由度 p の **χ^2 (カイ二乗と読む)分布**と呼ばれる分布 を p で割った分布に従う。

また、もしε_i が正規分布に従う場合（そう仮定して問題ない時には）F 検定統計量は 第1自由度 p, 第2自由度 $n-k$ の ***F* 分布**と呼ばれる分布に従う。

以下カイ二乗分布との F 分布について簡単に説明する

回帰分析(F 検定)

- χ^2 分布

m 個の確率変数 $X_1, X_2, \ldots, X_m$ がそれぞれ**独立に**標準正規分布に従うとする。この時、それらの2乗の和

$$Y = X_1^{\,2} + X_2^{\,2} + \ldots + X_m^{\,2}$$

の分布を**自由度 m の χ^2 分布**といい $Y \sim \chi^2(m)$ と表記する。

カイ二乗分布の期待値は $E(Y) = m$, 分散は$\mathrm{Var}(Y) = 2m$ となる。

回帰分析(F 検定)

- F 分布

2 つの確率変数 X_1 と X_2 がそれぞれ**独立に** $X_1 \sim \chi^2(m)$ と$X_2 \sim \chi^2(n)$ であるとする。この時

$$W = \frac{X_1 / m}{X_2 / n}$$

と定義される確率変数 W の分布を第1自由度 m, 第2自由度 n のF 分布、もしくは単に**自由度 (m, n) の F 分布**と呼び、$W \sim F(m, n)$と表記する。

回帰分析(F 検定)

- F 検定の棄却域

F 検定は $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ が十分 0 から離れているかをみるための統計量であるので、本質的には $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ の2乗である F 検定統計量は $\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r}$ が 0 から(正もしくは負の方向に)離れるにつれ大きくなっていく。よって F 検定における棄却域は、有意水準を α とすると、F分布に従う確率変数 W に対して

$$\Pr(W > c) = \alpha$$

となる値 c を分布表から求め、実際のF 検定統計量の値がこの値を超えれば有意水準 α で帰無仮説を棄却するという事になる。

回帰分析(F 検定)

- F 検定の例

係数 $\boldsymbol{\beta}$ を $\boldsymbol{\beta} = [\,\boldsymbol{\beta}_1^{\mathrm{T}}, \boldsymbol{\beta}_2^{\mathrm{T}}\,]^{\mathrm{T}}$, 説明変数 $\mathbf{X}$ を $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ と分割しよう。ここで $\boldsymbol{\beta}_1 : k_1 \times 1$ ベクトル、$\boldsymbol{\beta}_2 : k_2 \times 1$ ベクトル、$\mathbf{X}_1: n \times k_1$ 行列、$\mathbf{X}_2: n \times k_2$ 行列、$k_1 + k_2 = k$ とする。帰無仮説 $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ を検定する F 検定統計量はp.7の F 検定統計量を計算する式において

$$\underset{k_2\times k}{\mathbf{C}} = [\ \underset{k_2\times k_1}{\mathbf{0}}\ ,\ \underset{k_2\times k_2}{\mathbf{I}_{k_2}}\ ],\quad \mathbf{r} = \underset{k_2\times 1}{\mathbf{0}}\ ,\quad p = k_2\,,$$

である事に注意すると

$$\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{r} = [\ \mathbf{0},\ \ \mathbf{I}_{k_2}\ ]\begin{bmatrix}\hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2\end{bmatrix} = \hat{\boldsymbol{\beta}}_2$$

を得る。また分割行列の逆行列の公式より

回帰分析(F 検定)

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \left(\begin{bmatrix} \mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2 \\ \mathbf{X}_2^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \mathbf{F}_1 & -(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_1(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1} & \mathbf{F}_2 \end{bmatrix}$$

である。ここで $\mathbf{F}_1 = (\mathbf{X}_1{}^{\mathrm{T}}\mathbf{X}_1 - \mathbf{X}_1{}^{\mathrm{T}}\mathbf{X}_2(\mathbf{X}_2{}^{\mathrm{T}}\mathbf{X}_2)^{-1}\mathbf{X}_2{}^{\mathrm{T}}\mathbf{X}_1)^{-1}$,
および $\mathbf{F}_2 = (\mathbf{X}_2{}^{\mathrm{T}}\mathbf{X}_2 - \mathbf{X}_2{}^{\mathrm{T}}\mathbf{X}_1 (\mathbf{X}_1{}^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1{}^{\mathrm{T}}\mathbf{X}_2)^{-1}$ である。

回帰分析(F 検定)

さらに

$$
\begin{aligned}
&\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}} \\
&= \begin{bmatrix} \mathbf{0}, & \mathbf{I}_{k_2} \end{bmatrix} \begin{bmatrix} \mathbf{F}_1 & -(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_1(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1} & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{k_2} \end{bmatrix} \\
&= \mathbf{F}_2
\end{aligned}
$$

も得る。よって

回帰分析(F 検定)

$$F=\frac{(\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r})^{\mathrm{T}}[s^2\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}}]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}-\mathbf{r})}{p}$$

$$=\frac{\hat{\boldsymbol{\beta}}_2^{\mathrm{T}}[s^2\mathbf{F}_2]^{-1}\hat{\boldsymbol{\beta}}_2}{k_2}$$

$$=\frac{\hat{\boldsymbol{\beta}}_2^{\mathrm{T}}[s^2(\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2-\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_1(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2)^{-1}]^{-1}\hat{\boldsymbol{\beta}}_2}{k_2}$$

$$=\frac{\hat{\boldsymbol{\beta}}_2^{\mathrm{T}}[\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2-\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_1(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2]\hat{\boldsymbol{\beta}}_2}{s^2k_2}$$

がこの帰無仮説を検定するF検定統計量である。

演習問題(F 検定)

問題1　$X \sim \chi^2(m)$ の時 $E(X) = m$, $\mathrm{var}(X) = 2m$ である事を確かめなさい。ただし$E(Z^4) = 3$, $Z \sim N(0, 1)$を用いてよい。

単回帰モデル: $Y_i = \alpha + \beta X_i + \varepsilon_i$, $i = 1, \ldots, n$, において、$\hat{\beta}$ と $\hat{\sigma}_\beta^2$ をそれぞれ β の 最小二乗推定量とその標準誤差すなわち$\hat{\sigma}_\beta^2 = \hat{\sigma}^2 / \sum_{i=1}^n (X_i - \bar{X})^2$; ここで $\hat{\sigma}^2 = (n-2)^{-1} SSR_{UR}$, SSR_{UR}はこの(無制約)モデルからの残差平方和とする。

問題2　帰無仮説 H_0: $\beta = 0$ を検定する F 検定統計量は $F = \hat{\beta}^2 / \hat{\sigma}_\beta^2$ と表す事が出来ることを示しなさい。

問題3　問題2のF検定統計量は $F = \dfrac{SSR_R - SSR_{UR}}{SSR_{UR}/(n-2)}$ とも表せることを示しなさい。ここで SSR_Rは帰無仮説を制約条件とした回帰式 ($Y_i = \alpha + \varepsilon_i$) からの残差平方和とする。