統計解析論

その6
様々な仮説検定

1．母平均に関する検定

- 母平均がある値と等しくないことを示す．
- 帰無仮説：母平均＝ある値（μ_0）
- X_iの母平均と母分散　$E[X_i]=\mu \quad V[X_i]=\sigma^2$
- モデル

$$X_i=\mu+\varepsilon_i \qquad E[\varepsilon_i]=0 \qquad V[\varepsilon_i]=\sigma^2$$

誤差項ε_iは正規分布

- 回帰係数μに関する検定

母平均に関する検定（2）

- t統計量
  - 母平均の推定　$\hat{\mu}=\overline{X}$
  - 推定値の分散，標準誤差の推定

$$V(\hat{\mu})=\frac{\sigma^2}{\sum_{i=1}^{n}1^2}=\frac{\sigma^2}{n} \qquad 推定値(\sigma^2)=S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$s.e.(\hat{\mu})=\frac{S}{\sqrt{n}}$$

  - t統計量　$t=\frac{\hat{\mu}-\mu_0}{s.e.(\hat{\mu})}=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$ は $t(n-1)$ という分布
    - 誤差項が正規分布ではない場合もt分布を近似として使用

2．母平均差に関する検定

- 二つの集団の母平均に差があることを示す．
- 帰無仮説：2集団の母平均差は0
- もう一つ仮定
  - 2集団の母分散は等しい．つまり，帰無仮説では完全に同一の母集団に属する．
  - 母分散が異なる場合は次の誤差項の問題，異分散の場合で扱う．
- ダミー変数
  - あるカテゴリーに属する場合に値が1になり，そうでないときには0となる変数．
  - 質的変数をモデル化するのに使う．
    - $D1_i$は集団1に属している場合1，それ以外0．
    - $D2_i$は集団2に属している場合1，それ以外0．

母平均差に関する検定（2）

- $E[X_i|集団1]=\mu_1 \quad E[X_i|集団2]=\mu_2 \quad V[X_i]=\sigma^2$
- モデル　$X_i=\mu_1 D1_i+\mu_2 D2_i+\varepsilon_i$

  集団1の標本数：n_1，集団2の標本数：n_2
- 集団ごとの平均μ_1，μ_2の推定値

$$\hat{\mu}_1=\frac{1}{n_1}\sum_{集団1}^{n_1}X_i=\overline{X}_1 \qquad \hat{\mu}_2=\frac{1}{n_2}\sum_{集団2}^{n_2}X_i=\overline{X}_2$$

理由：正規方程式（$D_{1i}D_{2i}=0$を利用）

$$\sum_{i=1}^{n}(X_i-\mu_1 D1_i-\mu_2 D2_i)D1_i=\sum_{i=1}^{n}X_i D1_i-n_1\mu_1=\sum_{集団1}^{n_1}X_i-n_1\mu_1=0$$

$$\sum_{i=1}^{n}(X_i-\mu_1 D1_i-\mu_2 D2_i)D2_i=\sum_{i=1}^{n}X_i D2_i-n_2\mu_2=\sum_{集団2}^{n_2}X_i-n_2\mu_2=0$$

母平均差に関する検定（3）

- 母平均差に関するモデルへの変形

$$D1_i+D2_i=1 \qquad D2_i=1-D1_i$$

$$X_i=\mu_1 D1_i+\mu_2(1-D1_i)+\varepsilon_i=\mu_2+(\mu_1-\mu_2)D1_i+\varepsilon_i$$

- 推定値とその分散

$$\mu_1-\mu_2の推定値=\hat{\mu}_1-\hat{\mu}_2=\overline{X}_1-\overline{X}_2$$

$$V(\hat{\mu}_1-\hat{\mu}_2)=\frac{\sigma^2}{\sum_{i=1}^{n}(D1_i-\overline{D1})^2}=\frac{\sigma^2}{\sum_{i=1}^{n}\left(D1_i-\frac{n_1}{n_1+n_2}\right)^2}$$

$$=\frac{\sigma^2}{\sum_{集団1}^{n_1}\left(\frac{n_2}{n_1+n_2}\right)^2+\sum_{集団2}^{n_2}\left(\frac{n_1}{n_1+n_2}\right)^2}=\frac{\sigma^2}{\frac{n_1 n_2^2+n_1^2 n_2}{(n_1+n_2)^2}}=\frac{\sigma^2}{\frac{n_1 n_2}{n_1+n_2}}$$

母平均差に関する検定（4）

- 分散の推定量

$$RSS = \sum_{\text{集団1}}^{n_1}\left(X_i - \overline{X}_1\right)^2 + \sum_{\text{集団2}}^{n_2}\left(X_i - \overline{X}_2\right)^2$$

$$S^2 = \frac{RSS}{n_1 + n_2 - 2} = \frac{\sum_{\text{集団1}}^{n_1}\left(X_i - \overline{X}_1\right)^2 + \sum_{\text{集団2}}^{n_2}\left(X_i - \overline{X}_2\right)^2}{n_1 + n_2 - 2}$$

- t統計量 $t = \dfrac{\hat{\mu}_1 - \hat{\mu}_2}{S / \sqrt{n_1 n_2 / (n_1 + n_2)}}$

分布 $t(n_1 + n_2 - 2)$

3．複数の仮説の検定

- 帰無仮説が複数の式に関わる場合
- 例 $\beta_1 = \beta_1^0, \beta_2 = \beta_2^0$
- 対立モデル

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

このRSSの値を RSS_a とする．

- 帰無モデル

$$y_i - \beta_1^0 x_{1i} - \beta_2^0 x_{2i} = \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

このRSSの値を RSS_0 とする．

複数の係数からなる仮説の検定（2）

- 検定統計量

$$F = \frac{(RSS_0 - RSS_a)/(\text{制約式の数} = 2)}{RSS_a / (\text{標本数} - \text{対立仮説における係数の数})}$$

- 帰無仮説における統計量の分布

$F(\text{制約式の数} = 2, \text{標本数} - \text{対立モデルにおける係数の数})$

- 片側検定を用いる
  - 統計量の値が大きい範囲でのみ棄却

4．例1（収穫一定性の検定）

- 対立モデル（コブ－ダグラス型生産関数）

$$Y_i = \gamma + \alpha \log K + \beta \log L + \varepsilon_i$$

- 帰無仮説 $\alpha + \beta = 1$
  - この場合，証明したいことが帰無仮説に来ているので，あまり望ましくないが，やむを得ない．
- 帰無モデル

$$\begin{aligned}\log Y_i &= \gamma + \alpha \log K_i + (1-\alpha)\log L_i + \varepsilon_i \\ &= \log L_i + \gamma + \alpha(\log K_i - \log L_i) + \varepsilon_i\end{aligned}$$

より

$$\log Y_i - \log L_i = \gamma + \alpha(\log K_i - \log L_i) + \varepsilon_i$$

例1（収穫一定性の検定）（2）

- 統計量

$$F = \frac{(RSS_0 - RSS_a)/1}{RSS_a / (\text{標本数} - \text{対立仮説における係数の数})}$$

- 帰無仮説における統計量の分布

$F(\text{制約式の数} = 1, \text{標本数} - \text{対立モデルにおける係数の数})$

- F検定を使用しない方法

$$\log Y_i = \gamma + (\alpha + \beta - 1)\log K_i - (\beta - 1)\log K_i + \beta \log L_i + \varepsilon_i$$

$$\log Y_i - \log K_i = \gamma + (\alpha + \beta - 1)\log K_i + \beta(\log L_i - \log K_i) + \varepsilon_i$$

5．例2（多重比較）

- 3群以上の集団の平均比較
- 対立モデル

$$X_i = \mu_1 D1_i + \mu_2 D2_i + \cdots + \mu_k Dk_i + \varepsilon_i$$

  - $D1_i$は集団1に属している場合1，それ以外0．
  - $D2_i$は集団2に属している場合1，それ以外0．
  - 以下同文．．．．．

$E[X_i | \text{集団1}] = \mu_1 \quad E[X_i | \text{集団2}] = \mu_2 \quad E[X_i | \text{集団3}] = \mu_3$

- 帰無仮説 $\mu_1 = \mu_2 = \cdots = \mu_k$
- 帰無モデル $X_i = \mu + \varepsilon_i$

例2（多重比較）（2）

- F統計量

$$RSS_0 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

$$RSS_a = \sum_{\text{集団}1}^{n_1} \left(X_i - \overline{X}_1\right)^2 + \sum_{\text{集団}2}^{n_2} \left(X_i - \overline{X}_2\right)^2 + \cdots + \sum_{\text{集団}k}^{n_k} \left(X_i - \overline{X}_k\right)^2$$

$$RSS_0 - RSS_a = \sum_{\text{集団}1}^{n_1} \left(\overline{X} - \overline{X}_1\right)^2 + \sum_{\text{集団}2}^{n_2} \left(\overline{X} - \overline{X}_2\right)^2 + \cdots + \sum_{\text{集団}k}^{n_k} \left(\overline{X} - \overline{X}_k\right)^2$$

$$F = \frac{\left(RSS_0 - RSS_a\right)/(k-1)}{RSS_a/(n-k)}$$

- 統計量の分布　$F(k-1, n-k)$

6．回帰の有意性検定

- 回帰式中の定数項以外の係数のどれかが有意かどうか＝回帰に意味がある
  定数項は被説明変数の平均を多くは表すから，
  平均0でなければ有意になって当たり前．
- 帰無仮説 $\beta_2 = \beta_3 = \cdots = \beta_K = 0$
- 帰無モデル　$y_i = \beta_1 + \varepsilon_i$

$$RSS_0 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = TSS$$

- F統計量 $F = \frac{(RSS_0 - RSS_a)/(K-1)}{RSS_a/(n-K)} = \frac{(TSS - RSS_a)/(K-1)}{RSS_a/(n-K)}$
  - 分布はF（K－1，n－K）

7．t統計量とF統計量

- 1変数のみの回帰を考える．
- 回帰の有意性検定を$y_i = \alpha + \beta x_i + \varepsilon_i$に適用
  - 回帰の有意性検定は帰無仮説 $\beta = 0$ を検定

$$F = \frac{TSS - RSS}{RSS/(n-1)}$$

  - β=0はt検定でも検定できる．
  - 計算結果を見比べると，$t^2 = F$. 一般化すると，

$$t^2 = \frac{RSS(\text{t 値の回帰係数無モデル}) - RSS(\text{t 値の回帰係数有モデル}}{RSS(\text{t 値の回帰係数有モデル})/(n-K)}$$