統計解析論　その5

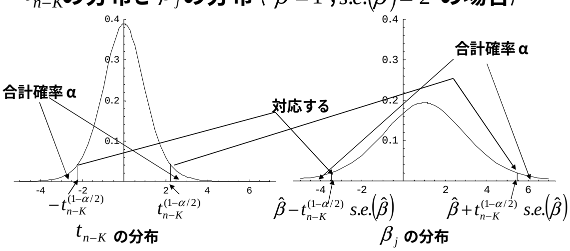
区間推定と検定の基本

1．係数の区間推定

- いままでの復習
  $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i$(誤差項)
  - 誤差項は平均0，分散一定
  - そのとき，$\dfrac{\hat{\beta}_j - \beta_j}{s.e.\left(\hat{\beta}_j\right)}$ は近似的に自由度n-Kのt分布に従う．
  - t_{n-K} を分布が自由度n-Kのt分布になるような変数とすると，$\dfrac{\hat{\beta}-\beta}{s.e.\left(\hat{\beta}\right)} = t_{n-K}$ とかける．
- 未知のパラメータ β_j を推定値 $\hat{\beta}_j$ から推測する．

2．区間推定（準備）

- β_jの表現　　$\beta = \hat{\beta} + t_{n-K}\ s.e.\left(\hat{\beta}\right)$
- t_{n-K}の分布と β_jの分布（ $\hat{\beta} = 1$ ，$s.e.\left(\hat{\beta}\right) = 2$ の場合）



2.区間推定

- (1-α)信頼区間
  - その範囲にパラメータが入る確率が1-αの区間
  - その下限　$\beta - t_{n-K}^{(1-\alpha/2)}\ s.e.\left(\hat{\beta}\right) = \beta + t_{n-K}^{(\alpha/2)}\ s.e.\left(\hat{\beta}\right)$
  - その上限　$\beta + t_{n-K}^{(1-\alpha/2)}\ s.e.\left(\hat{\beta}\right) = \beta - t_{n-K}^{(\alpha/2)}\ s.e.\left(\hat{\beta}\right)$
  - t分布の（α/2）分位点と(1-α/2）分位点
    $\Pr\left[t_{n-K} \le t_{n-K}^{(\alpha/2)}\right] = \alpha/2$　　$\Pr\left[t_{n-K} \ge t_{n-K}^{(\alpha/2)}\right] = 1-\alpha/2$
    $\Pr\left[t_{n-K} \ge t_{n-K}^{(1-\alpha/2)}\right] = \alpha/2$　　$\Pr\left[t_{n-K} \le t_{n-K}^{(1-\alpha/2)}\right] = 1-\alpha/2$
    $t_{n-K}^{(\alpha/2)} = -t_{n-K}^{(1-\alpha/2)}$

3．検定の基本的考え

- 検定は何を問題にしているか
  - パラメータβ_jの推定値がある値と異なっているときその異なり方が，本当にβ_jの値がその値と異なっているといえるほど十分大きいのか？
- 有意
  - 本当にある値と異なっているといえるほど十分大きい＝有意
  - 逆に小さいと偶然，誤差項の影響で異なっているだけと考える．

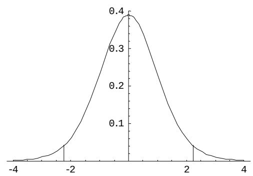
4．帰無仮説と対立仮説

- 有意性を考えるときは否定をまず考える．
  - 実際には異なってはいないと仮定するところから出発する．（言いたいことの否定を仮定することが多い）
  - この仮説を**帰無仮説**と呼ぶ．回帰係数に関しては，「βj＝ある値」が**帰無仮説**である．
- 帰無仮説が反証されたときにはその否定を受け入れる（受容する）．
  - これ以後は帰無仮説がデータによる反証を受ける．データが反証するに十分なものなら，その（一般的には）否定＝**対立仮説**を受け入れる．帰無仮説は棄却される．
  - 対立仮説は一般的には「βj≠ある値」となる

5．有意水準（その1）

- 「本当にある値と異なっている」といえるほど十分大きいとは？その基準＝反証の基準
- 差ではなくt統計量で比較する
- t統計量の分布は帰無仮説が正しければ，自由度n-Kのt分布である．
- 十分異なっている場合はt統計量の値は自由度n-Kのt分布のもとであまり起きそうもない値となる．

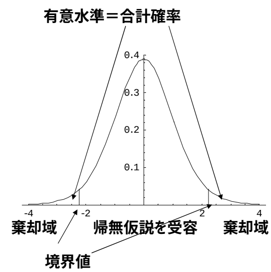
統計量の分布



5．有意水準（その2）

- 棄却域
  - 起きそうもない値と考える範囲（図の縦線の外側）
  - この線の外側の値では対立仮説を受け入れる．
- 有意水準
  - 外側の確率
  - 有意水準としては1%，5%が使われる．
- 境界値
  - 帰無仮説を受け入れる境界

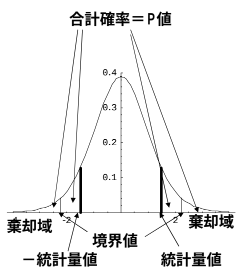
帰無仮説が正しいときの統計量の分布



6．P値

- 統計量の値を境界値とした検定を仮に考える．
- その検定の有意水準をP値と呼ぶ
- P値＜有意水準なら，帰無仮説が棄却され，対立仮説が採択される．
- P値≧有意水準なら，帰無仮説が受容される．

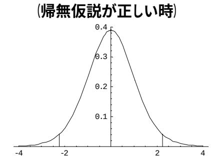
統計量の分布



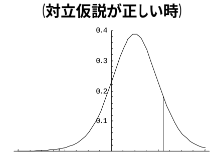
7．2つの過誤

- 第1種の過誤
  - 帰無仮説が正しいのに帰無仮説を棄却する．
  - 回帰では係数はある値なのに，その値ではないとすること．
  - 有意水準は第1種過誤の確率を示す．
- 第2種の過誤
  - 対立仮説が正しいのに帰無仮説を採択する．
  - 回帰では係数はその値ではないのに，その値だとしてしまうこと．
- 統計では第1種の過誤を小さくする慣行がある
  - 対立仮説を反証するために帰無仮説を設定することが多い．
  - 反証＝帰無仮説の採択の可能性を大きく取る．
  - 従って，有意水準として小さな値を使う．

統計量の分布
（帰無仮説が正しい時）



（対立仮説が正しい時）



8．片側検定

- 示したい仮説が「β」＜ある値」，「β」>ある値」のとき＝片側検定
  - 帰無仮説は「β」≦ある値」か「β」≧ある値」
  - このときt統計量の分布の計算は「β」＝ある値」で計算しても良いことがわかっている．
  - 従って，棄却域は対立仮説の側にだけ＝<u>片側</u>に取られる．
  - 同じ有意水準なら片側検定の境界値の絶対値は小さくなる
- 対立仮説が「β」≠ある値」のとき＝両側検定
  - 片側検定のα%有意水準での境界値は，2α%有意水準の両側検定での境界値

8．片側検定（図解）

$H_a(H_1\text{対立仮説}):\beta>\text{ある値}$

$H_0(\text{帰無仮説}):\beta\leq\text{ある値}$

または

$\beta=\text{ある値}$

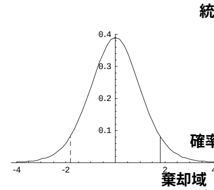
$H_a(H_1\text{対立仮説}):\beta<\text{ある値}$

$H_0(\text{帰無仮説}):\beta\geq\text{ある値}$

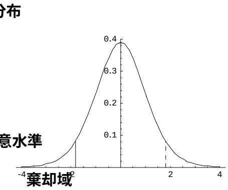
または

$\beta=\text{ある値}$

統計量の分布



確率α＝有意水準



9．検定の実際

- 帰無仮説が「β」＝０以外の値」の場合
  - t統計量を計算し直す．
  - Ｐ値を計算し直す．
  - 実例をよく見ておいてください．
- 片側検定の場合
  - Ｅｘｃｅｌの分析ツールの出力では，t検定のＰ値は両側検定で計算される．
  - 片側検定でのＰ値はExcelで出力されたＰ値の半分である．したがって，Ｐ値／２を有意水準と比較