統計解析論特殊講義　4

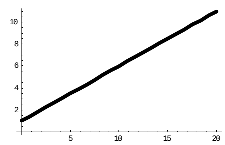
最小二乗法の統計学

1．最小二乗法がよく使われる理由

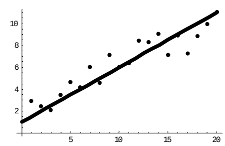
- 問題
  - 最小二乗推定量がよく使用されるのにはどんな理由があるのだろうか？どのような望ましい性質があるのか？
  - 「望ましい」ことの基準は何か？

1．0　考え方（1）

- データの成り立ち
  - 回帰直線を想定

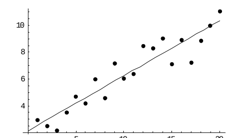


  - 回帰直線＋確率的に決まる誤差

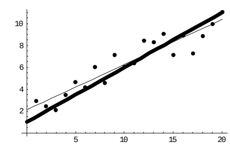


1．0　考え方（2）

- 最小二乗法による推定



- データの成り立ちの回帰直線と推定された回帰直線の比較



1．1　設定

- 今まではデータの値と最小二乗推定量との関係を論じてきた．
- 今後はデータの値だけではなく，データの成り立ちを考慮する．
  - データの成り立ち
    - 被説明変数＝説明変数の線形和(線形結合)＋誤差
    - 誤差の確率分布＝各データ毎に確率的に決まるので異なる
      - 普通は平均0，分散σ^2（一定）と仮定
      - さらに特別な場合は誤差が正規分布に従うと仮定
      - i番目のデータとj番目のデータは全く関連がない＝独立．
        - 無作為抽出を想定
      - 説明変数Xは確率変動しない．

確率の復習（1．1用＜1＞）

- 確率変数
  - 確率的に決まる変数のこと
- 確率変数の平均＝分布の平均＝母平均
  - 期待値のこと
  - 確率変数の中心を表す
  - E［確率変数］と表記
- 分散
  - （確率変数ーその平均）の二乗の期待値
  - 確率変数の散らばり方を示す
  - V［確率変数］と表記

１．１．１　設定の数式表現

- ＤＧＰ (Data Generating Process)
  - データ発生の仕組み
  - 被説明変数＝説明変数の線形和(線形結合)＋誤差
  - 数式化（K個の説明変数の場合）
    - 変数の場合
      $Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$
    - データの値の場合
      $y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i} + \varepsilon_i$
    - 誤差項　$\varepsilon, \varepsilon_i$（残差項 $\hat{u}_i = y_i - \left(\hat{\beta}_1 x_{1,i} + \cdots + \hat{\beta}_K x_{K,i}\right)$ と区別すること）
      - 平均0　$E[\varepsilon] = E[\varepsilon_i] = 0$
      - 分散一定値　σ^2
      - 独立性＝iに無関係に決まる＋Xは確率変動しないのでXと独立

１．１．２　注意

- ＤＧＰと回帰モデルの違い
  - ＤＧＰはデータの決まり方を示す．
    - 理論によって与えられる．
    - 未知なのはβjのパラメータの値（真の値と呼ぶ）である．
      - 真の値は確率変動しない定数
    - 理論がないケースでは，本当はわからない．
      - その場合一つの想定に過ぎない．
  - 回帰モデル
    - 最小自乗法によって推定する回帰式
      - この式の誤差の二乗和を最小化するパラメータを決める．
    - 回帰モデルとＤＧＰが同じ式であると想定する．
    - 異なった場合を考えることもある．（今回は考えないが）

１．２　DGPの変形

- $y_i = \beta_1 x_{1,i} + \cdots + \beta_K x_{K,i} + \varepsilon_i$ の変形

$$y_i = y_i^* + proj\left(y_i \middle| X_j\text{以外の説明変数}\right)$$

などより

$$\begin{aligned} y_i^* + proj\left(y_i \middle| X_j\text{以外の説明変数}\right) &= \beta_1 x_{1,i} + \cdots + \beta_{j-1} x_{j-1,i} \\ &+ \beta_j \left(x_{j,i}^* + proj\left(x_{j,i} \middle| X_j\text{以外の説明変数}\right)\right) \\ &+ \beta_{j+1} x_{j+1,i} + \cdots + \beta_K x_{K,i} + \varepsilon_i \end{aligned}$$

ここからさらに変形して

$$\begin{aligned} y_i^* = \beta_j x_{j,i}^* + \{&\beta_1 x_{1,i} + \cdots + \beta_{j-1} x_{j-1,i} + \beta_{j+1} x_{j+1,i} + \cdots + \beta_K x_{K,i} \\ &- proj\left(y_i \middle| X_j\text{以外の説明変数}\right) + \beta_j proj\left(x_{j,i} \middle| X_j\text{以外の説明変数}\right)\} \\ &+ \varepsilon_i \end{aligned}$$

１．３　最小二乗推定量と真の値の関係

- 最小二乗推定量を真の値と誤差項で表す．

$$\hat{\beta}_j = \sum_{i=1}^n x_{j,i}^* y_i^* \Big/ \sum_{i=1}^n \left(x_{j,i}^*\right)^2$$

これに

$$\begin{aligned} y_i^* = \beta_j x_{j,i}^* + \{&\beta_1 x_{1,i} + \cdots + \beta_{j-1} x_{j-1,i} + \beta_{j+1} x_{j+1,i} + \cdots + \beta_K x_{K,i} \\ &- proj\left(y_i \middle| X_j\text{以外の説明変数}\right) + \beta_j proj\left(x_{j,i} \middle| X_j\text{以外の説明変数}\right)\} + \varepsilon_i \end{aligned}$$

を代入すると

$$\begin{aligned} \hat{\beta}_j &= \sum_{i=1}^n x_{j,i}^* y_i^* \Big/ \sum_{i=1}^n \left(x_{j,i}^*\right)^2 \\ &= \sum_{i=1}^n \beta_j \left(x_{j,i}^*\right)^2 \Big/ \sum_{i=1}^n \left(x_{j,i}^*\right)^2 + \sum_{i=1}^n x_{j,i}^* \varepsilon_i \Big/ \sum_{i=1}^n \left(x_{j,i}^*\right)^2 \\ &= \beta_j + \sum_{i=1}^n \left(x_{j,i}^* \Big/ \sum_{i=1}^n \left(x_{j,i}^*\right)^2 \right) \varepsilon_i \end{aligned}$$

１．３．１　補足説明

- 中括弧と $x_{j,i}^*$ の積の合計が０になる理由

$$\begin{aligned} \sum_{i=1}^n \{&\beta_1 x_{1,i} + \cdots + \beta_{j-1} x_{j-1,i} + \beta_{j+1} x_{j+1,i} + \cdots + \beta_K x_{K,i} \\ &- proj\left(y_i \middle| X_j\text{以外の説明変数}\right) \\ &+ \beta_j proj\left(x_{j,i} \middle| X_j\text{以外の説明変数}\right)\} x_{j,i}^* = 0 \end{aligned}$$

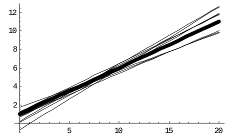
  - なぜならば，$x_{j,i}^*$ は X_j を X_j 以外の説明変数に回帰した残差だから，残差と回帰時の説明変数の積の合計は０となるから．

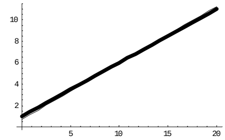
１．４　不偏性（０）

- 望ましい性質（１）＝不偏性
  - 推定値と真の値との間に何か関係があるか？
    - 推定値の期待値が真の値になるという性質
    - 推定値が平均的には真の値になっている．
  - 数式表現
    $E\left[\hat{\beta}_1\right] = \beta_1, \cdots, E\left[\hat{\beta}_K\right] = \beta_K$
  - 最小二乗推定量は不偏である．

1．4 不偏性(1)

- 実験
  - 誤差項を10セット用意．
  - これを元に説明変数と被説明変数の組み合わせを10セット用意．
  - この10セットそれぞれについて最小二乗法でパラメータを推定
  - 上図
    - 太線：真の回帰直線群，細線：推定された回帰直線
  - 下図
    - 細線：推定された回帰直線の平均
    - 真の回帰直線とほとんど同じである．





確率の復習(1．4用)

- 期待値に関する演算
  - 確率変数の線形和(一次結合)の期待値は確率変数の期待値の線形和である．

$$E[c_1\varepsilon_1 + c_2\varepsilon_2] = c_1E[\varepsilon_1] + c_2E[\varepsilon_2]$$

$$E\left[\sum_{i=1}^{n} c_i\varepsilon_i\right] = \sum_{i=1}^{n} c_iE[\varepsilon_i]$$

    - ただし，$c_1, c_2, \cdots, c_n$は確率変動しない定数

1．4 不偏性(2)

- 不偏性の説明
  - 最小二乗推定量と真の値の関係から

$$E[\hat{\beta}_j] = E[\beta_j] + E\left[\sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)\varepsilon_i\right]$$

  - β jは確率変動しないので $E[\beta_j] = \beta_j$．
  - 説明変数は確率変動しないと仮定しているので

$$E\left[\sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)\varepsilon_i\right] = \sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)E[\varepsilon_i]$$

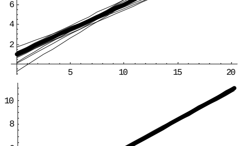
  - $E[\varepsilon_i] = 0$なので $E\left[\sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)\varepsilon_i\right] = 0$ より $E[\hat{\beta}_j] = \beta_j$

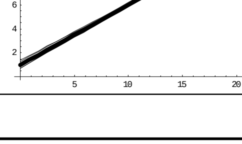
1．5 一致性(0)

- 不偏性で十分か？
  - 半々の確率で1と-1の値をとる確率変数の期待値は0
    - 推定値の期待値が真の値と等しくても推定値が真の値に近づくという保証はない．(標本数が増えたとき)
  - 「近づく」というのにふさわしい概念が必要．
- 望ましい性質(2)＝一致性
  - 標本数が増えると推定誤差がどんどん小さくなる．
    - 推定誤差は確率変動するから小さくなるというのをうまく表現しなければならない．
  - 標本数が増えると「推定誤差の2乗の期待値」＝平均二乗誤差が0に近づく．平均二乗誤差の極限が0．

1．5 一致性(1)

- 実験
  - 同じデータの成り立ちの場合についてデータ数が増加したときの推定された回帰直線を比較
  - 上図
    - 標本数20の場合の推定された回帰直線群(細線)と真の回帰直線群(太線)
  - 下図
    - 標本数1000の場合
    - 太線と細線がほぼ一致





1．5 一致性(2)

- 推定誤差

$$\hat{\beta}_j = \beta_j + \sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)\varepsilon_i \text{ より}$$

$$\hat{\beta}_j - \beta_j = \sum_{i=1}^{n}\left(x_{j,i}^{*} \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2\right)\varepsilon_i$$

- 平均二乗誤差＝推定誤差の分散(最小二乗法の場合)

$\beta_j = E[\hat{\beta}_j]$ より $\hat{\beta}_j - \beta_j = \hat{\beta}_j - E[\hat{\beta}_j]$ となり

$$E\left[\left(\hat{\beta}_j - \beta_j\right)^2\right] = E\left[\left(\hat{\beta}_j - E[\hat{\beta}_j]\right)^2\right] = V(\hat{\beta}_j)$$

確率の復習（1．5用）

- 分散に関する演算
  - もし確率変数同士が独立ならば，確率変数の線形和（一次結合）の分散は確率変数の分散の線形和である．
    - 期待値と違って分散の係数は確率変数の係数の2乗である．

$$V[c_1\varepsilon_1 + c_2\varepsilon_2] = c_1^2 V[\varepsilon_1] + c_2^2 V[\varepsilon_2]$$

$$V\left[\sum_{i=1}^{n} c_i\varepsilon_i\right] = \sum_{i=1}^{n} c_i^2 V[\varepsilon_i]$$

    - ただし，$c_1, c_2, \cdots, c_n$は確率変動しない定数で，各 ε_i は独立．

1．5　一致性（3）

- 一致性の証明（1）
  - $w_i = x_{j,i}^* \Big/ \sum_{i=1}^{n}(x_{j,i}^*)^2$ とおくと $\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_i\varepsilon_i$
  - 各 ε_i は独立だから，

$$E\left[(\hat{\beta}_j - \beta_j)^2\right] = V[\hat{\beta}_j] = V[\beta_j] + \sum_{i=1}^{n} w_i^2 V[\varepsilon_i]$$

  - $V[\beta_j] = 0$，$V[\varepsilon_i] = \sigma^2$ より

$$\sum_{i=1}^{n} w_i^2 V[\varepsilon_i] = \sum_{i=1}^{n} w_i^2\sigma^2 = \sigma^2 \sum_{i=1}^{n} w_i^2 = \sigma^2 \sum_{i=1}^{n}(x_{j,i}^*)^2 \Big/ \left(\sum_{i=1}^{n}(x_{j,i}^*)^2\right)^2$$

$$= \sigma^2 \Big/ \sum_{i=1}^{n}(x_{j,i}^*)^2$$

  - よって，$E\left[(\hat{\beta}_j - \beta_j)^2\right] = V[\hat{\beta}_j] = \sigma^2 \Big/ \sum_{i=1}^{n}(x_{j,i}^*)^2$

1．5　一致性（4）

- 一致性の証明（2）
  - 仮定

$$\lim_{n\to\infty}\sum_{i=1}^{n}(x_{j,i}^*)^2 = \infty$$

  - 一致性

$$\lim_{n\to\infty} E\left[(\hat{\beta}_j - \beta_j)^2\right] = \lim_{n\to\infty}\left\{\sigma^2 \Big/ \sum_{i=1}^{n}(x_{j,i}^*)^2\right\} = 0$$

  - 最小二乗推定量の分散の極限

$$\lim_{n\to\infty} V[\hat{\beta}_j] = 0$$

1．5　一致性（5）

- 仮定の意味
  - $\lim_{n\to\infty}\sum_{i=1}^{n}(x_{j,i}^*)^2 = \infty$ を満たす必要条件
    - $\lim_{n\to\infty}\sum_{i=1}^{n}(x_{j,i}^*)^2$ が有限である為の必要条件は $\lim_{n\to\infty}(x_{j,n}^*)^2 = 0$
    - 逆に $\lim_{n\to\infty}(x_{j,n}^*)^2 > 0$ なら一致性の条件を満たす．
      - 例　$x_{j,i}^* = \alpha i + \beta,\quad \lim_{n\to\infty} x_{j,n}^* = \alpha$
    - $\lim_{n\to\infty}(x_{j,n}^*)^2 = 0$ でも一致性を持つ例

$$\lim_{n\to\infty}(x_{j,i}^*)^2 = \frac{\alpha}{i^{2-\beta}} \quad \beta > 0$$

1．6　最良線形不偏推定量

- 不偏性，一致性で十分望ましいか？
  - 他の種類の推定方法と比較してどうか？
    - 実は他の推定量の中に不偏性，一致性を満たすものがある．
  - 最小二乗推定量が優位に立つ切り札はなにか？
- 効率性
  - 効率性とは，最小の分散をもつこと．
    - 分散が小さいほど，推定量のばらつきが少ない＝より正確．
  - 不偏性をもち，被説明変数の観測値の線形和（一次結合）で表現される推定量のなかで最も分散が小さい＝効率的なのが最小二乗推定量である．

2．最小二乗推定量の分布（0）

- 問題
  - 最小二乗推定量の信頼度を知る
    - その中心とばらつきを知る
      - 期待値と分散は？
  - 信頼区間，検定ができるか？
    - そのための仮定は？

2．最小二乗推定量の分布 (0)

- **期待値**
  - 不偏性の議論から

$$E\left[\hat{\beta}_1\right]=\beta_1,\cdots,E\left[\hat{\beta}_K\right]=\beta_K$$

  - 真のパラメータの値が期待値
- **分散**
  - 一致性の議論から

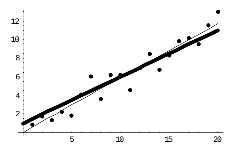
$$E\left[\left(\hat{\beta}_j-\beta_j\right)^2\right]=V\left[\hat{\beta}_j\right]=\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2$$

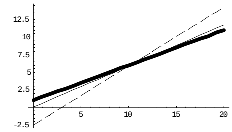
  - 単回帰の場合は，

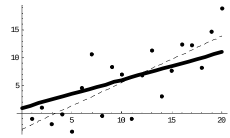
$$V\left[\hat{\beta}\right]=\sigma^2\Big/\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\sigma^2\Big/\left[n\left\{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2/n\right\}\right]$$

2．最小二乗推定量の分布 (1)

- 分散に関する実験 (1)
  - 誤差項の分散の推定量の分散に及ぼす効果
    - データの値：点
    - 真の回帰直線：太線
    - 推定された回帰直線：細線
  - 右上図　誤差分散　1
  - 右下図　誤差分散　16

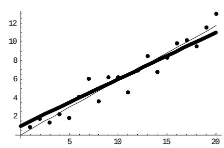


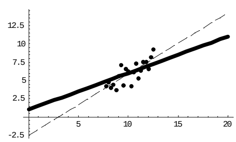




2．最小二乗推定量の分布 (2)

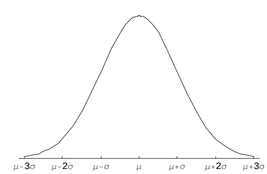
- 分散に関する実験 (2)
  - 説明変数の散らばり方＝標本分散が推定量の分散に及ぼす効果
    - 誤差の分散は標本数は同じ
    - データの値：点
    - 真の回帰直線：太線
    - 推定された回帰直線：細線
- 他は同じで標本数が多くなる場合
  - 標本数に反比例





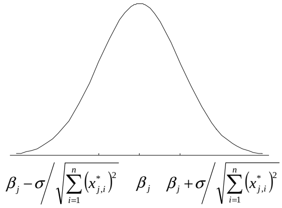
2．最小二乗推定量の分布 (3)

- 追加の仮定
  - 誤差項の分布が平均0，分散σ^2の正規分布
- 確率の復習
  - 正規分布
    - 平均μ，分散σ^2の正規分布の確率密度関数は右図
    - 確率密度関数とはある区間内の確率をその区間の長さで割ったもので，確率の密度を表す．ただし，ある点での密度を表す．



2．最小二乗推定量の分布 (4)

- 最小二乗推定量の分布
  - 平均が真の値βj，分散が $\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2$ の正規分布
  - さらに，もし誤差項が正規分布に従っていなくても，適当な条件を満たしていれば，中心極限定理から，標本数nが大きくなれば，この分布で近似できる．
    - 近似できる目安はnが30以上といわれている．



3．係数推定値同士の相関 (0)

- **問題**
  - 異なった係数の推定値同士は関連があるのか？
    - 相関，共分散は？
  - 異なった係数の推定値同士の相関がなくなるのはどういう場合か？

3．係数推定値同士の相関（1）

- **確率復習（1）**
  - XとYの（分布，母）共分散
    - 定義　$Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$
    - 性質
      - $Cov(X,X)=V(X),\ Cov(Y,Y)=V(Y)$
      - $Cov(aX+bY,cX+dY)$
        $=acCov(X,X)+(ad+bc)Cov(X,Y)+bdCov(Y,Y)$
        $=acV(X)+(ad+bc)Cov(X,Y)+bdV(Y)$
      - $Cov\left[\sum_{i=1}^{n}v_i\varepsilon_i,\sum_{i=1}^{n}w_i\varepsilon_i\right]=\sum_{i=1}^{n}v_iw_iV[\varepsilon_i]+\sum_{i\neq j}v_iw_jCov[\varepsilon_i,\varepsilon_j]$

3．係数推定値同士の相関（2）

- **確率の復習（2）**
  - もしε_i が独立で同一の分散 σ^2を持つなら，

$$Cov\left[\sum_{i=1}^{n}v_i\varepsilon_i,\sum_{i=1}^{n}w_i\varepsilon_i\right]=\sum_{i=1}^{n}v_iw_iV[\varepsilon_i]+\sum_{i\neq j}v_iw_jCov[\varepsilon_i,\varepsilon_j]$$

$$=\sigma^2\sum_{i=1}^{n}v_iw_i$$

  - XとYの（分布，母）相関
    - 定義　$\rho_{X,Y}=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$
    - 無相関　$Cov(X,Y)=0$　$\rho_{X,Y}=0$　（独立ならば無相関）

3．係数推定値同士の相関（3）

- **βj，βkの推定量同士の相関**
  - 共分散

$$Cov(\hat\beta_j,\hat\beta_k)=Cov\left[\sum_{i=1}^{n}\left\{x^*_{j,i}\Big/\sum_{i=1}^{n}(x^*_{j,i})^2\right\}\varepsilon_i,\sum_{i=1}^{n}\left\{x^*_{k,i}\Big/\sum_{i=1}^{n}(x^*_{k,i})^2\right\}\varepsilon_i\right]$$

$$=\sum_{i=1}^{n}\left\{x^*_{j,i}\Big/\sum_{i=1}^{n}(x^*_{j,i})^2\right\}\left\{x^*_{k,i}\Big/\sum_{i=1}^{n}(x^*_{k,i})^2\right\}Cov(\varepsilon_i,\varepsilon_i)+$$

$$\sum_{i\neq m}^{n}\left\{x^*_{j,i}\Big/\sum_{i=1}^{n}(x^*_{j,i})^2\right\}\left\{x^*_{k,m}\Big/\sum_{i=1}^{n}(x^*_{k,i})^2\right\}Cov(\varepsilon_i,\varepsilon_m)$$

$Cov(\varepsilon_i,\varepsilon_i)=V(\varepsilon_i)=\sigma^2$ と誤差項の独立性から$Cov(\varepsilon_i,\varepsilon_m)=0$とにより

$$Cov(\hat\beta_j,\hat\beta_k)=\sum_{i=1}^{n}\left\{x^*_{j,i}x^*_{k,i}\Big/\left(\sum_{i=1}^{n}(x^*_{j,i})^2\sum_{i=1}^{n}(x^*_{k,i})^2\right)\right\}\sigma^2$$

$$=\sigma^2\left(\sum_{i=1}^{n}x^*_{j,i}x^*_{k,i}\right)\Big/\left(\sum_{i=1}^{n}(x^*_{j,i})^2\sum_{i=1}^{n}(x^*_{k,i})^2\right)$$

3．係数推定値同士の相関（3）

- **係数推定値同士の相関**

$$\rho_{\hat\beta_j,\hat\beta_k}=\frac{\sigma^2\left(\sum_{i=1}^{n}x^*_{j,i}x^*_{k,i}\right)\Big/\left(\sum_{i=1}^{n}(x^*_{j,i})^2\sum_{i=1}^{n}(x^*_{k,i})^2\right)}{\sqrt{\sigma^2\Big/\sum_{i=1}^{n}(x^*_{j,i})^2}\sqrt{\sigma^2\Big/\sum_{i=1}^{n}(x^*_{k,i})^2}}$$

$$=\sum_{i=1}^{n}x^*_{j,i}x^*_{k,i}\Big/\left(\sqrt{\sum_{i=1}^{n}(x^*_{j,i})^2}\sqrt{\sum_{i=1}^{n}(x^*_{k,i})^2}\right)$$

- **メッセージ**
  - もしすべての説明変数が標本として無相関ならば，係数推定値どうしは分布として無相関

3．係数推定値同士の相関（4）

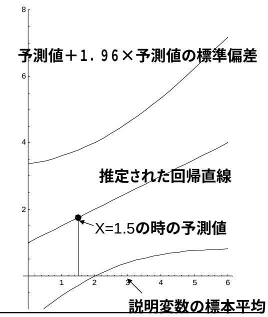
- **メッセージ**
    - どの説明変数同士を掛けて合計しても0になる（積和をとると0，直交する，$\sum_{i=1}^{n}x_{j,i}x_{k,i}=0$）ならば，係数推定値同士は分布として無相関
  - 理由
    - どの説明変数同士を掛けて合計しても0なら，説明変数同士は影響を与え合わないから，他の説明変数の影響を取り除いても変わらない（$x^*_{j,i}=x_{j,i},x^*_{k,i}=x_{k,i}$）
    - $Cov[\hat\beta_j,\hat\beta_k]=\frac{\sigma^2\sum_{i=1}^{n}x^*_{j,i}x^*_{k,i}}{\sum_{i=1}^{n}(x^*_{j,i})^2\sum_{i=1}^{n}(x^*_{k,i})^2}=\frac{\sigma^2\sum_{i=1}^{n}x_{j,i}x_{k,i}}{\sum_{i=1}^{n}x^2_{j,i}\sum_{i=1}^{n}x^2_{k,i}}=0$

4．予測

- **問題**
  - 説明変数の値を色々と想定して，その値に対する被説明変数の値を予測したい．
    - どの様な予測値が考えられるか？
    - 予測値の持つ望ましい性質は？
    - 予測値の信頼性は？
      - 予測値の分散は？
        - パラメータの推定誤差と誤差項の両方を考える必要がある．

４．単回帰での予測（1）

- グラフ例
  - 標本数=20
  - 推定された回帰式
  - 予測値の９５%信頼区間を描いた．
    - （予測値＋1．96×予測値の標準偏差）と（予測値－1．96×予測値の標準偏差）の間
    - この間に予測値がくる確率は９５%
  - 信頼区間が最も狭いのは説明変数の標本平均のところ



４．１　単回帰での予測（2）

- 例の設定（単回帰）
  - 回帰モデル $Y=\alpha+\beta X+\varepsilon$ を考える．（DGPも一致）
  - これと同じ推定量を与える回帰モデルは
    $Y-\bar{y}=\hat{\beta}(X-\bar{x})+\text{残差}$ または$Y=\bar{y}+\hat{\beta}(X-\bar{x})+\text{残差}$
  - 第２式に対応する回帰モデルを考える．
    $Y=\delta+\beta(X-\bar{x})+\varepsilon \qquad \hat{\delta}=\bar{y}$
  - 元のDGPを変形すると$Y=\alpha+\beta\bar{x}+\beta(X-\bar{x})+\varepsilon$
  - これより変形した回帰式に対応するDGPでは，
    $\delta=\alpha+\beta\,\bar{x}$

４．１　単回帰での予測（3）

- 推定値の誤差の表現
  - $Y=\alpha+\beta X+\varepsilon$ より $\hat{\delta}=\bar{y}=\alpha+\beta\,\bar{x}+\bar{\varepsilon}$ であるから，
    $\delta-\hat{\delta}=(\alpha+\beta\,\bar{x})-(\alpha+\beta\,\bar{x}+\bar{\varepsilon})=-\bar{\varepsilon}$
  - $\hat{\beta}-\beta=\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2$
- X=xのときの予測値　$\hat{y}=\hat{\delta}+\hat{\beta}(x-\bar{x})$
- 予測誤差
  - $y-\hat{y}=(\delta-\hat{\delta})+(\beta-\hat{\beta})(x-\bar{x})+\varepsilon_{\text{次の観測時}}$

$$=-\bar{\varepsilon}+(x-\bar{x})\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\right)\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2+\varepsilon_{\text{次の観測時}}$$

４．１　単回帰での予測（4）

- 予測誤差の期待値は０
  - $E[y-\hat{y}]=E(\delta-\hat{\delta})+(x-\bar{x})E(\beta-\hat{\beta})+E[\varepsilon_{\text{次の観測時}}]$
    - 不偏性より $E[\hat{\delta}-\delta]=E[\hat{\beta}-\beta]=0$
    - 誤差項は期待値０なので $E[\varepsilon_{\text{次の観測時}}]=0$

    よって $E[y-\hat{y}]=0$
- 予測誤差の分散＜$\hat{\delta}$と$\hat{\beta}$との相関が０になる＞

$$V[y-\hat{y}]=V[-\bar{\varepsilon}]+(x-\bar{x})^2V[\hat{\beta}]+V[\varepsilon_{\text{次の観測時}}]+2(x-\bar{x})Cov[\hat{\delta},\hat{\beta}]$$

$$=\sigma^2/n+(x-\bar{x})^2\sigma^2/\sum_{i=1}^{n}(x_i-\bar{x})^2+\sigma^2$$

$$=\sigma^2\left(1/n+(x-\bar{x})^2/\sum_{i=1}^{n}(x_i-\bar{x})^2+1\right)$$

４．２　予測の一般論（1）

- 予測
  - 説明変数が特定の値をとったとしたときに，被説明変数の値を予測する．
  - 予測値は回帰値となる．　$\hat{y}_{X_1=x_1,\cdots,X_K=x_k}=\sum_{j=1}^{K}\hat{\beta}_j x_j$
- 予測値の期待値
  - $E[\hat{y}_{X_1=x_1,\cdots,X_K=x_k}]=E\left[\sum_{j=1}^{K}\hat{\beta}_j x_j\right]=\sum_{j=1}^{K}E[\hat{\beta}_j]x_j$
  - 不偏性より $E[\hat{\beta}_j]=\beta_j$ だから $E[\hat{y}_{X_1=x_1,\cdots,X_K=x_k}]=\sum_{j=1}^{K}\beta_j x_j$
  - $y_{X_1=x_1,\cdots,X_K=x_k}=\sum_{j=1}^{K}\beta_j x_j+\varepsilon$ より $E[y_{X_1=x_1,\cdots,X_K=x_k}]=\sum_{j=1}^{K}\beta_j x_j+E[\varepsilon]=\sum_{j=1}^{K}\beta_j x_j$
  - したがって，$E[\hat{y}_{X_1=x_1,\cdots,X_K=x_k}]=\sum_{j=1}^{K}\beta_j x_j=E[y_{X_1=x_1,\cdots,X_K=x_k}]$

４．２　予測の一般論（2）

- 予測値の分散

$$V[\hat{y}_{X_1=x_1,\cdots,X_K=x_k}]=V\left[\sum_{j=1}^{K}\hat{\beta}_j x_j\right]=Cov\left[\sum_{j=1}^{K}\hat{\beta}_j x_j,\sum_{j=1}^{K}\hat{\beta}_j x_j\right]$$

$$=\sum_{j=1}^{n}x_j^2Cov[\hat{\beta}_j,\hat{\beta}_j]+\sum_{j\neq k}x_j x_k Cov[\hat{\beta}_j,\hat{\beta}_k]$$

$$=\sum_{j=1}^{n}x_j^2V[\hat{\beta}_j]+\sum_{j\neq k}x_j x_k Cov[\hat{\beta}_j,\hat{\beta}_k]$$

5．誤差項の分散の推定

- **問題**
  - 推定値の分散 $\sigma^2 \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2$ を知るには誤差項の分散を知る必要がある．
  - 誤差項の分散の推定方法は？
    - 残差と誤差は違うが，残差から誤差を推定できないか？
    - 残差の標本分散で推定できないか？
      - できないなら，それを修正することで対応できないか？
    - どういう基準を推定値にもとめるか？
      - 一致性？不偏性？

5．1　残差二乗和の利用(1)

- **単回帰の例(1)**

$Y=\delta+\beta(X-\bar{x})+\varepsilon$ を考える．

$$\hat{u}_i = y_i-\hat{y}_i=\left\{\delta+\beta(x_i-\bar{x})+\varepsilon_i\right\}-\left\{\hat{\delta}+\hat{\beta}(x_i-\bar{x})\right\}$$
$$=\left(\delta-\hat{\delta}\right)+\left(\beta-\hat{\beta}\right)(x_i-\bar{x})+\varepsilon_i$$
$$\varepsilon_i=\left(\hat{\delta}-\delta\right)+\left(\hat{\beta}-\beta\right)(x_i-\bar{x})+\hat{u}_i$$
$$\sum_{i=1}^{n}\varepsilon_i^2=\left(\hat{\delta}-\delta\right)^2 n+\left(\hat{\beta}-\beta\right)^2\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}\hat{u}_i^2$$
$$+2\left(\hat{\delta}-\delta\right)\sum_{i=1}^{n}\hat{u}_i+2\left(\hat{\beta}-\beta\right)\sum_{i=1}^{n}(x_i-\bar{x})\hat{u}_i$$

5．1　残差二乗和の利用(2)

- **単回帰の例(2)**
  - 残差と説明変数の関係から $\sum_{i=1}^{n}\hat{u}_i=0$　$\sum_{i=1}^{n}(x_i-\bar{x})\hat{u}_i=0$

よって $\sum_{i=1}^{n}\varepsilon_i^2=\left(\hat{\delta}-\delta\right)^2 n+\left(\hat{\beta}-\beta\right)^2\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}\hat{u}_i^2$

$$E\left[\sum_{i=1}^{n}\varepsilon_i^2\right]=E\left[\left(\hat{\delta}-\delta\right)^2\right]n+E\left[\left(\hat{\beta}-\beta\right)^2\right]\sum_{i=1}^{n}(x_i-\bar{x})^2+E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]$$
$$E\left[\left(\hat{\delta}-\delta\right)^2\right]=E\left[\left(\hat{\delta}-E\left[\hat{\delta}\right]\right)^2\right]=V\left[\hat{\delta}\right]=V[\bar{\varepsilon}]=\sigma^2/n$$
$$E\left[\left(\hat{\beta}-\beta\right)^2\right]=E\left[\left(\hat{\beta}-E\left[\hat{\beta}\right]\right)^2\right]=V\left[\hat{\beta}\right]=\sigma^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2$$

5．1　残差二乗和の利用(3)

- **単回帰の例(3)**

$$E\left[\sum_{i=1}^{n}\varepsilon_i^2\right]=\sum_{i=1}^{n}E\left[\varepsilon_i^2\right]=\sum_{i=1}^{n}E\left[\left(\varepsilon_i-E[\varepsilon_i]\right)^2\right]=\sum_{i=1}^{n}V(\varepsilon_i)=n\sigma^2$$

  - これらから $n\sigma^2=\left(\sigma^2+\sigma^2\right)+E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]$
  - したがって $E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]=(n-2)\sigma^2$
- **一般の場合(K個の説明変数がある場合)**
  - $n\sigma^2=K\sigma^2+E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]$ より $E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]=(n-K)\sigma^2$

5．1　残差二乗和の利用(5)

- **誤差分散の不偏推定量**
  - $E\left[\sum_{i=1}^{n}\hat{u}_i^2\Big/(n-K)\right]=\sigma^2$ より $\sum_{i=1}^{n}u_i^2\Big/(n-K)$
- **1に回帰して母平均を推定した場合** $Y=\mu+\varepsilon$
  - 説明変数は1個なので，誤差分散の不偏推定量は

    $\sum_{i=1}^{n}\hat{u}_i^2\Big/(n-1)$　　$\hat{u}_i=y_i-\bar{y}$
  - $V[Y]=V[\varepsilon]=\sigma^2$ で誤差分散は分散そのもの．つまり，分散の不偏推定量＝標本不偏分散

5．1　残差二乗和の利用(6)

- **確率の復習**
  - 大数の法則
    - ある適当な条件の下で，標本数が大きくなると，ある変数の標本平均はその変数の期待値に近づいていく．

      $$\frac{1}{n}\sum_{i=1}^{n}Z_i\xrightarrow{n\to\infty}E[Z]$$
    - 大数の法則(別バージョン)

      $$\frac{1}{n}\sum_{i=1}^{n}c_iZ_i\xrightarrow{n\to\infty}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}c_i\right)E[Z]$$

5．1　残差二乗和の利用（7）

- **一致性（1）**
  - $\sum_{i=1}^{n}\varepsilon_i^2=\left(\hat{\delta}-\delta\right)^2 n+\sum_{i=1}^{n}\left(\hat{\beta}-\beta\right)^2\left(x_i-\bar{x}\right)^2+\sum_{i=1}^{n}\hat{u}_i^2$ **より**

    $\sum_{i=1}^{n}\varepsilon_i^2\Big/n=\left(\hat{\delta}-\delta\right)^2+\sum_{i=1}^{n}\left(\hat{\beta}-\beta\right)^2\left(x_i-\bar{x}\right)^2\Big/n+\sum_{i=1}^{n}\hat{u}_i^2\Big/n$
  - **大数の法則よりnが大きくなると**

    $\hat{\delta}-\delta=\bar{\varepsilon}\longrightarrow E[\varepsilon]=0$

    $\sum_{i=1}^{n}\left(\hat{\beta}-\beta\right)^2\left(x_i-\bar{x}\right)^2\Big/n\longrightarrow\lim_{n\to\infty}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)E\left[\left(\hat{\beta}-\beta\right)^2\right]\right\}$

    $=\lim_{n\to\infty}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)\left(\sigma^2\Big/\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)\right\}=\lim_{n\to\infty}\sigma^2/n=0$

5．1　残差二乗和の利用（8）

- **一致性（2）**
  - **大数の法則より** $\sum_{i=1}^{n}\varepsilon_i^2\Big/n\longrightarrow E\left[\varepsilon_i^2\right]=V\left[\varepsilon_i\right]=\sigma^2$
  - **よって，** $\sum_{i=1}^{n}\hat{u}_i^2\Big/n\longrightarrow\sigma^2$
  - $\sum_{i=1}^{n}\hat{u}_i^2\Big/(n-K)=\lim\left(\sum_{i=1}^{n}\hat{u}_i^2\Big/n\right)\lim\{n/(n-K)\}=\sigma^2$
  - **問題点**
    - **一致推定量は残差二乗和をnで割ってもn－Kで割っても得られる．**
    - **従って，決め手は不偏性となる．**

6．自由度（1）

- **問題**
  - **説明変数がK個ある回帰の場合は** $E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]=(n-K)\sigma^2$
  - **つまり，**$E[RSS]=(n-K)\sigma^2$ **であるから，n-K個の独立な平均0，分散**σ^2 **な確率変数の2乗の和であると考えられる．**
  - **このようなことはESSやTSSでもいえるのだろうか？**
  - **もともと，Yのn個の値はn個の誤差項＋回帰直線でできているが，RSSでK個だけ少ない確率変数の和になっている．これはどういうことか？**

6．自由度（2）

- **自由度**
  - **ある変数の期待値が誤差項の分散のp倍＋定数で表されたとき，その変数は自由度pであるとよぶ．**
  - $E[M]=p\sigma^2+$定数
  - **この変数は定数と平均0で分散**σ^2 **の互いに独立な確率変数の2乗がp個が合計されたものと考えることが出来る．（確率変数の2乗の期待値は**σ^2 **だから．）**
  - **RSSの場合は定数0とn-K個の確率変数の2乗の合計＝RSSの自由度はn-K**
  - **誤差項の分散の不偏推定量はRSSをその自由度で割る**

6．自由度（3）

- **TSSの場合（1）**
  - **回帰式** $y_i=\alpha+\beta x_i+\varepsilon_i$ **を例に考える**
    - $\bar{y}=\alpha+\beta\bar{x}+\bar{\varepsilon}$
    - $y_i-\bar{y}=\beta\left(x_i-\bar{x}\right)+\varepsilon_i-\bar{\varepsilon}$
  - $\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\beta^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+\sum_{i=1}^{n}\left(\varepsilon_i-\bar{\varepsilon}\right)^2$

    $+2\beta\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(\varepsilon_i-\bar{\varepsilon}\right)$

6．自由度（4）

- **TSSの場合（2）**
  - $\sum_{i=1}^{n}\left(\varepsilon_i-\bar{\varepsilon}\right)^2=\sum_{i=1}^{n}\varepsilon_i^2-2\bar{\varepsilon}\sum_{i=1}^{n}\varepsilon_i+\sum_{i=1}^{n}\bar{\varepsilon}^2$

    $=\sum_{i=1}^{n}\varepsilon_i^2-2\bar{\varepsilon}\left(n\bar{\varepsilon}\right)+n\bar{\varepsilon}^2=\sum_{i=1}^{n}\varepsilon_i^2-n\bar{\varepsilon}^2$
  - $E\left[\sum_{i=1}^{n}\left(\varepsilon_i-\bar{\varepsilon}\right)^2\right]=E\left[\sum_{i=1}^{n}\varepsilon_i^2\right]-nE\left[\bar{\varepsilon}^2\right]=\sum_{i=1}^{n}E\left[\varepsilon_i^2\right]-nV\left[\bar{\varepsilon}\right]$

    $=n\sigma^2-n\frac{\sigma^2}{n}=(n-1)\sigma^2$

6．自由度（5）

- TSSの場合（3）　TSSの自由度はn-1
  - $E\left[\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})\right]=\sum_{i=1}^{n}(x_i-\bar{x})E[\varepsilon_i-\bar{\varepsilon}]$
    $=\sum_{i=1}^{n}(x_i-\bar{x})\{E[\varepsilon_i]-E[\bar{\varepsilon}]\}=0$
  - $E[TSS]=E\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right]$
    $=\beta^2\sum_{i=1}^{n}(x_i-\bar{x})^2+E\left[\sum_{i=1}^{n}(\varepsilon_i-\bar{\varepsilon})^2\right]+2\beta E\left[\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})\right]$
    $=\beta^2\sum_{i=1}^{n}(x_i-\bar{x})^2+(n-1)\sigma^2$

6．自由度（6）

- ESSの場合（1）

$$ESS=\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$
$$\hat{\beta}=\beta+\left(\hat{\beta}-\beta\right)$$
$$\hat{\beta}^2=\beta^2+\left(\hat{\beta}-\beta\right)^2+2\beta\left(\hat{\beta}-\beta\right)$$
$$E\left[\hat{\beta}^2\right]=\beta^2+E\left[\left(\hat{\beta}-\beta\right)^2\right]+2\beta E\left[\hat{\beta}-\beta\right]$$
$$E\left[\hat{\beta}-\beta\right]=0$$

6．自由度（7）

- ESSの場合（2）

$$E\left[\hat{\beta}^2\right]=\beta^2+E\left[\left(\hat{\beta}-\beta\right)^2\right]=\beta^2+V\left[\hat{\beta}\right]=\beta^2+\sigma^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2$$
$$E[ESS]=E\left[\hat{\beta}^2\right]\sum_{i=1}^{n}(x_i-\bar{x})^2=\left(\beta^2+\sigma^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\sum_{i=1}^{n}(x_i-\bar{x})^2$$
$$=\sigma^2+\beta^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

- 一般の場合　ESSの自由度はK-1

$$E[ESS]=(K-1)\sigma^2+定数$$

6．自由度（8）

- TSS=ESS+RSSの帰結
  - $E[TSS]=E[ESS]+E[RSS]$
  - これを σ^2 の係数について比較すると
  - TSSの自由度＝ESSの自由度＋RSSの自由度
  - 実際，
    - TSSの自由度＝n-1
    - ESSの自由度＝K-1
    - RSSの自由度＝n-K

7．標準化統計量（1）

- 分布，平均，分散が変わらない分布を求める．
- 誤差項が正規分布の場合
  - $\hat{\beta}_j$ は$N\left(\beta_j,\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2\right)$に従う．（平均βj,分散 $\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2$ の正規分布）
  - なぜならば，$\hat{\beta}_j=\beta_j+\sum_{i=1}^{n}\left(x_{j,i}^*\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2\right)\varepsilon_i$ であるが

    独立な正規分布に従う確率変数の線形和は，正規分布だから．

7．標準化統計量（2）

- 標準正規分布になる統計量（1）
  - $E\left[\hat{\beta}_j\right]=\beta_j$ より，$E\left[\hat{\beta}_j-\beta_j\right]=E\left[\hat{\beta}_j\right]-E\left[\beta_j\right]=E\left[\hat{\beta}_j\right]-\beta_j=0$
  - $V\left[\hat{\beta}_j-\beta_j\right]=V\left[\hat{\beta}_j\right]+V\left[\beta_j\right]=V\left[\hat{\beta}_j\right]+0=V\left[\hat{\beta}_j\right]$
  - 従って，$\hat{\beta}_j-\beta_j$ の分布は $N\left(0,\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2\right)$
  - $V\left[\dfrac{\hat{\beta}_j-\beta_j}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2}}\right]=\dfrac{V\left[\hat{\beta}_j-\beta_j\right]}{\left\{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2}\right\}^2}=\dfrac{\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2}{\sigma^2\Big/\sum_{i=1}^{n}\left(x_{j,i}^*\right)^2}=1$

7．標準化統計量（3）

- 標準正規分布になる統計量（2）
  - $E\left[\dfrac{\hat{\beta}_j-\beta_j}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}}\right]=\dfrac{E\left[\hat{\beta}_j-\beta_j\right]}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}}=\dfrac{0}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}}=0$
  - 従って，$\dfrac{\hat{\beta}_j-\beta_j}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}}$ の分布は $N(0,1)$
  - σが解っている場合には有効

7．標準化統計量（4）

- 誤差項が正規分布ではない場合
  - 中心極限定理を使うとnが大きい（nが30以上）場合 $\dfrac{\hat{\beta}_j-\beta_j}{\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}}$ の分布は標準正規分布N（0，1）で近似可能
  - この場合，誤差分散の推定量の一致性を利用できる

    ので，σを $\hat{\sigma}=\sqrt{\sum_{i=1}^{n}\hat{u}_i^2\Big/n}$ で置き換えて計算してもよい．
    - この場合は，σは未知でも良い．

7．1　t統計量

- σが不明の場合＝実証分析では一般的
- σの推定が必要＝ σ^2 の不偏推定量を用いる．

  $S^2=\sum_{i=1}^{n}\hat{u}_i^2\Big/(n-K) \quad S=\sqrt{S^2}=\sqrt{\sum_{i=1}^{n}\hat{u}_i^2\Big/(n-K)}$
- 標準化統計量のσをSで置き換えたものをt統計量と呼ぶ．

  $t=\left(\hat{\beta}_j-\beta_j\right)\Big/\left(S\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}\right)$
  - σをSで置き換えたことで分布は標準正規分布ではなくなる．

7．1．1　t分布

- t統計量の分布
  - 誤差項が正規分布の時は，t分布になる．
  - t分布
    - Z_j が各々独立で，平均0，分散1の正規分布に従うとする．
    - $Z_{j+1}\Big/\sqrt{\sum_{j=1}^{p}Z_j^2\Big/p}$ を自由度pのt分布と呼ぶ．

$$E\left[\sum_{j=1}^{p}Z_j^2\right]=\sum_{i=1}^{p}E\left[Z_j^2\right]=p$$

$$t=\left\{\left(\hat{\beta}_j-\beta_j\right)\Big/\left(\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}\right)\right\}\Big/\sqrt{\sum_{i=1}^{n}\left(\hat{u}_i/\sigma\right)^2\Big/(n-K)}$$

7．1．2　t分布になる理由（1）

- 理由（1）
  - $t=\left\{\left(\hat{\beta}_j-\beta_j\right)\Big/\left(\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}\right)\right\}\Big/\sqrt{\sum_{i=1}^{n}\left(\hat{u}_i/\sigma\right)^2\Big/(n-K)}$
  - $\left(\hat{\beta}_j-\beta_j\right)\Big/\left(\sigma\Big/\sqrt{\sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2}\right)=Z_{n-K+1}\sim N(0,1)$
  - $E\left[\sum_{i=1}^{n}\left(\hat{u}_i/\sigma\right)^2\right]=E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]/\sigma^2=n-K=E\left[\sum_{j=1}^{n-K}Z_j^2\right]$ より，

  $\sum_{i=1}^{n}\left(\hat{u}_i/\sigma\right)^2=\sum_{j=1}^{n-K}Z_j^2=\sum_{j=1}^{n-K}N(0,1)^2$ と書けるとみられる．

7．1．2　t分布になる理由（2）

- 理由（2）
  - 分母と分子の独立性
    - Zn-k+1と $\hat{u}_i$ の独立性がいえれば，Zn-k+1と $\sum_{i=1}^{n}\hat{u}_i^2$ の独立性もいえる．
    - Zn-k+1と $\hat{u}_i$ はともに正規分布なのでその共分散が0であれば独立
    - $Cov\left[\hat{\beta}_j-\beta_j,\hat{u}_i\right]=0$ をいえばよい．
    - 残差と係数推定値の独立性をいえば，分子と分母の独立性が言える．

7．1．2　t分布になる理由（3）

- 理由（3）
  - 残差と係数推定値の独立性（1）
    - 例として $y_i = \alpha + \beta x_i + \varepsilon_i$ を考える．
    - 残差を考えるには例の変換をした方がいい．

      $y_i = (\alpha + \beta \bar{x}) + \beta (x_i - \bar{x}) + \varepsilon_i$
    - $\gamma = \alpha + \beta \bar{x}$ とおくと $\hat{\gamma} = \hat{\alpha} + \hat{\beta}\bar{x} = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta}\bar{x} = \bar{y}$
    - $y_i = \gamma + \beta (x_i - \bar{x}) + \varepsilon_i$
    - $\hat{u}_i = -(\hat{\gamma} - \gamma) - (\hat{\beta} - \beta)(x_i - \bar{x}) + \varepsilon_i$
    - $\hat{\gamma} - \gamma = \bar{y} - (\alpha + \beta \bar{x}) = (\alpha + \beta \bar{y} + \bar{\varepsilon}) - (\alpha + \beta \bar{x}) = \bar{\varepsilon}$

7．1．2　t分布になる理由（4）

- 理由（4）
  - 残差と係数推定値の独立性（2）
    - $Cov[\hat{u}_i, \hat{\gamma} - \gamma] = -Cov[\hat{\gamma} - \gamma, \hat{\gamma} - \gamma]$
      $- (x_i - \bar{x}) Cov[\hat{\beta} - \beta, \hat{\gamma} - \gamma] + Cov[\varepsilon_i, \hat{\gamma} - \gamma]$
    - $\sum_{i=1}^{n} (1 \cdot (x_i - \bar{x})) = 0$ なので説明変数同士の積和が0である．この場合，その係数推定値同士は無相関なので，

      $Cov[\hat{\beta} - \beta, \hat{\gamma} - \gamma] = 0$
    - $Cov[(\hat{\gamma} - \gamma), (\hat{\gamma} - \gamma)] = V[\hat{\gamma}] = \sigma^2 / n$

7．1．2　t分布になる理由（5）

- 理由（5）
  - 残差と係数推定値の独立性（3）
    - $Cov[\varepsilon_i, \hat{\gamma} - \gamma] = Cov\left[\varepsilon_i, \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j\right] = \frac{1}{n}\sum_{i=1}^{n} Cov[\varepsilon_i, \varepsilon_j]$

      $= \frac{1}{n} Cov[\varepsilon_i, \varepsilon_i] = \sigma^2 / n$
    - $Cov[\hat{u}_i, \hat{\gamma} - \gamma] = -Cov[\hat{\gamma} - \gamma, \hat{\gamma} - \gamma] - (x_i - \bar{x}) Cov[\hat{\beta} - \beta, \hat{\gamma} - \gamma]$
      $+ Cov[\varepsilon_i, \hat{\gamma} - \gamma] = 0$

7．1．2　t分布になる理由（6）

- 理由（6）
  - 残差と係数推定値の独立性（3）
    - $Cov[\hat{u}_i, (\hat{\beta} - \beta)] = -Cov[(\hat{\gamma} - \gamma), (\hat{\beta} - \beta)]$
      $- (x_i - \bar{x}) Cov[\hat{\beta} - \beta, \hat{\beta} - \beta] + Cov[\varepsilon_i, \hat{\beta} - \beta]$
      $= -(x_i - \bar{x}) Cov[\hat{\beta} - \beta, \hat{\beta} - \beta] + Cov[\varepsilon_i, \hat{\beta} - \beta]$
    - $Cov[\hat{\beta} - \beta, \hat{\beta} - \beta] = V[\hat{\beta}] = \sigma^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$

7．1．2　t分布になる理由（7）

- 理由（7）
  - 残差と係数推定値の独立性（4）
    - $Cov[\varepsilon_i, \hat{\beta} - \beta] = Cov\left[\varepsilon_i, \sum_{j=1}^{n} (x_j - \bar{x}) \varepsilon_j \Big/ \sum_{j=1}^{n} (x_j - \bar{x})^2\right]$

      $= \sum_{i=1}^{n} (x_j - \bar{x}) Cov[\varepsilon_i, \varepsilon_j] \Big/ \sum_{j=1}^{n} (x_j - \bar{x})^2$

      $= (x_j - \bar{x}) Cov[\varepsilon_i, \varepsilon_i] \Big/ \sum_{j=1}^{n} (x_j - \bar{x})^2$

      $= (x_j - \bar{x}) \sigma^2 \Big/ \sum_{j=1}^{n} (x_j - \bar{x})^2$
    - $Cov[\hat{u}_i, (\hat{\beta} - \beta)] = -(x_i - \bar{x}) \sigma^2 \Big/ \sum_{j=1}^{n} (x_i - \bar{x})^2 + (x_i - \bar{x}) \sigma^2 \Big/ \sum_{j=1}^{n} (x_i - \bar{x})^2 = 0$

7．1．2　t分布になる理由（7）

- まとめ
  - $t = \left\{ (\hat{\beta}_j - \beta_j) \Big/ \left(\sigma \Big/ \sqrt{\sum_{i=1}^{n} (x_{j,i}^{*})^2} \right) \right\} \Big/ \sqrt{\sum_{i=1}^{n} (\hat{u}_i / \sigma)^2 \Big/ (n - K)}$
  - $\sum_{i=1}^{n-K} (\hat{u}_i / \sigma)^2 = \sum_{i=1}^{n-K} Z_i^2 \quad Z_i \sim N(0,1)$, 各独立
  - $(\hat{\beta}_j - \beta_j) \Big/ \left(\sigma \Big/ \sqrt{\sum_{i=1}^{n} (x_{j,i}^{*})^2} \right) \equiv Z_{t-K+1} \sim N(0,1)$
  - 分母と分子が独立（残差と係数推定値の独立性から）
  - $t = Z_{n-K+1} \Big/ \sqrt{\sum_{i=1}^{n-K} Z_i^2 \Big/ (n - K)} \quad Z_i \sim N(0,1)$, 各独立