統計解析論特殊講義３

多変数回帰

1．説明変数を増やす必要性（1）

- 例：コブダグラス型生産関数

  $Y(\text{生産量}) = AK^{\alpha}L^{\beta}$

  ただし，Yは生産量，Kは資本設備量，
  Lは労働投入量，Aは生産性を表す定数

  データでこの関係を表すと

  $y_i = a_i k_i^{\alpha} l_i^{\beta}$

  両辺の自然対数をとると

  $\log y_i = \log a_i + \alpha \log k_i + \beta \log l_i$

  さらに $\log a_i = \gamma + \text{誤差}_i$ とすると

  $\log y_i = \gamma + \alpha \log k_i + \beta \log l_i + \text{誤差}_i$

1．説明変数を増やす必要性（2）

- GW中の旅客数
  - 今年のGWは飛び石でしょぼかった．
    - GW中のＪＲ旅客数は当然減る（前年同期比で２％減）
  - ２％減の原因
    - ＳＡＲＳ，経済不況の影響を受けているのか？
    - 連休の構成，東北の桜花見シーズンが例年より速いことだけで説明できるのか？
  - 以下のようなモデルをたてる（The First Insightのために）

    $$\text{旅客数} = \delta + \alpha(\text{連休中の最大連続休日数}) + \beta(GDP) + \gamma(\text{連休中桜開花中の東北県数}) + \text{誤差}$$

    - βが正であれば経済不況の効果がある．どの程度の貢献かもわかる．
    - さらに，これで予測した回帰値よりどの位今年の旅客数が減ったかを調べれば今年特有の効果がどう効いたかをみることができる．

2．3説明変数の回帰式

- 回帰式

  $y_i = \alpha + \beta x_i + \gamma z_i + \text{誤差}_i$

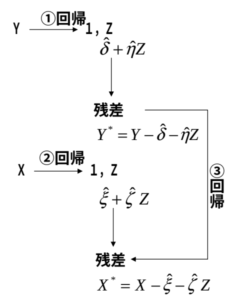
- 問題
  - α，β，γの最小二乗推定量をデータの値の式で表す．
  - 残差と説明変数の関係は？
  - 全変動の分解は２説明変数の場合と同様か？
  - Ｘのみの回帰の場合とＺを加えた回帰の関係
    - βの最小二乗推定量の変化は何によるのか？
    - 残差変動はどう変わるか？

3．単回帰の反省

- 二つの事実
  - 回帰式 $y_i = \alpha + \beta x_i + \text{誤差}_i$ のβの最小二乗推定量と残差は回帰式 $y_i - \bar{y} = \beta(x_i - \bar{x}) + \text{誤差}_i$ のβの最小二乗推定量と残差と同じである．
  - 回帰式 $x_i = \gamma + \text{誤差}_i$ のγの最小二乗推定量は $\bar{x}$，よって，$x_i - \bar{x}$ は回帰式 $x_i = \gamma + \text{誤差}_i$ の最小二乗残差．同様に，$y_i - \bar{y}$ は $y_i = \delta + \text{誤差}_i$ の最小二乗残差
- 含意
  - βの推定値や残差は，被説明変数Ｙと問題の説明変数Ｘをもう一つの説明変数１に回帰した残差同士の回帰の推定値や残差と同じ．
  - αに関しても同様のことがいえる．

4．3説明変数への一般化

- 予想
  - ２変数の場合の拡張
  - βの最小二乗推定量や残差は，被説明変数Ｙと問題の説明変数Ｘを残りの説明変数１とＺとに回帰した残差同士の回帰の推定値や残差と同じ．
  - ２変数の回帰の場合はＺがないので，残りの説明変数は１のみになるので，前のスライドのようになる．



4．1　予想の証明(1)

- **被説明変数のX以外の説明変数への回帰と残差**

  $y_i \xrightarrow{\text{回帰}} 1, z_i$ **を考える**

  - **回帰式** $y_i = \delta + \eta\, z_i + \text{誤差}_i$
  - $\hat{\eta} = \dfrac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}, \quad \hat{\delta} = \bar{y} - \hat{\eta}\,\bar{z}$
  - **回帰値を** $proj(y_i | 1, Z) \equiv \hat{\delta} + \hat{\eta}\, z_i$ **とかくことにする**
    - **単回帰の** $\bar{y}$ **に対応**
  - **残差を**$y_i^* \equiv y_i - proj(y_i | 1, Z) = y_i - \hat{\delta} - \hat{\eta}\, z_i$
    - **単回帰の** $y_i - \bar{y}$ **に対応**

4．1　予想の証明(2)

- **説明変数のX以外の説明変数への回帰と残差**
  - **回帰式**$x_i = \xi + \zeta\, z_i + \text{誤差}_i$
  - $\hat{\zeta} = \dfrac{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}, \quad \hat{\xi} = \bar{x} - \hat{\zeta}\,\bar{z}$
  - **回帰値** $proj(x_i | 1, Z) \equiv \hat{\xi} + \hat{\zeta}\, z_i$　($\bar{x}$ **に対応**)
  - **残差** $x_i^* \equiv x_i - proj(x_i | 1, Z) = x_i - \hat{\xi} - \hat{\zeta}\, z_i$　($x_i - \bar{x}$**に対応**)

4．1　予想の証明(3)

- **残差の性質**　($\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ **,** $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ **に対応**)
  - $\sum_{i=1}^{n} y_i^* = 0,\ \sum_{i=1}^{n} y_i^* z_i = 0,\ \sum_{i=1}^{n} y_i^*\, proj(y_i | 1, Z) = 0,\ \sum_{i=1}^{n} y_i^*\, proj(x_i | 1, Z) = 0$
    - **理由**
      - **残差と説明変数の積の合計は0になる**
      - $proj(y_i | 1, Z) \equiv \hat{\delta} + \hat{\eta} z_i$ **より**

        $\sum_{i=1}^{n} y_i^*\, proj(y_i | 1, Z) = \sum_{i=1}^{n} y_i^* (\hat{\delta} + \hat{\eta} z_i) = \hat{\delta} \sum_{i=1}^{n} y_i^* + \hat{\eta} \sum_{i=1}^{n} y_i^* z_i = 0$
  - $\sum_{i=1}^{n} x_i^* = 0,\ \sum_{i=1}^{n} x_i^* z_i = 0,\ \sum_{i=1}^{n} x_i^*\, proj(y_i | 1, Z) = 0,\ \sum_{i=1}^{n} x_i^*\, proj(x_i | 1, Z) = 0$

4．1　予想の証明(4)

- **誤差の分解**
  - $$\begin{aligned}\text{誤差}_i &= y_i - (\alpha + \beta\, x_i + \gamma\, z_i) \\ &= y_i^* + proj(y_i | 1, Z) - \alpha - \beta(x_i^* + proj(x_i | 1, Z)) - \gamma\, z_i \\ &= y_i^* - \beta\, x_i^* + (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i)\end{aligned}$$
    - $\text{誤差}_i = (y_i - \bar{y}) - \beta(x_i - \bar{x}) + (\bar{y} - \alpha - \beta\bar{x})$ **に対応**

4．1　予想の証明(5)

- **誤差の二乗和の計算(1)**

$$\begin{aligned}\sum_{i=1}^{n} \text{誤差}_i^2 &= \sum_{i=1}^{n} \{y_i^* - \beta\, x_i^* + (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i)\}^2 \\ &= \sum_{i=1}^{n} (y_i^*)^2 - 2\beta \sum_{i=1}^{n} y_i^* x_i^* + \beta^2 \sum_{i=1}^{n} (x_i^*)^2 \\ &\quad + \sum_{i=1}^{n} (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i)^2 \\ &\quad + 2 \sum_{i=1}^{n} y_i^* (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i) \\ &\quad - 2\beta \sum_{i=1}^{n} x_i^* (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i)\end{aligned}$$

4．1　予想の証明(6)

- **誤差の二乗和の計算(2)**
  - **0になる項**
    - $\sum_{i=1}^{n} y_i^* (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i)$

      $= \sum_{i=1}^{n} y_i^*\, proj(y_i | 1, Z) - \alpha \sum_{i=1}^{n} y_i^* - \beta \sum_{i=1}^{n} y_i^*\, proj(x_i | 1, Z) - \gamma \sum_{i=1}^{n} y_i^* z_i = 0$
    - $\sum_{i=1}^{n} x_i^* (proj(y_i | 1, Z) - \alpha - \beta\, proj(x_i | 1, Z) - \gamma\, z_i) = 0$

4．1　予想の証明（7）

- **平方完成**

$$\sum_{i=1}^{n} 誤差_i^2 = \sum_{i=1}^{n}\left(y_i^*\right)^2 - 2\beta\sum_{i=1}^{n} y_i^* x_i^* + \beta^2\sum_{i=1}^{n}\left(x_i^*\right)^2$$
$$+\sum_{i=1}^{n}\left(proj\left(y_i \middle| 1, Z\right) - \alpha - \beta\ proj\left(x_i \middle| 1, Z\right) - \gamma z_i\right)^2$$
$$= \sum_{i=1}^{n}\left(x_i^*\right)^2 \left\{\beta - \frac{\sum_{i=1}^{n} y_i^* x_i^*}{\sum_{i=1}^{n}\left(x_i^*\right)^2}\right\}^2 + \left\{\sum_{i=1}^{n}\left(y_i^*\right)^2 - \frac{\left(\sum_{i=1}^{n} y_i^* x_i^*\right)^2}{\sum_{i=1}^{n}\left(x_i^*\right)^2}\right\}$$
$$+\sum_{i=1}^{n}\left(proj\left(y_i \middle| 1, Z\right) - \alpha - \beta\ proj\left(x_i \middle| 1, Z\right) - \gamma z_i\right)^2$$

4．1　予想の証明（8）

- **βの決定**　（ $\hat{\beta} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2$ **に対応）**
  - **まず，βを決め，それに基づいてα，γをきめればよいので，**
    $$\hat{\beta} = \sum_{i=1}^{n} y_i^* x_i^* \Big/ \sum_{i=1}^{n}\left(x_i^*\right)^2$$
- **αとγの決定**　（ $n\left(\bar{y} - \alpha - \beta\,\bar{x}\right)^2 = 0$ **に対応）**
  - $\sum_{i=1}^{n}\left(proj\left(y_i \middle| 1, Z\right) - \alpha - \beta\ proj\left(x_i \middle| 1, Z\right) - \gamma z_i\right)^2$ **を0にする．**
    $$proj\left(y_i \middle| 1, Z\right) - \alpha - \beta\ proj\left(x_i \middle| 1, Z\right) - \gamma z_i = \hat{\delta} + \hat{\eta} z_i - \alpha - \hat{\beta}\left(\hat{\xi} + \hat{\zeta}\, z_i\right) - \gamma z_i$$
    $$= \left(\hat{\delta} - \alpha - \hat{\beta}\hat{\xi}\right) + \left(\hat{\eta} - \hat{\beta}\hat{\zeta} - \gamma\right) z_i$$
    より $\hat{\alpha} = \hat{\delta} - \hat{\beta}\hat{\xi}$ **と** $\hat{\gamma} = \hat{\eta} - \hat{\beta}\hat{\zeta}$ **として0にする．**

4．1　予想の証明（9）

- $y_i = \alpha + \beta x_i + \gamma z_i + 誤差_i$ **の変形**
  - $y_i = y_i^* + \hat{\delta} + \hat{\eta} z_i$ **と** $x_i = x_i^* + \hat{\xi} + \hat{\zeta}\, z_i$ **から**
    $$y_i^* + \hat{\delta} + \hat{\eta} z_i = \alpha + \beta\left(x_i^* + \hat{\xi} + \hat{\zeta}\, z_i\right) + \gamma z_i + 誤差_i$$
  - **移項**
    $$y_i^* = \left(\alpha - \hat{\delta} + \beta\hat{\xi}\right) + \beta x_i^* + \left(\gamma - \hat{\eta} + \beta\hat{\zeta}\right) z_i + 誤差_i$$
  - $\hat{\alpha} = \hat{\delta} - \hat{\beta}\hat{\xi}$ **と**　$\hat{\gamma} = \hat{\eta} - \hat{\beta}\hat{\zeta}$ **とより**
    $$y_i^* = \beta x_i^* + 誤差_i$$
      - **βの推定と回帰の残差についてはこれで十分**
    - $y_i - \bar{y} = \beta\left(x_i - \bar{x}\right) + 誤差_i$　**に対応**

4．2　含意

- **βはXとYから1，Zの影響を取り除いた回帰の係数**
- **説明変数を増やしても推定値が変わらないケースがある．**
  - **回帰式** $y_i^* = y_i - \hat{\delta} - \hat{\eta}\, z_i = \beta\ x_i^* + 誤差_i$ **の変形**
    - $y_i = \beta\ x_i^* + \hat{\delta} + \hat{\eta}\, z_i + 誤差_i$
    - **これをδ，ηも推定する形にすれば回帰式は**
      $$y_i = \beta\ x_i^* + \delta + \eta\, z_i + 誤差_i$$
  - **δ，ηの由来**
    $$y_i = \delta + \eta\, z_i + 誤差_i$$
  - **回帰式への** x_i^* **の追加はδ，ηの値を変えない．**

4．3　推定量の展開（1）

- **問題**
  - **他の変数との関係で推定量はどう変わるのか？**
    - **説明変数Zを追加しても係数の推定値に影響を与えない場合がある．どういう場合か？**
    - **Xの回帰係数推定量が0なのに，説明変数Z追加で0ではなくなるような場合はあるのか？**
    - **上記の逆．Xの回帰係数推定量が0ではないのに，説明変数Z追加で0になるような場合はあるのか？**
  - **分散，共分散だけで最小二乗推定量を表してみたい．**

4．3　推定量の展開（2）

- **出発点**
  $$\hat{\beta} = \sum_{i=1}^{n} y_i^* x_i^* \Big/ \sum_{i=1}^{n}\left(x_i^*\right)^2$$
- **分散･共分散の記号**
  - **Xの分散**　$S_{XX} \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})$
  - **XとYの共分散**　$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
  - **Xの分散はXとXとの共分散と見なせる→** S_{XX}

4．4　推定量の展開（3）

- $\sum_{i=1}^{n}\left(x_i^*\right)^2$ の展開（1）
  - $\sum_{i=1}^{n}\left(x_i^*\right)^2=\sum_{i=1}^{n}x_i^*x_i^*=\sum_{i=1}^{n}\left(x_i-\hat{\xi}-\hat{\zeta}z_i\right)x_i^*=\sum_{i=1}^{n}x_ix_i^*-\sum_{i=1}^{n}\hat{\xi}x_i^*-\sum_{i=1}^{n}\hat{\zeta}z_ix_i^*$
    $=\sum_{i=1}^{n}x_ix_i^*-\hat{\xi}\sum_{i=1}^{n}x_i^*-\hat{\zeta}\sum_{i=1}^{n}z_ix_i^*$
  - x_i^* は回帰式 $x_i=\xi+\zeta z_i+誤差_i$ の残差だから，
    $\sum_{i=1}^{n}x_i^*=\sum_{i=1}^{n}x_i^*z_i=0$ である．よって，

$$\sum_{i=1}^{n}\left(x_i^*\right)^2=\sum_{i=1}^{n}x_ix_i^*-\hat{\xi}\sum_{i=1}^{n}x_i^*-\hat{\zeta}\sum_{i=1}^{n}z_ix_i^*=\sum_{i=1}^{n}x_ix_i^*$$

4．4　推定量の展開（4）

- $\sum_{i=1}^{n}\left(x_i^*\right)^2$ の展開（2）

$$\begin{aligned}\sum_{i=1}^{n}\left(x_i^*\right)^2&=\sum_{i=1}^{n}x_ix_i^*=\sum_{i=1}^{n}\{(x_i-\bar{x})+\bar{x}\}x_i^*=\sum_{i=1}^{n}(x_i-\bar{x})x_i^*+\bar{x}\sum_{i=1}^{n}x_i^*\\&=\sum_{i=1}^{n}(x_i-\bar{x})x_i^*=\sum_{i=1}^{n}(x_i-\bar{x})\left\{(x_i-\bar{x})+\bar{x}-\hat{\xi}-\hat{\zeta}(z_i-\bar{z})-\hat{\zeta}\bar{z}\right\}\\&=\sum_{i=1}^{n}(x_i-\bar{x})^2-\hat{\zeta}\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})+\left(\bar{x}-\hat{\xi}-\hat{\zeta}\bar{z}\right)\sum_{i=1}^{n}(x_i-\bar{x})\\&=\sum_{i=1}^{n}(x_i-\bar{x})^2-\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})\left\{\sum_{i=1}^{n}(z_i-\bar{z})^2\right\}^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})\\&=n\left\{S_{XX}-S_{XZ}S_{ZZ}^{-1}S_{ZX}\right\}\end{aligned}$$

4．3　推定量の展開（5）

- $\sum_{i=1}^{n}y_i^*x_i^*$ の展開

$$\begin{aligned}\sum_{i=1}^{n}y_i^*x_i^*&=\sum_{i=1}^{n}\left(y_i-\hat{\delta}-\hat{\eta}z_i\right)x_i^*=\sum_{i=1}^{n}\left\{(y_i-\bar{y})+\bar{y}-\hat{\delta}-\hat{\eta}z_i\right\}x_i^*\\&=\sum_{i=1}^{n}(y_i-\bar{y})x_i^*+\sum_{i=1}^{n}\left(\bar{y}-\hat{\delta}\right)x_i^*-\sum_{i=1}^{n}\hat{\eta}z_ix_i^*\\&=\sum_{i=1}^{n}(y_i-\bar{y})x_i^*+\left(\bar{y}-\hat{\delta}\right)\sum_{i=1}^{n}x_i^*-\hat{\eta}\sum_{i=1}^{n}z_ix_i^*=\sum_{i=1}^{n}(y_i-\bar{y})x_i^*\\&=\sum_{i=1}^{n}(y_i-\bar{y})\left(x_i-\hat{\xi}-\hat{\zeta}z_i\right)=\sum_{i=1}^{n}(y_i-\bar{y})\left\{(x_i-\bar{x})+\bar{x}-\hat{\xi}-\left(\hat{\zeta}z_i-\hat{\zeta}\bar{z}\right)-\hat{\zeta}\bar{z}\right\}\\&=\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\hat{\zeta}\sum_{i=1}^{n}(y_i-\bar{y})(z_i-\bar{z})+\left(\bar{x}-\hat{\xi}-\hat{\zeta}\bar{z}\right)\sum_{i=1}^{n}(y_i-\bar{y})\\&=\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})\Big/\sum_{i=1}^{n}(z_i-\bar{z})^2\\&=n\left\{S_{XY}-S_{XZ}S_{ZZ}^{-1}S_{ZY}\right\}\end{aligned}$$

4．3　推定量の展開（6）

- 最終結果

$$\hat{\beta}=\frac{\sum_{i=1}^{n}y_i^*x_i^*}{\sum_{i=1}^{n}\left(x_i^*\right)^2}=\frac{S_{xy}-S_{xz}S_{zz}^{-1}S_{zy}}{S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}}$$

- 回帰係数の最小二乗推定量は無相関の別の説明変数から影響を受けない
  $S_{xz}=0$ のケースを考えると $\hat{\beta}=S_{XY}/S_{XX}$
  - 無相関変数は回帰に含まれても含まれなくても同じ
  - 誤差の新解釈

4．4　以上からの教訓（1）

- 無相関な説明変数同士はお互いの係数推定に影響を与えない
  - $y_i=\alpha+\beta x_i+\gamma z_i+誤差_i$ のβの最小二乗推定量は，$y_i=\alpha+\beta x_i+誤差_i$ のβの最小二乗推定量と変わらない．
  - γz_i は説明変数と無相関のため誤差に含めることが可能である．
  - 逆に誤差とは説明変数と相関がない要因をまとめているものともいえる．
    - 逆に誤差は全ての説明変数との相関が0とならなければならないともいえる．

4．4　以上からの教訓（2）

- 変数を回帰式に追加するときの教訓（1）
  - XとYが無相関でも $y_i=\alpha+\beta x_i+\gamma z_i+誤差_i$ のβの最小二乗推定量は0にならないかもしれない．
    （$S_{XY}=0$より）$\hat{\beta}=\dfrac{-S_{xz}S_{zz}^{-1}S_{zy}}{S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}}$
  - 例
    - $Z\cong c_1Y+c_2X$ のとき
      - 余計な変数を加えることの危険性を示している

4．4　以上からの教訓（3）

- 変数を回帰式に追加するときの教訓（2）
  - 逆に他の変数を加えたことで被説明変数への影響がなくなることがある．

    $S_{xy} - S_{xz}S_{zz}^{-1}S_{zy} = 0$のとき
  - 例
    - $Y \cong c_1 Z$，$X \cong c_2 Z$ とすると $Y \cong (c_1/c_2)X$ だから，YのZへの回帰係数は$S_{zz}^{-1}S_{zy} \cong c_1$となる．同様に，XのZへの回帰係数から $S_{zz}^{-1}S_{xz} \cong c_2$ ．さらに，YのXへの回帰から $S_{xx}^{-1}S_{xy} \cong c_1/c_2$．また，$X \cong c_2 Z$から，$S_{xx} \cong c_2^2 S_{zz}$．
    - $$S_{xy} - S_{xz}S_{zz}^{-1}S_{zy} = \left\{\left(S_{xy}S_{xx}^{-1}\right)\left(S_{xx}S_{zz}^{-1}\right) - \left(S_{xz}S_{zz}^{-1}\right)\left(S_{zy}S_{zz}^{-1}\right)\right\}S_{zz}$$
      $$\cong \left((c_1/c_2)c_2^2 - c_1c_2\right) = 0$$

4．5　説明力の指標（0）

- 問題
  - TSS＝ESS＋RSSの分解はこの場合もうまくいくか？
    - うまくいけば説明力の指標として決定係数＝ESS／TSSを考えることができる．
  - 新たな説明変数の追加によって決定係数はどう変わるのか？
    - 増加するのか，減少するのか，変わらないのか
    - 変化分はどのように決まるのか？

4．5　説明力の指標（1）

- 残差
  - $\hat{u}_i = y_i^* - \hat{\beta}x_i^*$
- 回帰値
  - $\hat{y}_i = y_i - \hat{u}_i = y_i - y_i^* + \hat{\beta}x_i^*$

    $= \hat{\delta} + \hat{\eta}z_i + \hat{\beta}x_i^* = proj(y_i|1,Z) + \hat{\beta}x_i^*$
- 残差二乗和（RSS）＜誤差二乗和の最小値だから＞

$$RSS = \sum_{i=1}^{n}\hat{u}_i^2 = \sum_{i=1}^{n}(y_i^*)^2 - \frac{\left(\sum_{i=1}^{n}y_i^*x_i^*\right)^2}{\sum_{i=1}^{n}(x_i^*)^2} = \sum_{i=1}^{n}(y_i^*)^2 - \sum_{i=1}^{n}(\hat{\beta}x_i^*)^2$$

4．5　説明力の指標（2）

- TSSの分解（1）＜RSSの計算＞

$$RSS = \sum_{i=1}^{n}(y_i^*)^2 - \left(\sum_{i=1}^{n}y_i^*x_i^*\right)^2 \Big/ \sum_{i=1}^{n}(x_i^*)^2$$
$$= \sum_{i=1}^{n}(y_i^*)^2 - \left(\sum_{i=1}^{n}y_i^*x_i^*\right)^2 \Big/ \left(\sum_{i=1}^{n}(x_i^*)^2\right)^2 \cdot \left(\sum_{i=1}^{n}(x_i^*)^2\right)$$
$$= \sum_{i=1}^{n}(y_i^*)^2 - \hat{\beta}^2\sum_{i=1}^{n}(x_i^*)^2 = \sum_{i=1}^{n}(y_i^*)^2 - \sum_{i=1}^{n}(\hat{\beta}x_i^*)^2$$
$$= \sum_{i=1}^{n}(y_i^* - y_i^*\text{の標本平均})^2 - \sum_{i=1}^{n}(\hat{\beta}x_i^* - \hat{\beta}x_i^*\text{の標本平均})^2$$
$$= y_i^*\text{の変動} - (\hat{\beta}x_i^*)\text{の変動}$$

（$y_i^*\text{の標本平均} = \sum_{i=1}^{n}y_i^*/n = 0$，$x_i^*\text{の標本平均} = \frac{1}{n}\sum_{i=1}^{n}x_i^* = 0$より）

4．5　説明力の指標（3）

- TSSの分解（2）＜y_i^*の変動 の計算＞
  - y_i^*の変動は回帰式 $y_i = \delta + \eta z_i + 誤差_i$ だから，この回帰式のTSS，ESS，RSSの関係より

    $$TSS = proj(y_i|1,Z)\text{の変動} + y_i^*\text{の変動}$$
  - $RSS = y_i^*\text{の変動} - (\hat{\beta}x_i^*)\text{の変動}$より，

    $$TSS = RSS + (\hat{\beta}x_i^*)\text{の変動} + proj(y_i|1,Z)\text{の変動}$$

4．5　説明力の指標（4）

- TSSの分解（3）＜ESSの計算（1）＞
  - 回帰値の計算

    $\hat{y}_i = y_i - \hat{u}_i = y_i - y_i^* + \hat{\beta}x_i^* = proj(y_i|1,Z) + \hat{\beta}x_i^*$
  - 回帰値の標本平均

    $\hat{y}_i\text{の標本平均} = proj(y_i|1,Z)\text{の標本平均} + \hat{\beta}x_i^*\text{の標本平均}$

    $\hat{y}_i - \hat{y}_i\text{の標本平均}$

    $= proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均} + \hat{\beta}x_i^* - \hat{\beta}x_i^*\text{の標本平均}$

4．5　説明力の指標 (5)

- TSSの分解 (4) ＜ESSの計算 (2) ＞
  - ESS

$$ESS = \hat{y}_i \text{の変動} = \sum_{i=1}^{n}\left(\hat{y}_i - \hat{y}_i \text{の標本平均}\right)^2$$

$$= \sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)^2 + \sum_{i=1}^{n}\left(\hat{\beta} x_i^* - \hat{\beta} x_i^* \text{の標本平均}\right)^2$$

$$+ 2\sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)\left(\hat{\beta} x_i^* - \hat{\beta} x_i^* \text{の標本平均}\right)$$

4．5　説明力の指標 (6)

- TSSの分解 (5) ＜ESSの計算 (3) ＞
  - クロス項が0になる

$$\sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)\left(\hat{\beta} x_i^* - \hat{\beta} x_i^* \text{の標本平均}\right)$$

$$= \sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)x_i^*$$

$$-\left(\hat{\beta} x_i^* \text{の標本平均}\right)\sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)$$

$$= \sum_{i=1}^{n} x_i^* \, proj(y_i|1,Z) - \left(proj(y_i|1,Z)\text{の標本平均}\right)\sum_{i=1}^{n} x_i^*$$

$$= 0$$

4．5　説明力の指標 (7)

- TSSの分解 (6) ＜ESSの計算 (4) ＞

ESS

$$= \sum_{i=1}^{n}\left(proj(y_i|1,Z) - proj(y_i|1,Z)\text{の標本平均}\right)^2 + \sum_{i=1}^{n}\left(\hat{\beta} x_i^* - \hat{\beta} x_i^* \text{の標本平均}\right)^2$$

$$= proj(y_i|1,Z)\text{の変動} + \left(\hat{\beta} x_i^*\right)\text{の変動}$$

- TSSの分解
  - $TSS = RSS + \left(\hat{\beta} x_i^*\right)\text{の変動} + proj(y_i|1,Z)\text{の変動}$ **とこの結果より** $TSS = RSS + ESS$
- 決定係数

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

4．6　追加説明変数と決定係数(1)

- 説明変数追加の効果 (1)
  - $ESS = proj(y_i|1,Z)\text{の変動} + \left(\hat{\beta} x_i^*\right)\text{の変動}$ **の意味**
    - $proj(y_i|1,Z)\text{の変動}$ **の意味**
      回帰式 $Y = \delta + \eta Z$ **のESS**
    - $\left(\hat{\beta} x_i^*\right)\text{の変動}$ **の意味**
      回帰式 $y_i = \beta x_i^* + \text{誤差}_i$ **のESS**
    - $\left(\text{回帰式}Y = \alpha + \beta X + \gamma Z\text{の}ESS\right)$

$$= \left(\text{回帰式}Y = \delta + \eta Z\text{の}ESS\right)$$

$$+ \left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}ESS\right)$$

4．6　追加説明変数と決定係数(2)

- 説明変数追加の効果 (2)

$$R^2\left(\text{回帰式}Y = \alpha + \beta X + \gamma Z\right)$$

$$= \frac{\left(\text{回帰式}Y = \alpha + \beta X + \gamma Z\text{の}ESS\right)}{TSS}$$

$$= \frac{\left(\text{回帰式}Y = \delta + \eta Z\text{の}ESS\right) + \left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}ESS\right)}{TSS}$$

$$= \left(\text{回帰式}Y = \delta + \eta Z\text{の}R^2\right) + \left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}R^2\right)$$

  - Xという変数を説明変数に追加したときの決定係数の増加は他の変数の影響を除いたX固有の変動がどの程度Yの変動を説明するかで決まる.

4．6　追加説明変数と決定係数(3)

- 注意
  - $\left(\text{回帰式}y_i = \beta x_i + \text{誤差}_i\text{の}R^2\right)$ **と**
    $\left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}R^2\right)$ **とは同じではない**
  - 決定係数の分解

$$R^2 = \left(\text{回帰式}Y = \phi + \psi X\text{の}R^2\right) + \left(\text{回帰式}Y = \gamma z_i^*\text{の}R^2\right)$$

$$R^2 = \left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}R^2\right)$$

$$+ \left(\text{回帰式}Y = \delta + \eta Z\text{の}R^2\right)$$

  - しかし

$$R^2 = \left(\text{回帰式}y_i = \beta x_i^* + \text{誤差}_i\text{の}R^2\right)$$

$$+ \left(\text{回帰式}y_i = \gamma z_i^* + \text{誤差}_i\text{の}R^2\right)$$

ではない.

4．7　偏相関係数

- Y^* と X^* との相関係数

$$r_{x,y|1,Z} = \frac{\sum_{i=1}^{n} x_i^* y_i^*}{\sqrt{\sum_{i=1}^{n} \left(x_i^*\right)^2}\sqrt{\sum_{i=1}^{n} \left(y_i^*\right)^2}}$$

- 偏相関係数の2乗は $y_i^* = \beta x_i^* + 誤差$ の決定係数
- XとYが，1とZの影響を取り除いた上で，どの程度相関があるかを示す．

4．7．1　偏相関係数と変数追加（1）

- 出発点

$$\left(回帰式 Y = \alpha + \beta X + \gamma Z のESS\right)$$
$$= \left(回帰式 Y = \delta + \eta Z のESS\right) + \left(回帰式 y_i = \beta x_i^* + 誤差_i のESS\right)$$

- RSSの変化

$$TSS - ESS\left(回帰式 Y = \alpha + \beta X + \gamma Z\right)$$
$$= TSS - ESS\left(回帰式 Y = \delta + \eta Z\right) - ESS\left(回帰式 y_i = \beta x_i^* + 誤差_i\right)$$
$$RSS\left(回帰式 Y = \alpha + \beta X + \gamma Z\right)$$
$$= RSS\left(回帰式 Y = \delta + \eta Z\right) - ESS\left(回帰式 y_i = \beta x_i^* + 誤差_i\right)$$

4．7．1　偏相関係数と変数追加（2）

- 変形

$$ESS\left(回帰式 y_i = \beta x_i^* + 誤差_i\right) = \sum_{i=1}^{n} \hat{\beta}^2 \left(x_i^*\right)^2$$
$$= ESS\left(回帰式 y_i^* = \beta x_i^* + 誤差_i\right)$$
$$R^2\left(回帰式 y_i^* = \beta x_i^* + 誤差_i\right)$$
$$= r_{XY|Z}^* = ESS\left(回帰式 y_i^* = \beta x_i^* + 誤差_i\right) \Big/ \sum_{i=1}^{n} \left(y_i^*\right)^2$$
$$\sum_{i=1}^{n} \left(y_i^*\right)^2 = RSS\left(回帰式 Y = \delta + \eta Z\right)$$
$$ESS\left(回帰式 y_i = \beta x_i^* + 誤差_i\right) = RSS\left(回帰式 Y = \delta + \eta Z\right) r_{XY|Z}^*$$

4．7．1　偏相関係数と変数追加（3）

- 変数追加によるRSSの減少と偏相関係数

$$RSS\left(回帰式 Y = \alpha + \beta X + \gamma Z\right)$$
$$= RSS\left(回帰式 Y = \delta + \eta Z\right) - ESS\left(回帰式 y_i = \beta x_i^* + 誤差_i\right)$$
$$= RSS\left(回帰式 Y = \delta + \eta Z\right) - RSS\left(回帰式 Y = \delta + \eta Z\right) r_{XY|Z}^*$$
$$= RSS\left(回帰式 Y = \delta + \eta Z\right)\left(1 - r_{XY|Z}^2\right)$$

4．8　残りの係数の決定

- 既に述べた方法

$$\hat{\alpha} = \hat{\delta} - \hat{\beta}\hat{\xi} \qquad \hat{\gamma} = \hat{\eta} - \hat{\beta}\hat{\zeta}$$

- 回帰式の変形

$$y_i = \alpha + \beta x_i + \gamma z_i + 誤差_i$$

まず，βを決めてから，α，γを決めることにする．この場合，α，βに関しても誤差の二乗和を最小にしなければならないから，最初の式を誤差が変わらないように移項して変形すると

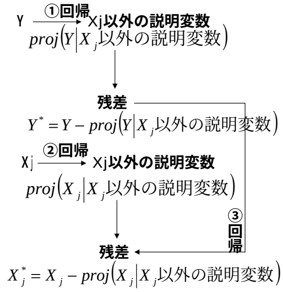
$$y_i - \hat{\beta} x_i = \alpha + \gamma z_i + 誤差_i$$

5．K個の説明変数の場合

- 回帰式
  - $Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + 誤差$
  - $y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i} + 誤差_i$
  - $x_{1,i}$ ：i番目のデータに関する1番目の説明変数の値
  - $x_{j,i}$ ：i番目のデータに関するj番目の説明変数の値
- 注意
  - 今までは説明変数には1が入っていた．つまり，定数項があったが，一般論では入っても入らなくても良い．
- 問題
  - β_j の推定法は？

5．1　3変数までの方法の延長(1)

- 再帰的方法
  - 3説明変数の場合の拡張
  - βjの最小二乗推定量や残差は，被説明変数Yと問題の説明変数Xjを残りの説明変数たちに回帰した残差同士の回帰の推定値や残差と同じ．
- K説明変数の回帰をK-1説明変数の回帰に還元
  - これを繰り返せば最終的に1説明変数の回帰に還元
- 実際はこれより直接的な方法がある



5．1　3変数までの方法の延長(2)

- βjに関する簡略化された回帰式

$$y_i^* = \beta_j x_{j,i}^* + 誤差_i$$

- 他の係数を推定するための回帰式

$$\begin{aligned} y_i - \hat{\beta}_j x_{j,i} &\\ = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_{j-1} x_{j-1,i} + \beta_{j+1} x_{j+1,i} + \cdots + \beta_K x_{K,i} + 誤差 \end{aligned}$$

5．2　直接的な方法(1)

- βjの満たすべき連立方程式を導く
  - 普通は微分を使って導出するが，本講義では最後まで二次関数アプローチを使う．
  - 誤差の二乗和は係数βjたちの2次式になり，2次の係数は $\sum_{i=1}^{n} x_{j,i}^2 > 0$ なのでその最小値は必ず存在する．
  - そのような最小値により最小二乗推定量 $\hat{\beta}_j$ たちがきまる．そのときの残差を $\hat{u}_i$ とする．
  - βjだけ $\hat{\beta}_j$ 以外の値にする．

5．2　直接的な方法(2)

- 誤差の分解

$$\begin{aligned} 誤差_i &= y_i - \left(\hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_{j-1} x_{j-1,i} + \beta_j x_{j,i} + \hat{\beta}_{j+1} x_{j+1,i} + \cdots + \hat{\beta}_K x_{K,i}\right) \\ &= \left\{y_i - \left(\hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_j x_{j,i} + \cdots + \hat{\beta}_K x_{K,i}\right)\right\} + \hat{\beta}_j x_{j,i} - \beta_j x_{j,i} \\ &= \hat{u}_i - \left(\beta_j - \hat{\beta}_j\right) x_{j,i} \end{aligned}$$

- 誤差の二乗和

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

5．2　直接的な方法(3)

- 結論

$$\sum_{i=1}^{n} 誤差_i^2 = \sum_{i=1}^{n} x_{j,i}^2 \left(\beta_j - \hat{\beta}_j - \frac{\sum_{i=1}^{n} x_{j,i} \hat{u}_i}{\sum_{i=1}^{n} x_{j,i}^2}\right)^2 - \frac{\left(\sum_{i=1}^{n} x_{j,i} \hat{u}_i\right)^2}{\sum_{i=1}^{n} x_{j,i}^2} + \sum_{i=1}^{n} \hat{u}_j^2$$

において，$\hat{\beta}_j$ で最小値を取るためには，$\sum x_{j,i} \hat{u}_i = 0$ となる必要がある．よって，ここでも残差と説明変数の積の合計は0となる．

5．3　正規方程式

- βjの決定
  - すべての説明変数について残差と説明変数の積の合計が0になるという連立方程式．これを解くと β_j が得られる．
  - 式で表すと，

$$\sum_{i=1}^{n} x_{1,i} \hat{u}_i = \sum_{i=1}^{n} \left\{y_i - \left(\beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i}\right)\right\} x_{1,i} = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{K,i} \hat{u}_i = \sum_{i=1}^{n} \left\{y_i - \left(\beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i}\right)\right\} x_{K,i} = 0$$

5．K説明変数の場合のまとめ（1）

- 係数推定値の決定
  - 再帰的アプローチ　　2説明変数，3説明変数と同じ
  - 直接的アプローチ　　正規方程式（2説明変数，3説明変数と同じ）
- 特定の係数だけの推定するための回帰式
  - 残りの係数を求める回帰式
- 残差と説明変数の関係
  - 残差と説明変数の積の合計＝0
  - 2説明変数，3説明変数と同じ

5．K説明変数の場合のまとめ（2）

- 説明力の指標（1）
  - RSS

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

  - 説明変数に1，つまり定数項が入っていない場合 $\sum_{i=1}^{n}\hat{u}_i=0$ とは限らないので，$\sum_{i=1}^{n}\hat{u}_i^2=$ 残差変動

とは限らない．したがって，この項がESSの一部とならない．だから，TSS＝RSS＋ESSとならない．よって決定係数は意味を持たない．

5．K説明変数の場合のまとめ（3）

- 説明力の指標（2）
  - 説明変数に定数項がある場合は3変数の場合と同様にTSS＝ESS＋RSSの分解ができるので，決定係数を説明力の指標とすることが出来る．
  - 説明変数を新たに加えたときの決定係数の増加分の説明である下の式が成立する．

$$R^2\left(\text{回帰式}Y=\beta_1X_1+\cdots+\beta_KX_K\right)$$
$$=\left(\text{回帰式}Y=\beta_1X_1+\cdots+\beta_{j-1}X_{j-1}+\beta_{j+1}X_{j+1}+\cdots+\beta_KX_K\text{の}R^2\right)$$
$$+\left(\text{回帰式}y_i=\beta x_{ji}^*+\text{誤差}_i\text{の}R^2\right)$$

6．マルチコリニアリティ（多重共線性）

- 説明変数の間に線形の関係，または，それに近い関係がある場合
  - cjを選ぶと $\sum_{j=1}^{K}c_jX_j=0$ と出来るような場合．
- 例
  - 国民所得＝消費＋投資であるが，それを考えずに説明変数に国民所得，消費，投資を入れてしまった場合
  - 定数項のある回帰で説明変数1＋説明変数2＝定数の場合
- 統計ソフトは結果を出さないか異常な結果を出す．
  - 係数値の符号が理論とあわない．また，t値が異常に小さい．
- 対策
  - 不要な変数を取り除く（相関係数が大きいもの同士を回帰に入れない）

6．1　多重共線性の理論的説明

- 3変数の場合

$Y=\alpha+\beta X+\gamma Z+$誤差とする

もし $X=2Z$ とすると，これを上の式に代入して

$Y=\alpha+(2\beta+\gamma)Z+$誤差　を得るから，

2β＋γは推定できても，βとγを別個に推定することはできない．