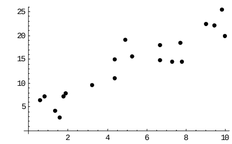
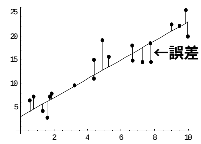
統計解析論特殊講義2

最小二乗法

1．最小二乗基準

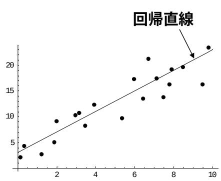
- 問題
  - 与えられたデータの組にもっとも当てはまる直線を求めよ
- 最小二乗基準
  - 「もっとも当てはまる」の意味？
  - その一つの答え
    - 誤差の二乗の合計が最小
    - 誤差とは右図の縦線の長さ





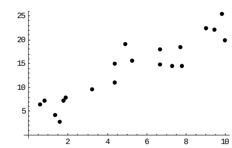
2．用語（1）

- 回帰直線
- 回帰式
  - 回帰直線を式で表したもの
  - 一般的には$Y=\alpha+\beta X$と表すことができる
  - 誤差を含めると
    $Y=\alpha+\beta X+$誤差項
- 説明変数　X
  - $Y=\alpha+\beta X=\alpha\times 1+\beta\times X$と解釈するとXと1が説明変数とも理解できる．
- 被説明変数　Y
  - X，Yは例えば所得，消費などの変数を表す



2．用語（2）

- 前のスライドは説明変数，被説明変数間の関係を扱った
- データの値を表す
  - データの値は小文字で表す
  - 何番目のデータかはその添え字で表す
    $(x_1,y_1),(x_2,y_2),\cdots$
    $y_i=\alpha+\beta x_i+$誤差項
- データ数＝標本数　n



i	x_i	X	Y	y_i
1	x_1	7.27	14.50	y_1
2	x_2	6.66	17.95	y_2
3	x_3	4.36	11.00	y_3
20	x_{20}	1.86	7.89	y_{20}

3.最小二乗法

- 最小二乗法
  - 最小二乗基準に従ってデータに最もフィットする回帰直線を求める．
- 問題の変形
  - 最小二乗基準→誤差の二乗の合計の最小化
  - 変化させるパラメータ＝係数→
    $Y=\alpha+\beta X+$誤差項 のα，β
  - 誤差の二乗の和をα，βの式で表す
    誤差$=y_i-(\alpha+\beta x_i)$

4．最小二乗法の解（1）

- 変形された問題
  - $\sum_{i=1}^{n}$誤差$_i^2=\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2$を最小化するα，βを求める．
  - このようなα，βをデータの値だけの式で表す．
- 解答のための方針
  - 最小化の二つの方向
    - 微分
    - 二次関数化

4. 最小二乗法の解 (2)

- **標本平均**
  - **例**
    - **Xのデータの標本平均** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
    - **Yのデータの値の二乗の標本平均** $\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$
  - **性質**

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} x_i - n\overline{x}$$

$$= n\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \overline{x}\right) = 0$$

$$\sum_{i=1}^{n} \text{定数}(x_i - \overline{x}) = \text{定数}\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

4. 最小二乗法の解 (3)

- **誤差の変形**

$$\text{誤差}_i = y_i - \alpha - \beta x_i$$
$$= (y_i - \overline{y}) + \overline{y} - \alpha - \beta(x_i - \overline{x}) - \beta\overline{x}$$
$$= (y_i - \overline{y}) - \beta(x_i - \overline{x}) + (\overline{y} - \alpha - \beta\overline{x})$$

4. 最小二乗法の解 (4)

- **誤差の二乗和の合計の展開**

$$\text{誤差}^2\text{の合計} = \sum_{i=1}^{n} \text{誤差}_i$$
$$= \sum_{i=1}^{n}\{(y_i - \overline{y}) - \beta(x_i - \overline{x}) + (\overline{y} - \alpha - \beta\overline{x})\}^2$$
$$= \sum_{i=1}^{n}\{(y_i - \overline{y})^2 + \beta^2(x_i - \overline{x})^2 + (\overline{y} - \alpha - \beta\overline{x})^2$$
$$-2\beta(y_i - \overline{y})(x_i - \overline{x}) + 2(y_i - \overline{y})(\overline{y} - \alpha - \beta\overline{x}) - 2\beta(x_i - \overline{x})(\overline{y} - \alpha - \beta\overline{x})\}^2$$
$$= \sum_{i=1}^{n}(y_i - \overline{y})^2 + \beta^2\sum_{i=1}^{n}(x_i - \overline{x})^2 + \sum_{i=1}^{n}(\overline{y} - \alpha - \beta\overline{x})^2$$
$$-2\beta\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) + 2\sum_{i=1}^{n}(y_i - \overline{y})(\overline{y} - \alpha - \beta\overline{x}) - 2\beta\sum_{i=1}^{n}(x_i - \overline{x})(\overline{y} - \alpha - \beta\overline{x})$$

4. 最小二乗法の解 (5)

- **ゼロになる項**
  - $\overline{y} - \alpha - \beta\overline{x}$ **はi毎に変わらない定数だから**
  - $\sum_{i=1}^{n}(x_i - \overline{x})(\overline{y} - \alpha - \beta\overline{x}) = (\overline{y} - \alpha - \beta\overline{x})\sum_{i=1}^{n}(x_i - \overline{x}) = 0$
  - $\sum_{i=1}^{n}(y_i - \overline{y})(\overline{y} - \alpha - \beta\overline{x}) = (\overline{y} - \alpha - \beta\overline{x})\sum_{i=1}^{n}(y_i - \overline{y}) = 0$
- **さらに変形できる項**

$$\sum_{i=1}^{n}(\overline{y} - \alpha - \beta\overline{x})^2 = (\overline{y} - \alpha - \beta\overline{x})^2\sum_{i=1}^{n} 1 = n(\overline{y} - \alpha - \beta\overline{x})^2$$

4. 最小二乗法の解 (6)

- **二次関数化と平方完成**

$$\text{誤差の２乗の和} = n(\alpha - \overline{y} + \beta\overline{x})^2 + \beta^2\sum_{i=1}^{n}(x_i - \overline{x})^2 - 2\beta\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) + \sum_{i=1}^{n}(y_i - \overline{y})^2$$

$$A \equiv \sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$B \equiv \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$ **とすると**

$$C \equiv \sum_{i=1}^{n}(y_i - \overline{y})^2$$

$$\text{誤差の２乗の和} = n(\alpha - \overline{y} + \beta\overline{x})^2 + A\beta^2 - 2B\beta + C$$
$$= n(\alpha - \overline{y} + \beta\overline{x})^2 + A\left(\beta - \frac{B}{A}\right)^2 - \frac{B^2}{A} + C$$

4. 最小二乗法の解 (7)

- **解＝最小二乗推定量**
  - **2乗の項を0にすれば最小になるので**

$$\alpha - \overline{y} + \beta\overline{x} = 0$$

$$\beta - \frac{B}{A} = 0$$

  - **よって，**

$$\beta = \frac{B}{A} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \equiv \hat{\beta}$$

$$\alpha = \overline{y} - \beta\,\overline{x} \equiv \hat{\alpha}$$

で最小．
  - **これらを回帰直線の傾きと切片の最小二乗推定量と呼ぶ**

4．最小二乗法の解ー練習問題ー

- **元の回帰式を**$Y=\alpha+\beta X$ **のαとβの最小二乗推定量を** $\hat{\alpha},\hat{\beta}$ **．新たに被説明変数を**$2Y+X$ **，説明変数を** $X-1$ **（と1）とした回帰式**

  $$2Y+X=\gamma+\delta(X-1)$$

  のγとδの最小二乗推定量を求めよ．

4．練習問題の答え（1）

- **まず** $2Y+X\text{の標本平均}=2\bar{y}+\bar{x}$

  $X-1\text{の標本平均}=\bar{x}-1$
- **よって，**

$$\hat{\delta}=\frac{\sum_{i=1}^{n}\{2y_i+x_i-(2Y+X)\text{の標本平均}\}\{x_i-1-(X-1)\text{の標本平均}\}}{\sum_{i=1}^{n}\{x_i-1-(X-1)\text{の標本平均}\}^2}$$

$$=\frac{\sum_{i=1}^{n}\{2y_i+x_i-2\bar{y}-\bar{x}\}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}\{2(y_i-\bar{y})+(x_i-\bar{x})\}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\frac{2\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})+\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=2\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+1=2\hat{\beta}+1$$

4．練習問題の答え（2）

- **つづき**

  $$\hat{\gamma}=\text{被説明変数の標本平均}-\hat{\delta}\times\text{説明変数の標本平均}$$

  $$=2\bar{y}+\bar{x}-(2\hat{\beta}+1)(\bar{x}-1)=2\bar{y}+\bar{x}-2\hat{\beta}\bar{x}+2\hat{\beta}-\bar{x}+1$$

  $$=2(\bar{y}-\hat{\beta}\bar{x})+2\hat{\beta}+1=2\hat{\alpha}+2\hat{\beta}+1$$
- **その他の練習問題**
  - **教科書p．35　7，8，9**
  - **元の回帰式Y＝α＋βXのαとβの最小二乗推定量を** $\hat{\alpha},\hat{\beta}$ **．新たに被説明変数をY－5X＋1，説明変数を2X＋1（と1）とした回帰式（Y－5X＋1）＝γ＋δ（2X＋1）のγとδの最小二乗推定量を求めよ．**

5．回帰値と残差（用語）

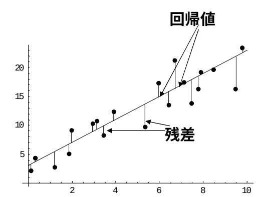
- **回帰値**

  $$\hat{y}_i\equiv\hat{\alpha}+\hat{\beta}x_i$$
- **残差**

  $$\hat{u}_i=y_i-\hat{y}_i=y_i-(\hat{\alpha}+\hat{\beta}x_i)$$

  ＝実際の値－回帰値
- **残差変動＝残差二乗和**

  $$RSS=\sum_{i=1}^{n}\hat{u}_i^2\geq 0$$
  - **実は誤差の2乗の合計の最小値**



6．被説明変数の変動の分解（0）

- **問題**
  - **残差変動はどのようにしてきまるか？**
  - **残差変動と被説明変数Yの変動との関係は？**
  - **残差変動と回帰値の変動との関係は？**
    - **ここで変動とは標本平均からの乖離の二乗の合計**
  - **回帰の成功度（説明力）を測る指標は？**
  - **説明変数と被説明変数の関連の程度を計るもう一つの指標相関係数との関係は？**

6．被説明変数の変動の分解（1）

- **残差変動の計算**
  - **残差変動は誤差の二乗和の最小値だから，**

    $\text{誤差の 2 乗の和}=n(\alpha-\bar{y}+\beta\bar{x})^2+A\left(\beta-\frac{B}{A}\right)^2-\frac{B^2}{A}+C$**より**

$$\text{残差変動}=C-\frac{B^2}{A}=\sum_{i=1}^{n}(y_i-\bar{y})^2-\frac{\left\{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})\right\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\sum_{i=1}^{n}(y_i-\bar{y})^2-\left\{\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$=\sum_{i=1}^{n}(y_i-\bar{y})^2-\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

6. 被説明変数の変動の分解 (2)

- $\sum_{i=1}^{n}(y_i-\bar{y})^2$ の解釈
  - 被説明変数の変動
    - これを**総変動** (TSS) と呼ぶ
      - 変動→標本平均からの乖離の二乗和を変動と呼ぶから
      - 「総」→これは後でわかる
    - 当然0以上

6. 被説明変数の変動の分解 (3)

- $\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$ の解釈 (何かの変動？)

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

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

$$\hat{\alpha}+\hat{\beta}\bar{x}=\hat{\alpha}+\hat{\beta}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}+\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}x_i=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha}+\hat{\beta}x_i\right)$$

$$=\text{回帰値の標本平均}$$

$$\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}\{\text{回帰値}-\text{回帰値の標本平均}\}^2$$

$$=\text{回帰値の変動}\equiv\text{回帰変動}\equiv ESS\geq 0$$

16. 被説明変数の変動の分解 (4)

- 残差変動＝総変動ー回帰変動

$$\text{残差変動}(RSS)=\sum_{i=1}^{n}(y_i-\bar{y})^2-\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$=\text{総変動}(TSS)-\text{回帰変動}(ESS)$$

- 総変動＝回帰変動＋残差変動
  - 被説明変数の変動は回帰値の変動と残差変動に分解できる．
  - 回帰変動は回帰で説明できる成分を表す．
  - 残差変動は回帰で説明できない成分を表す．

7. 回帰の説明力の指標

- 決定係数 (R^2)
  - 被説明変数の変動の内回帰値の変動で説明できる成分の割合 (大きければ説明力が大きい)

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

- 性質

$$0\leq R^2\leq 1$$

$ESS=\hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2\geq 0, TSS\geq 0$ より $R^2=\frac{ESS}{TSS}\geq 0$

$RSS\geq 0, TSS\geq 0$ より $R^2=1-\frac{RSS}{TSS}\leq 1$

8. 標本相関係数と決定係数の関係

- 決定係数＝標本相関係数の2乗

$$R^2=\frac{ESS}{TSS}=\frac{\hat{\beta}^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$=\left\{\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^2\times\frac{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$=\frac{\left\{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})\right\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}=\left\{\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right\}^2$$

8. 残差の性質 (0)

- 問題
  - 残差と被説明変数の値，回帰値の関係は定義から明らかである．では，説明変数との間になにか関係はないか？
  - $\sum_{i=1}^{n}\hat{u}_i^2$ が変動であるなら，残差 $\hat{u}_i$ の標本平均からの乖離の2乗の合計になっているはずである．だとすれば，残差 $\hat{u}_i$ は標本平均からの乖離となり，残差の標本平均は0のはずだが本当か？

9．残差の性質（1）

- （残差の値×説明変数の値）の合計＝0

  $\bar{y}-\hat{\alpha}-\hat{\beta}\bar{x}=0$ を利用すると

$$
\begin{aligned}
\sum_{i=1}^{n}\hat{u}_i x_i &= \sum_{i=1}^{n}\left(y_i-\hat{\alpha}-\hat{\beta}x_i\right)x_i \\
&= \sum_{i=1}^{n}\left\{(y_i-\bar{y})-\hat{\beta}(x_i-\bar{x})+\left(\bar{y}-\hat{\alpha}-\hat{\beta}\bar{x}\right)\right\}\left\{(x_i-\bar{x})+\bar{x}\right\} \\
&= \sum_{i=1}^{n}\left\{(y_i-\bar{y})-\hat{\beta}(x_i-\bar{x})\right\}\left\{(x_i-\bar{x})+\bar{x}\right\} \\
&= \sum_{i=1}^{n}\left\{(y_i-\bar{y})-\hat{\beta}(x_i-\bar{x})\right\}(x_i-\bar{x})+\bar{x}\sum_{i=1}^{n}\left\{(y_i-\bar{y})-\hat{\beta}(x_i-\bar{x})\right\} \\
&= \sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\hat{\beta}\sum_{i=1}^{n}(x_i-\bar{x})^2+\bar{x}\sum_{i=1}^{n}(y_i-\bar{y})-\hat{\beta}\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})
\end{aligned}
$$

9．残差の性質（2）

- 右辺各項の変形

$$\bar{x}\sum_{i=1}^{n}(y_i-\bar{y})=0$$

$$\hat{\beta}\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})=0$$

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$ より $$\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\hat{\beta}\sum_{i=1}^{n}(x_i-\bar{x})^2=0$$

- よって（残差×説明変数）の合計＝0

9．残差の性質（3）

- 残差の合計＝0（標本平均も0）

$$
\begin{aligned}
\sum_{i=1}^{n}\hat{u}_i &= \sum_{i=1}^{n}\left(y_i-\hat{\alpha}-\hat{\beta}x_i\right)=\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\hat{\alpha}-\sum_{i=1}^{n}\hat{\beta}x_i \\
&= \sum_{i=1}^{n}y_i-\hat{\alpha}\sum_{i=1}^{n}1-\hat{\beta}\sum_{i=1}^{n}x_i \\
&= n\bar{y}-\hat{\alpha}n-\hat{\beta}(n\bar{x})=n\left(\bar{y}-\hat{\alpha}-\hat{\beta}\bar{x}\right) \\
&= n\left\{\bar{y}-\left(\bar{y}-\hat{\beta}\bar{x}\right)-\hat{\beta}\bar{x}\right\}=0
\end{aligned}
$$

  - （残差×1）の合計＝0とも解釈できる
  - こちらも（残差×説明変数）の合計＝0

10．係数の変換と回帰式の変換（0）

- 問題
  - 回帰式の係数を求めたい式に変換すると便利な場合が特に検定の際にでてくる．それはどのような場合に可能で，どのような式ならそれを係数にもつ回帰式を作ることができるのか？
    - $Y=\alpha+\beta X$ から$\alpha+\beta-1$を係数とする回帰式が作れるか？
  - 被説明変数に説明変数の定数倍を加える，または，被説明変数からそれを引くと，回帰式はどのようになるか？あるいは被説明変数を定数倍すると何が起こるか？
  - 同様のことを説明変数に行ったら何が起きるか？
  - 被説明変数と説明変数の変換の組み合わせで何が起きる？
  - $\alpha+\beta-1$を係数とする回帰式を推定するとその係数の推定値は $\hat{\alpha}+\hat{\beta}-1$ か？

10．係数の変換と回帰式の変換（1）

- 係数を変換する
  - 元の回帰式　$Y=\alpha\times 1+\beta X$
  - 係数が $a\alpha+b\beta+c$ と $d\alpha+e\beta+f$ となる回帰式を作る．（ae-bdが0ではない場合）

$$
\begin{aligned}
Y&=\alpha+\beta X \\
&=(a\alpha+b\beta+c)\left(\frac{e-dX}{ae-bd}\right)+(d\alpha+e\beta+f)\left(\frac{-b+aX}{ae-bd}\right)-\frac{ce-fb+(af-cd)X}{ae-bd}
\end{aligned}
$$

  - 未知係数を持たない部分を左辺に移項する．

$$Y+\frac{ce-fb+(af-cd)X}{ae-bd}=(a\alpha+b\beta+c)\left(\frac{e-dX}{ae-bd}\right)+(d\alpha+e\beta+f)\left(\frac{-b+aX}{ae-bd}\right)$$

10．係数の変換と回帰式の変換（2）

- 係数変換の例
  - 元の回帰式　$Y=\alpha+\beta X+$誤差
  - 係数として $\alpha+\beta-1$を持つ回帰式を作る
    - まず右辺に$\alpha+\beta-1$が入るが，右辺全体としては元の回帰式と同じになるようにする

      $Y=(\alpha+\beta-1)-\beta+1+\beta X+$誤差
    - 右辺のほしい式以外に元の回帰式の係数α，βと関係ない定数が係数として入っている場合は，それを左辺に移項して被説明変数から引く

      $Y-1=(\alpha+\beta-1)-\beta+\beta X+$誤差
    - ほしい式以外の部分について元の回帰式の係数でくくる．

      $Y-1=(\alpha+\beta-1)+\beta(X-1)+$誤差

10．係数の変換と回帰式の変換（3）

- 別のやり方
  - まず右辺のXの係数に $\alpha+\beta-1$ が入るが，右辺全体としては元の回帰式と同じになるようにする．
    $$Y=\alpha+(\alpha+\beta-1)X-\alpha X+X+誤差$$
  - 右辺のほしい式以外に元の回帰式の係数α，βと関係ない定数が係数として入っている場合は，それを左辺に移項して被説明変数から引く
    $$Y-X=\alpha+(\alpha+\beta-1)X-\alpha X+誤差$$
  - ほしい式以外の部分について元の回帰式の係数でくくる．
    $$Y-X=\alpha(1-X)+(\alpha+\beta-1)X+誤差$$
  - 誤差が元の回帰式と同じになる様変換していることに注意

10．係数の変換と回帰式の変換（4）

- 被説明変数を変換する
  $$Y=\alpha+\beta X \rightarrow aY+b+cX=a\alpha+b+a\beta X+cX$$
  $$=(a\alpha+b)\times 1+(a\beta+c)X$$
  - 例
    被説明変数を $2Y-3X+5$
    $$2Y-3X+5=(2\alpha+5)+(2\beta-3)X$$
    （以下しばらく誤差を省略して書く，回帰直線の式と思ってほしい）

10.係数の変換と回帰式の変換（5）

- 被説明変数の単位変換
  - １００億円単位のデータを１０兆円に換える
    - データの桁数が大きいので3桁分へらす
    - 被説明変数を１／１０００倍する
- 数式で表現
  - $Y=\alpha+\beta X+誤差$ において $Y\rightarrow cY$
    $$cY=c\alpha+(c\beta)X+c誤差$$
    なので回帰係数はc倍される
- ３桁減らすと回帰係数は全て３桁減る

10．係数の変換と回帰式の変換（6）

- 説明変数の変換
  $$Y=\alpha\times 1+\beta X$$
  $$=\left(\frac{d\alpha-c\beta}{ad-bc}\right)(a+bX)+\left(\frac{-b\alpha+a\beta}{ad-bc}\right)(c+dX)$$
  （ad-bcが0ではない場合）
- 説明変数と被説明変数の組み合わせの時
  - どちらかを先にやって，後からもう一つの変換すればよい．

10．係数の変換と回帰式の変換（7）

- 説明変数の単位変換
  - １００億円単位のデータを１０兆円に換える
    - データの桁数が大きいので3桁分へらす
    - 説明変数を１／１０００倍する
- 数式で表現
  - $Y=\alpha+\beta X+誤差$ において $X\rightarrow cX$
    $$Y=\alpha+(\beta/c)(cX)+誤差$$
    なのでc倍された変数の回帰係数は1/c倍
- ３桁減らすとその回帰係数のみ３桁増える

10．係数の変換と回帰式の変換（8）

- 被説明変数と説明変数の両方をc倍したら？
  - 例
    - 両方を３桁減らす
    - Ｃ＝１／１０００
- 式変形
  $$cY=c\alpha+(c\beta)X+c誤差$$
  $$cY=c\alpha+(c\beta/c)(cX)+c誤差$$
- 定数項はc倍，傾きは同じになる
  - 定数項は３桁減る，傾きはそのまま

10. 係数の変換と回帰式の変換

- 以上の変換によって係数が $f(\alpha,\beta)$ になった場合，係数の最小二乗法による推定値は，$f(\hat{\alpha},\hat{\beta})$
  - 例
    - $\alpha+\beta-1$ を係数とする回帰式を推定するとその係数の推定値は $\hat{\alpha}+\hat{\beta}-1$ である．
  - 理由
    - 以上の変換は移項操作をするか，定数倍をしているだけだった．これは左辺と右辺の差＝誤差を元の回帰式と同じに保っている．従って，誤差の二乗は同じか定数倍されているだけである．
    - 本質的に同じものを最小化しているのだから，元の回帰式の係数＝パラメータα，βの関数として考えれば同じ値で最小値をとるはずである．
    - 従って，変換後の係数値の最小二乗推定量はα，βをその最小二乗推定量に置き換えたものになる．

11. 応用 (1)

- 回帰式 $Y=\alpha+\beta X$ のβの最小二乗推定量と残差は回帰式 $Y-\bar{y}=\beta(X-\bar{x})$ のβの最小二乗推定量と残差と同じである．
  - 理由
    - 先ほどの変換を施すと以下の様に回帰式を変換可
      $Y-\bar{y}=(\alpha+\beta\bar{x}-\bar{y})+\beta(X-\bar{x})$
    - これは元の式と残差は同じである．
    - 最小二乗法で回帰式を推定すると
      $\alpha+\beta\bar{x}-\bar{y}\to\hat{\alpha}+\hat{\beta}\bar{x}-\bar{y}=0$
    - よって $Y-\bar{y}=\beta(X-\bar{x})$

11. 応用 (2)

- 回帰式 $Y=\alpha+\beta X$ のαの最小二乗推定量と残差はβの最小二乗推定量を $\hat{\beta}$ としたとき，回帰式 $Y-\hat{\beta}X=\alpha+$誤差 のαの最小二乗推定量と残差と同じである．
  - 理由
    $Y-\hat{\beta}X=\alpha+(\beta-\hat{\beta})X+$誤差
    最小二乗法で推定すると $\beta-\hat{\beta}\to\hat{\beta}-\hat{\beta}=0$ なので，
    $Y-\hat{\beta}X=\alpha+$誤差

12. まとめ

- 回帰式とはなにか？
- 最小二乗法とはなにか？
- 最小二乗推定量を計算できるか？
- 回帰のあてはまりの良さの指標　　決定係数
- 残差とその性質
  - (残差×説明変数) の合計が0
- 回帰式の変形
  - 所望の係数，説明変数，被説明変数にする