Z_i は回帰式に含まれる説明変数でもよい。例えば，$Y_i = \alpha + \beta X_i + u_i$ について，u_i の平均はゼロ，分散は $\sigma_*^2 X_i^2$ とする場合，各変数を X_i で割って，

$$\begin{aligned}\frac{Y_i}{X_i} &= \alpha\frac{1}{X_i} + \beta + \frac{u_i}{X_i} \\ &= \alpha\frac{1}{X_i} + \beta + u_i^*\end{aligned}$$

を通常の最小二乗法で推定すればよい。β は定数項として推定されるが，意味は限界係数（すなわち，傾き）と同じなので注意すること。

● 不均一分散の例：　都道府県別『県民経済計算 2017 年版』

C_i = 家計最終消費支出（1 兆円）

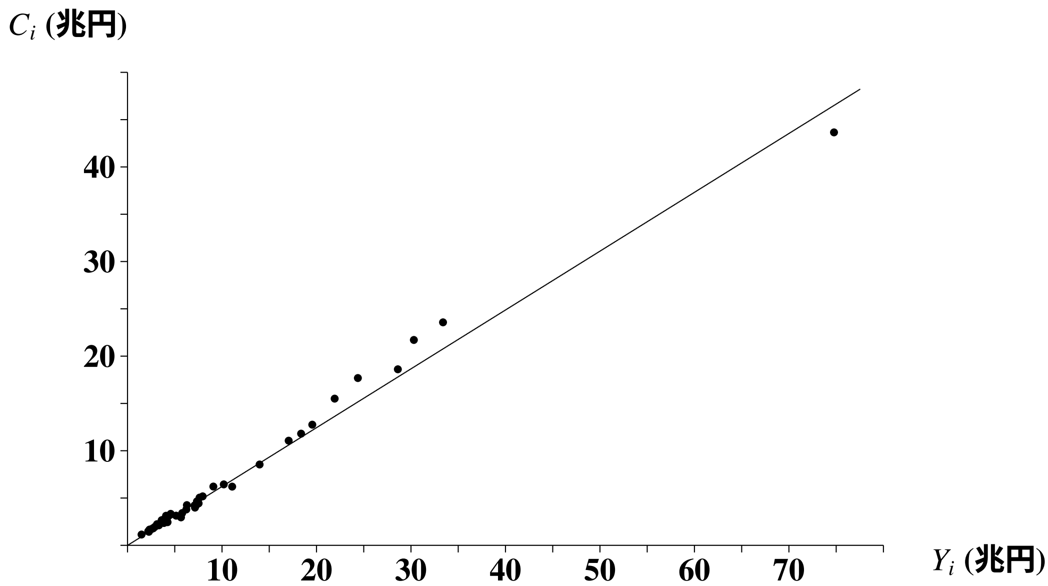
Y_i = 県民可処分所得（1 兆円）

L_i = 人口（千万人）

下記の消費関数を推定する。

$$C_i = \alpha + \beta Y_i + u_i, \qquad i = 1, 2, \cdots, n$$

$n = 47$ である。



$$C_i = \underset{(0.1684)}{\mathbf{0.0063}} + \underset{(0.0104)}{\mathbf{0.6219}}\, Y_i, \qquad R^2 = 0.9875, \quad \overline{R}^2 = 0.9872, \quad s = 0.8865$$

各係数推定値の下の ( ) 内は標準誤差を表す。

残差の二乗 $\hat{u}_i^2$ を人口の二乗 L_i に回帰させる。

$$\hat{u}_i^2 = -\underset{(\mathbf{0.1319})}{\mathbf{0.0838}} + \underset{(\mathbf{0.3720})}{\mathbf{5.6909}} L_i^2, \qquad R^2 = 0.8387, \quad \overline{R}^2 = 0.8351, \quad s = 0.8233$$

各係数推定値の下の ( ) 内は標準誤差を表す。

結果は，L_i^2 の係数推定値の t 値は 15.3 (= 5.6909/0.3720) で，L_i^2 の係数がゼロという仮説は棄却される。

よって，誤差項 u_i の分散 σ_i^2 は $\sigma_*^2 L_i^2$ と表される。

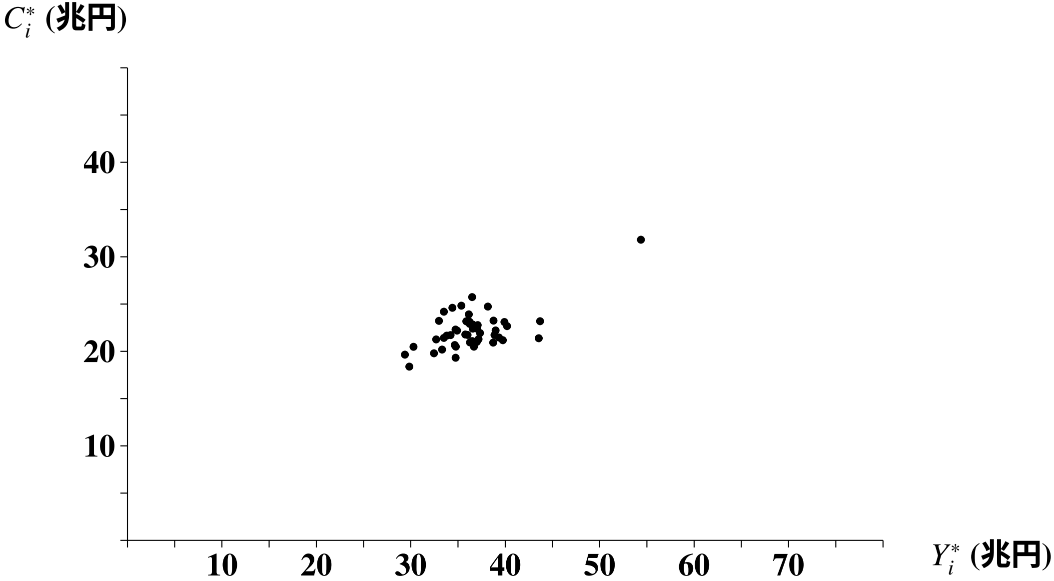
すなわち，$u_i \sim N(0, \sigma_*^2 L_i^2)$ となる。

データを変換して，下記の式を推定し直す。

$$C_i^* = \alpha L_i^* + \beta Y_i^* + u_i^*$$

ただし，$C_i^* = \dfrac{C_i}{L_i}$，$L_i^* = \dfrac{1}{L_i}$，$Y_i^* = \dfrac{Y_i}{L_i}$，$u_i^* = \dfrac{u_i}{L_i}$ とする。

αL_i^* は一定ではないので，C_i^* と Y_i^* の関係は直線ではないことに注意。



$$C_i^* = -\underset{(0.0647)}{\mathbf{0.1234}}\,L_i^* + \underset{(0.0137)}{\mathbf{0.6206}}\,Y_i^*, \qquad R^2 = 0.9933, \quad \overline{R}^2 = 0.9709, \quad s = 1.8671$$

各係数推定値の下の ( ) 内は標準誤差を表す。

7.3　誤差項：非正規分布のケース

(* 復習) 中心極限定理：

確率変数 X_1，X_2，$\cdots$，X_n は互いに独立とする。$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ とする。

$n \longrightarrow \infty$ のとき，$\dfrac{\overline{X} - \mathbf{E}(\overline{X})}{\sqrt{\mathbf{V}(\overline{X})}} \longrightarrow N(0,1)$ となる。ただし，$n\mathbf{V}(\overline{X}) < \infty$ とする。

(a)　すべての i について $\mathbf{E}(X_i) = \mu$，$\mathbf{V}(X_i) = \sigma^2$ とする。

$n \longrightarrow \infty$ のとき，$\dfrac{\overline{X} - \mathbf{E}(\overline{X})}{\sqrt{\mathbf{V}(\overline{X})}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$ となる。

$\mathbf{E}(\overline{X}) = \mu$，$\mathbf{V}(\overline{X}) = \dfrac{\sigma^2}{n}$ に注意。

(b) $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$ **は** $\sqrt{n}(\overline{X}-\mu) \longrightarrow N(0,\sigma^2)$ **と同じ意味である**

なぜなら，$\mathbf{V}\Big(\sqrt{n}(\overline{X}-\mu)\Big) = (\sqrt{n})^2\mathbf{V}(\overline{X}-\mu) = n\mathbf{V}(\overline{X}) = \sigma^2$ **となる。**

(c) $\mathbf{E}(X_i)=\mu$**，**$\mathbf{V}(X_i)=\sigma_i^2$ **（不均一分散），**$\sigma^2 = \lim_{n\to\infty}\dfrac{1}{n}\sum_{i=1}^{n}\sigma_i^2 < \infty$ **のとき，**

$\sqrt{n}(\overline{X}-\mu) \longrightarrow N(0,\sigma^2)$ **となる。**

中心極限定理を回帰分析に当てはめる。

単回帰：

$$Y_i = \alpha + \beta X_i + u_i$$

を考える。

- 誤差項 u_1，u_2，$\cdots$，u_n は互いに独立で，それぞれ平均ゼロ・分散 σ^2 とする（正規分布を仮定しない）。
- $n \longrightarrow \infty$ のとき，$\dfrac{1}{n}\sum_i (X_i - \overline{X})^2 \longrightarrow M < \infty$（一定の値）とする。

この場合，最小二乗推定量はどうなるか？

β の最小二乗推定量 $\hat{\beta}$ は次のように書き換えられる。

$$\hat{\beta} = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2} = \beta + \sum_i \omega_i u_i$$

ただし，$\omega_i = \dfrac{X_i - \overline{X}}{\sum_j (X_j - \overline{X})^2}$ とする。

$\hat{\beta}$ の分布を求めるために，まず $\hat{\beta} - \beta = \sum_i \omega_i u_i$ を利用して，$\sum_i \omega_i u_i$ の分布を求める。

$\sum_i \omega_i u_i$ が「(* 復習) 中心極限定理」の $\overline{X} = \frac{1}{n} \sum_i X_i$ に対応する。

すなわち，$n\omega_i u_i$ が「(* 復習) 中心極限定理」の X_i に対応する。

「(* 復習) 中心極限定理」の (c) に当てはめる。

$n\omega_i u_i$ の平均・分散は下記の通りとなる。

- $\mathbf{E}(n\omega_i u_i) = n\omega_i \mathbf{E}(u_i) = 0$
- $\mathbf{V}(n\omega_i u_i) = n^2\omega_i^2 \mathbf{V}(u_i) = n^2\omega_i^2\sigma^2$

$n^2\omega_i^2\sigma^2$ は「(* 復習) 中心極限定理」(c) の σ_i^2 に対応する。

まずは，「(* 復習) 中心極限定理」(c) の $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 < \infty$ が成り立つかどうかをチェックする。

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbf{V}(n\omega_i u_i) = \frac{1}{n}n^2\sigma^2\sum\omega_i^2 = \frac{\sigma^2}{\frac{1}{n}\sum(X_i-\overline{X})^2} \longrightarrow \frac{\sigma^2}{M} < \infty$$

$\sum_i \omega_i^2 = \dfrac{1}{\sum(X_i-\overline{X})^2}$ に注意。

よって，$\sqrt{n}\sum_i \omega_i u_i \longrightarrow N\left(0,\ \frac{\sigma^2}{M}\right)$ となる。

$\sqrt{n}\sum\omega_i u_i = \sqrt{n}(\hat{\beta}-\beta)$ なので，

$$\sqrt{n}(\hat{\beta}-\beta) \longrightarrow N\left(0,\ \frac{\sigma^2}{M}\right)$$

となる。

さらに，n が大きいとき，近似的に，

$$\hat{\beta} \sim N\left(\beta,\ \frac{\sigma^2}{\sum(X_i - \overline{X})^2}\right)$$

となる。

また，σ^2 をその推定量 s^2 に置き換えても，n が大きいとき，近似的に，

$$\hat{\beta} \sim N\left(\beta,\ \frac{s^2}{\sum(X_i - \overline{X})^2}\right)$$

となる（実践では，これが用いられる）。

結論：　誤差項 u_i の分布を仮定しなくても，データ数 n が大きければ，$\hat{\beta}$ の分布は正規分布で近似できる。

　重回帰 $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$ でも同様で，u_1，u_2，$\cdots$，u_n の分布を仮定しなくても（ただし，互いに独立で，平均ゼロ・分散 σ^2 の仮定は必要），n が大きいとき，$\hat{\beta}_j$ は近似的に下記のような正規分布を用いることができる。

$$\hat{\beta}_j \sim N(\beta_j,\ s^2 a_{jj})$$

$s\sqrt{a_{jj}}$ は $\hat{\beta}_j$ の標準誤差を表す。

a_{jj} については，11/10 講義ノート P.306 参照のこと。