$k=2$ の単純なモデル：

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \qquad i = 1, 2, \cdots, n$$

$\beta_1, \beta_2$ の最小二乗推定量は,

$$\min_{\beta_1,\, \beta_2} \sum_{i=1}^{n} (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i})^2$$

を解いて，$\hat{\beta}_1$，$\hat{\beta}_2$ が次のように得られる。

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \end{pmatrix}$$

$$= \frac{1}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i}X_{2i})^2} \begin{pmatrix} \sum X_{2i}^2 & -\sum X_{1i}X_{2i} \\ -\sum X_{1i}X_{2i} & \sum X_{1i}^2 \end{pmatrix} \begin{pmatrix} \sum X_{1i}Y \\ \sum X_{2i}Y \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{(\sum X_{2i}^2)(\sum X_{1i}Y_i) - (\sum X_{1i}X_{2i})(\sum X_{2i}Y_i)}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i}X_{2i})^2} \\ \dfrac{-(\sum X_{1i}X_{2i})(\sum X_{1i}Y_i) + (\sum X_{1i}^2)(\sum X_{2i}Y_i)}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i}X_{2i})^2} \end{pmatrix}$$

$\sum_{i=1}^{n} X_{ji}X_{li}$，$\sum_{i=1}^{n} X_{ji}Y_i$ をそれぞれ $\sum X_{ji}X_{li}$，$\sum X_{ji}Y_i$ と表記する。
ただし，$j = 1, 2$，$l = 1, 2$ とする。

一方，次の 2 つの回帰式を考える。

$$Y_i = \alpha_1 X_{2i} + v_i$$
$$X_{1i} = \alpha_2 X_{2i} + w_i$$

$\alpha_1$，$\alpha_2$ のそれぞれの最小二乗推定量を求めると，

$$\hat{\alpha}_1 = \frac{\sum X_{2i}Y_i}{\sum X_{2i}^2}, \qquad \hat{\alpha}_2 = \frac{\sum X_{2i}X_{1i}}{\sum X_{2i}^2}$$

となる。
$\hat{\alpha}_1$，$\hat{\alpha}_2$ を用いて，残差 $\hat{v}_i$，$\hat{w}_i$ を下記のようにそれぞれ求める。

$$\hat{v}_i = Y_i - \hat{\alpha}_1 X_{2i}, \qquad \hat{w}_i = X_{1i} - \hat{\alpha}_2 X_{2i}$$

$\hat{v}_i$，$\hat{w}_i$ は $Y_i$，$X_{1i}$ から $X_{2i}$ の影響を取り除いたものと解釈できる。

更に，次の回帰式を考える。

$$\hat{v}_i = \gamma \hat{w}_i + \epsilon_i$$

$\gamma$ の最小二乗推定量 $\hat{\gamma}$ は $\hat{\beta}_1$ に一致することを示す。

$$
\begin{aligned}
\hat{\gamma} &= \frac{\sum \hat{w}_i \hat{v}_i}{\sum \hat{w}_i^2} = \frac{\sum (X_{1i} - \hat{\alpha}_2 X_{2i})(Y_i - \hat{\alpha}_1 X_{2i})}{\sum (X_{1i} - \hat{\alpha}_2 X_{2i})^2} \\
&= \frac{\sum X_{1i} Y_i - \hat{\alpha}_1 \sum X_{1i} X_{2i} - \hat{\alpha}_2 \sum X_{2i} Y_i + \hat{\alpha}_1 \hat{\alpha}_2 \sum X_{2i}^2}{\sum X_{1i}^2 - 2\hat{\alpha}_2 \sum X_{1i} X_{2i} + \hat{\alpha}_2^2 \sum X_{2i}^2} \\
&= \frac{\sum X_{1i} Y_i - \dfrac{(\sum X_{2i} Y_i)(\sum X_{1i} X_{2i})}{\sum X_{2i}^2}}{\sum X_{1i}^2 - \dfrac{(\sum X_{1i} X_{2i})^2}{\sum X_{2i}^2}} \\
&= \frac{(\sum X_{2i}^2)(\sum X_{1i} Y_i) - (\sum X_{1i} X_{2i})(\sum X_{2i} Y_i)}{(\sum X_{1i}^2)(\sum X_{2i}^2) - (\sum X_{1i} X_{2i})^2} = \hat{\beta}_1,
\end{aligned}
$$

「$Y_i$ から $X_{2i}$ の影響を取り除いた変数」を被説明変数,「$X_{1i}$ から $X_{2i}$ の影響を取り除いた変数」を説明変数とした回帰係数が $\beta_1$ に等しい。

**一般化：** 次の回帰モデルを考える。

$$
Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}
$$

$j$ 番目の回帰係数 $\beta_j$ の意味は,「$Y_i$ から $X_{1i}, \cdots, X_{j-1,i}, X_{j+1,i}, \cdots, X_{ki}$ (すなわち，$X_{ji}$ 以外の説明変数) の影響を取り除いた変数」を被説明変数,「$X_{ji}$ から $X_{1i}, \cdots, X_{j-1,i}, X_{j+1,i}, \cdots, X_{ki}$ (すなわち，$X_{ji}$ 以外の説明変数) の影響を取り除いた変数」を説明変数とした回帰係数となる。

### 4.3.2　決定係数 $R^2$ と自由度修正済み決定係数 $\overline{R}^2$ について

また，決定係数 $R^2$ についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^n(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^n(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n(Y_i - \overline{Y})^2}$$

ただし，$\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$，$Y_i = \hat{Y}_i + \hat{u}_i$ である。

$R^2$ は，説明変数を増やすことによって，必ず大きくなる。なぜなら，説明変数が増えることによって，$\sum_{i=1}^n \hat{u}_i^2$ が必ず減少するからである。

$R^2$ を基準にすると，被説明変数にとって意味のない変数でも，説明変数が多いほど，よりよいモデルということになる。この点を改善するために，自由

度修正済み決定係数 $\overline{R}^2$ を用いる。

$$\overline{R}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2/(n-k)}{\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)},$$

$\sum_{i=1}^n \hat{u}_i^2/(n-k)$ は $u_i$ の分散 $\sigma^2$ の不偏推定量であり，$\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)$ は $Y_i$ の分散の不偏推定量である。分散や不偏推定量の意味は，統計学の知識を必要とし，後述する。

$R^2$ と $\overline{R}^2$ との関係は,

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k},$$

となる。さらに,

$$\frac{1 - \overline{R}^2}{1 - R^2} = \frac{n-1}{n-k} \geq 1,$$

という関係から，$\overline{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみに，等号が成り立つ。)

**数値例：** 今までと同じ数値例で，$\overline{R}^2$ を計算する。

| $i$ | $X_i$ | $Y_i$ | $X_i^2$ | $X_iY_i$ | $\hat{Y}_i$ | $\hat{u}_i$ | $X_i\hat{u}_i$ | $\hat{Y}_i\hat{u}_i$ | $\hat{u}_i^2$ | $Y_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 4 | 25 | 20 | 4.0 | 0.0 | 0.0 | 0.00 | 0.00 | 16 |
| 2 | 1 | 1 | 1 | 1 | 1.2 | −0.2 | −0.2 | −0.24 | 0.04 | 1 |
| 3 | 3 | 1 | 9 | 3 | 2.6 | −1.6 | −4.8 | −4.16 | 2.56 | 1 |
| 4 | 2 | 3 | 4 | 6 | 1.9 | 1.1 | 2.2 | 2.09 | 1.21 | 9 |
| 5 | 4 | 4 | 16 | 16 | 3.3 | 0.7 | 2.8 | 2.31 | 0.49 | 16 |
| 合計 | $\sum X_i$ | $\sum Y_i$ | $\sum X_i^2$ | $\sum X_iY_i$ | $\sum \hat{Y}_i$ | $\sum \hat{u}_i$ | $\sum X_i\hat{u}_i$ | $\sum \hat{Y}_i\hat{u}_i$ | $\sum \hat{u}_i^2$ | $\sum Y_i^2$ |
| | 15 | 13 | 55 | 46 | 13 | 0.0 | 0.0 | 0.0 | 4.3 | 43 |
| 平均 | $\overline{X}$ | $\overline{Y}$ | | | | | | | | |
| | 3 | 2.6 | | | | | | | | |

$\overline{Y} = 2.6$, $\displaystyle\sum_{i=1}^{n} \hat{u}_i^2 = 4.3$, $\displaystyle\sum_{i=1}^{n} Y_i^2 = 43$ なので，

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum Y_i^2 - n\overline{Y}^2} = 1 - \frac{4.3}{43 - 5 \times 2.6^2} = 1 - \frac{4.3}{9.2} = 0.5326$$

となり，$\overline{R}^2$ は，

$$\overline{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{(\sum Y_i^2 - n\overline{Y}^2)/(n-1)} = 1 - \frac{4.3/(5-2)}{9.2/(5-1)} = 0.3768$$

となる。

**自由度について：**　分子について，残差 $\hat{u}_i$ を求めるためには，$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ の $k$ 個の推定値を得なければならない。データ数 $n$ から推定値の数 $k$ を差し引いたものを**自由度** (degree of freedom) と呼ぶ。

一方，分母については，$X_{1i}$ が定数項だとして，$Y_i$ が定数項を除く $X_{2i}, X_{3i}, \cdots, X_{ki}$ に依存しない場合を考える。この場合，$\beta_2 = \beta_3 = \cdots = \beta_k = 0$ とするので，$\hat{u}_i = Y_i - \hat{\beta}_1$ となる。$\hat{u}_i$ を得るためには $\hat{\beta}_1$ だけを求めればよい。最小二乗法の考え方に沿って求めれば，$\hat{\beta}_1 = \overline{Y}$ となる（読者は確認すること）。すなわち，自由度は「データ数 − 推定値の数 $= n - 1$」ということになる。

このように，決定係数の第二項目の分子・分母をそれぞれの自由度で割ることによって，自由度修正済み決定係数が得られる。

**注意：**　$R^2$ や $\overline{R}^2$ を比較する場合，被説明変数が同じであることが重要である。被説明変数が対数かまたはそのままの値であれば，決定係数・自由度修正済み決定係数の大小比較は意味をなさない。ただし，被説明変数が異なる場合であっても，被説明変数を上昇率とするかそのままの値を用いるかの比較では，決定係数・自由度修正済み決定係数の大小比較はできないが，誤差項 $u_i$ の標準誤差での比較は可能である (標準誤差の小さいモデルを採用する)。$\Longrightarrow$ 関数型の選択

# 第5章　統計学の基礎：復習

## 5.1　確率変数，確率分布について

確率変数は，通常，大文字のアルファベット（例えば，$X$）で表すのに対して，実際に起こった値（すなわち，実現値）を小文字（例えば，$x$）で表す。

確率変数には離散型確率変数と連続型確率変数がある。まず，離散型確率変数 $X$ を考える。$X$ の取り得る値は分かっている。例えば，$X = x_1, x_2, \cdots, x_n$ の $n$ 通りの値を取るものとする。それぞれの値には確率が割り当てられる。すなわち，$\mathrm{Prob}(X = x_i) = p_i$ と表記し,「確率変数 $X$ が $x_i$ を取る確率は $p_i$ である」と読む。$p_i$ は確率であり，しかも，$X$ は $x_1, x_2, \cdots, x_n$ のいずれかの値を取るので，$\sum_{i=1}^{n} p_i = 1$ となる。また，$p_i$ は $x_i$ の関数であり，$f(x_i)$ と表すことができ

る。$f(x_i)$ を確率関数と呼ぶ。$f(x_i)$ は，(i) $f(x_i) \geq 0$，(ii) $\sum_{i=1}^{n} f(x_i) = 1$ を満たす関数でなければならない。

$X$ をサイコロを投げて出た目としよう。このとき，$X$ の取る値は 1，2，3，4，5，6 で，それぞれの目が出る確率は $\dfrac{1}{6}$ となる。したがって，$x_i = i$，$p_i = \dfrac{1}{6}$，$i = 1, 2, 3, 4, 5, 6$ となる。

$X$ が連続型確率変数の場合は，ある値 $a$ から別の値 $b$ までの区間に入る確率 $\mathrm{Prob}(a < X < b)$ という意味になる（ただし，$a < b$）。この場合，$f(x)$，$x = a$，$x = b$，$x$ 軸で囲まれた面積が確率を表すことになる。すなわち，

$$\mathrm{Prob}(a < X < b) = \int_a^b f(x)\mathrm{d}x,$$

となり，$f(x)$ を確率密度関数，または，密度関数と呼ぶ。$f(x)$ は，(i) $f(x) \geq 0$，(ii) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ を満たす連続関数でなければならない。

離散型の $f(\cdot)$ と連続型の $f(\cdot)$ の違いは，前者は $f(\cdot)$ そのものが確率を表すのに対して，後者の $f(\cdot)$ は面積が確率を表す（すなわち，連続型の $f(\cdot)$ の高さは確率を表さない）。

**分布関数（累積分布関数）：**　分布関数（累積分布関数）$F(x)$ は,

$$F(x) = \mathrm{Prob}(X \leq x) = \begin{cases} \displaystyle\sum_{i=1}^{r} f(x_i) & X \text{ が離散型確率変数のとき} \\ \displaystyle\int_{-\infty}^{x} f(t)\mathrm{d}t & X \text{ が連続型確率変数のとき} \end{cases}$$

ただし，離散型の場合，$r$ は $x_r \leq x < x_{r+1}$ となる $r$ である。すなわち，離散型の場合，$F(x)$ は 0 と 1 の間の階段状（階段関数）となる。

**同時確率分布：**　2 つの確率変数 $X, Y$ を考える。離散型の場合，$X$ の取る値を $x_1, X_2, \cdots, x_n$ とし，$Y$ の取る値を $y_1, y_2, \cdots, y_m$ としたとき，$X$ が $x_i$ を取り，かつ，$Y$ が $y_j$ を取る確率を同時確率分布と呼び，下記のように表す。

$$\mathrm{Prob}(X = x_i, Y = y_j) = p_{ij}$$

$p_{ij}$ は $x_i$, $y_j$ の関数となり，$p_{ij} = f(x_i, y_j)$ と表す。$f(x_i, y_j)$ を同時確率関数と呼ぶ。

連続型の場合は，$X$ が $c$ と $d$ の間の値（ただし，$a < b$）を取り，かつ，$Y$ が $c$ と $d$ の間の値（ただし，$c < d$）を取る確率は，下記のように表される。

$$\mathrm{Prob}(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y)\mathrm{d}y\mathrm{d}x$$

$f(x, y)$ を同時確率密度関数（または，同時密度関数）と呼ぶ。

## 5.2　期待値・分散・共分散の定義・定理

### 5.2.1　期待値の定義

**定義 (期待値，1 変数)：**　確率変数 $X$，ある関数 $g(\cdot)$ とするとき，$g(X)$ の期待値は次のように定義される。

$$\mathrm{E}(g(X)) = \begin{cases} \displaystyle\sum_{i=1}^{n} g(x_i)f(x_i), & X \text{ が離散型確率変数のとき} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x, & X \text{ が連続型確率変数のとき} \end{cases} \tag{5.1}$$

ただし，$f(\cdot)$ は確率関数（離散型のとき），または，密度関数（連続型のとき）を表す。

**定義 (期待値，2 変数)：**　確率変数 $X$，$Y$，ある関数 $g(\cdot,\cdot)$ とするとき，$g(X,Y)$ の期待値は次のように定義される。

$$
\mathrm{E}(g(X,Y)) = \begin{cases} \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i,y_j)f(x_i,y_j), \\ \qquad\qquad X,\ Y \text{ が離散型確率変数のとき} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\mathrm{d}y\mathrm{d}x, \\ \qquad\qquad X,\ Y \text{ が連続型確率変数のとき} \end{cases} \tag{5.2}
$$

ただし，$f(\cdot,\cdot)$ は確率関数（離散型のとき），または，密度関数（連続型のとき）を表す。

2 変数 $(X, Y)$ を $n$ 変数 $(X_1, X_2, \cdots, X_n)$ に拡張することも出来る。

### 5.2.2　期待値の定理

**定理 (1 変数)：**　$X$ を確率変数とする。$a+bX$ の期待値は,

$$\mathrm{E}(a+bX)=a+b\mathrm{E}(X), \tag{5.3}$$

となる。ただし，$a, b$ は定数とする。$g(X)=a+bX$ に対応する。

**定理 (2 変数)：**　$X, Y$ を確率変数とする。$X+Y$ の期待値は,

$$\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y), \tag{5.4}$$

となる。$g(X, Y)=X+Y$ に対応する。

**定理 (多変数)：**　$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ を考える。このとき，$\sum_{i=1}^{n} c_i X_i$ の平均は,

$$\mathrm{E}\Big(\sum_{i=1}^{n} c_i X_i\Big)=\sum_{i=1}^{n} c_i \mathrm{E}(X_i), \tag{5.5}$$

となる。

### 5.2.3 分散・共分散の定義・定理

**定義 (1 変数)：** $X$ を確率変数とする。$X$ の分散 $\sigma^2 = \mathrm{V}(X)$ は,

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}((X-\mu)^2), \tag{5.6}$$

である。ただし，$\mu = \mathrm{E}(X)$ とする。$g(X) = (X-\mu)^2$ に対応する。

**定義 (1 変数)：** $X$ を確率変数とする。$X$ の標準偏差 $\sigma$ は,

$$\sigma = \sqrt{\mathrm{V}(X)} \tag{5.7}$$

である。

**定理 (1 変数)：** $X$ を確率変数とする。$X$ の分散は,

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2, \tag{5.8}$$

と書き換えられる。ただし，$\mu = \mathrm{E}(X)$ とする。

**定理 (1 変数)：**　$X$ を確率変数とする。$a + bX$ の分散は,

$$\mathrm{V}(a + bX) = \mathrm{V}(bX) = b^2\mathrm{V}(X), \tag{5.9}$$

となる。ただし，$a, b$ は定数とする。

**定理 (1 変数)：**　$X$ を平均 $\mu$，分散 $\sigma^2$ の確率変数とする。$Z = \dfrac{X - \mu}{\sigma}$ について,

$$\mathrm{E}(\mathrm{Z}) = 0, \qquad \mathrm{V}(\mathrm{Z}) = 1, \tag{5.10}$$

となる。この変換を標準化，または，基準化と呼ぶ。

**定義 (2 変数)：**　$X$，$Y$ を確率変数とする。$X$ と $Y$ の共分散 $\sigma_{XY} = \mathrm{Cov}(X, Y)$ は,

$$\sigma_{XY} = \mathrm{Cov}(X, Y) = \mathrm{E}((X - \mu_X)(Y - \mu_Y)), \tag{5.11}$$

となる。$\mathrm{Cov}(X, Y)$ について，$g(X, Y) = (X - \mu_X)(Y - \mu_Y)$ に対応する。

**定義 (2 変数)：** $X$，$Y$ を確率変数とする。$X$ と $Y$ の相関係数 $\rho_{XY}$ は,

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \tag{5.12}$$

となる。ただし，$\sigma_X^2 = \mathrm{V}(X)$，$\sigma_Y^2 = \mathrm{V}(Y)$ とする。

**定理 (2 変数)：** $X$，$Y$ を確率変数とする。$X$ と $Y$ の共分散は,

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mu_X \mu_Y, \tag{5.13}$$

と書き換えられる。$\mathrm{E}(XY)$ について，$g(X, Y) = XY$ に対応する。

**定理 (2 変数)：** $X$，$Y$ を確率変数とする。$X + Y$ の分散は,

$$\mathrm{V}(X + Y) = \mathrm{V}(X) + 2\mathrm{Cov}(X, Y) + \mathrm{V}(Y), \tag{5.14}$$

となる。

**定理 (2 変数)：**　$X$，$Y$ を確率変数とする。$X$ と $Y$ が独立のとき，$X$ と $Y$ の共分散は,

$$\mathrm{Cov}(X, Y) = 0, \tag{5.15}$$

となる。

**定理 (2 変数)：**　$X$，$Y$ を確率変数とする。$X$ と $Y$ が独立のとき，$X + Y$ の分散は,

$$\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y), \tag{5.16}$$

となる。

**定理 (多変数)：**　$n$ 個の独立な確率変数 $X_1, X_2, \cdots, X_n$ を考える。このとき，$\sum_{i=1}^n c_i X_i$ の分散は,

$$\mathrm{V}(\sum_{i=1}^n c_i X_i) = \sum_{i=1}^n c_i^2 \mathrm{V}(X_i), \tag{5.17}$$

となる。

## 5.3　正規分布について

確率変数 $X$ の密度関数 $f(x)$ が,

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big),$$

となるとき，$f(x)$ を正規分布と呼ぶ。ただし，$\exp(x) = e^x$ である。$e$ は自然対数の底と呼ばれ，$e = \lim_{n\to\infty}\Big(1 + \frac{1}{n}\Big)^n = 2.718281828459045235360287 4713...$ と定義される。

上記の正規分布は,

$$\mathrm{E}(X) = \mu, \qquad \mathrm{V}(X) = \sigma^2,$$

となる（期待値の定義通りに計算すればよい）。

確率変数 $X$ が上記の密度関数 $f(x)$ となるとき，$X \sim N(\mu, \sigma^2)$ と表す。$X \sim N(\mu, \sigma^2)$ とは,「$X$ は平均 $\mu$，分散 $\sigma^2$ の正規分布に従う」と言う意味である。すなわち，$N$ は正規分布 (Normal distribution) のアルファベットの頭文字で，$\sim$ は「に従う」と読む。

**定理（標準化，基準化）：**　(5.10) のように $X$ を基準化する。

$$X \sim N(\mu, \sigma^2) \quad \text{のとき，} \quad Z = \frac{X-\mu}{\sigma} \sim N(0,1) \tag{5.18}$$

基準化によって，$X$ がどの分布に従う確率変数であっても，平均 0，分散 1 に変換することができるということを (5.10) の定理は示している。(5.18) では，さらに進んで，$X$ が正規分布であれば，$Z$ も正規分布となるということを言っている。この証明は，変数変換（置換積分）を利用して証明することになる（本書では証明略）。平均 0，分散 1 の正規分布 $N(0,1)$ は，標準正規分布と呼ばれる。

標準正規分布の確率分布表があれば，一般の正規分布の確率を得ることができる。すなわち，$\mu$ と $\sigma^2$ が既知とするとき，$Z$ が $z$ より大きい確率 $\mathrm{Prob}(Z>z)$ について，$\mathrm{Prob}(Z>z)=\mathrm{Prob}(X>\mu+z\sigma)$ となる。同様に，$X$ が $x$ より大きい確率 $\mathrm{Prob}(X>x)$ について，$\mathrm{Prob}(X>x)=\mathrm{Prob}\Big(Z>\dfrac{x-\mu}{\sigma}\Big)$ となる。320 ページの付表 1 を用いると，標準正規分布の確率，すなわち，$\mathrm{Prob}(Z>z)$ を求めることができる。

(5.5) 式と (5.16) 式によって，$n$ 個の独立な確率変数 $X_1, X_2, \cdots, X_n$ が同一の

分布（平均，分散が同じ分布）に従うとき，$\sum_{i=1}^{n} c_i X_i$ の平均，分散は，

$$E(\sum_{i=1}^{n} c_i X_i) = \mu \sum_{i=1}^{n} c_i, \qquad V(\sum_{i=1}^{n} c_i X_i) = \sigma^2 \sum_{i=1}^{n} c_i^2$$

となる。ただし，すべての $i$ について $\mu = E(X_i), \sigma^2 = V(X_i)$ とする。

$n$ 個の独立な確率変数 $X_1, X_2, \cdots, X_n$ が同一の正規分布に従うものとする。すなわち，すべての $i$ について $X_i \sim N(\mu, \sigma^2)$ とする。このとき，

$$\sum_{i=1}^{n} c_i X_i \sim N(\mu \sum_{i=1}^{n} c_i, \sigma^2 \sum_{i=1}^{n} c_i^2)$$

となる。すなわち，正規分布に従う確率変数の加重和もまた正規分布となる。この証明はそれほど簡単ではなく，積率母関数を利用して証明することになる（本書では証明略）。

特に，標本平均 $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ を考えると，

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

となる（すべての $i$ について，$c_i = \dfrac{1}{n}$ の場合を考えればよい）。

## 5.4　統計値・統計量，推定値・推定量について

1. 理論標本，理論観測値 $\Longrightarrow X_1, X_2, \cdots, X_n \Longrightarrow$ 確率変数
2. 実現された標本，実現された観測値，実現値，観測値 $\Longrightarrow x_1, x_2, \cdots, x_n$ $\Longrightarrow$ 観測データ

1. 理論観測値 $X_1, X_2, \cdots, X_n$ の関数 $\Longrightarrow$ 統計量
2. すべての $i$ について，$\mu = \mathrm{E}(X_i)$ と仮定する。
3. 母平均 $\mu$ の推定に使われる統計量 $\Longrightarrow \mu$ の推定量
   (a) $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ は $\mu$ の推定量

(b) $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ は $\sigma^2$ の推定量

4. 実現された標本を用いて実際に計算された推定量の値 $\Longrightarrow$ 推定値

(a) $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ は $\mu$ の推定値

(b) $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ は $\sigma^2$ の推定値

5. $\mu$ や $\sigma^2$ の推定量の候補は無数に考えられる。

# 5.5　大数の法則と中心極限定理

## 5.5.1　大数の法則

**大数の法則：その 1**　$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ は互いに独立ですべて同じ分布にしたがい，すべての $= 1, 2, \cdots, n$ について $\mathrm{E}(X_i) = \mu$ とする。$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$（すなわち，標本平均）とする。

$n \longrightarrow \infty$ のとき，

$$\overline{X} \longrightarrow \mu$$

となる。

**大数の法則：その 2**　$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ を考える（互いに独立である必要はなく，同じ分布である必要もない）。

$$\mu = \lim_{n\to\infty} \frac{1}{n}\mathrm{E}(\sum_{i=1}^{n} X_i) < \infty, \qquad \sigma^2 = \lim_{n\to\infty} \frac{1}{n}\mathrm{V}(\sum_{i=1}^{n} X_i) < \infty$$

とする。

　$n \longrightarrow \infty$ のとき,

$$\overline{X} \longrightarrow \mu$$

となる。

### 5.5.2　中心極限定理

**中心極限定理：その1**　$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ は互いに独立ですべて同じ分布にしたがい，すべての $= 1, 2, \cdots, n$ について $\mathrm{E}(X_i) = \mu$，$\mathrm{V}(X_i) = \sigma^2$ とする。$\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ とする。

　$n \longrightarrow \infty$ のとき,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1)$$

となる。$\mathrm{E}(\overline{X}) = \mu$，$\mathrm{V}(\overline{X}) = \sigma^2/n$ に注意せよ。

**中心極限定理：その 2**　$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ を考える（互いに独立である必要はなく，同じ分布である必要もない）。

$$\mu = \lim_{n\to\infty} \frac{1}{n} \mathrm{E}(\sum_{i=1}^{n} X_i) < \infty, \qquad \sigma^2 = \lim_{n\to\infty} \frac{1}{n} \mathrm{V}(\sum_{i=1}^{n} X_i) < \infty$$

とする。

$n \longrightarrow \infty$ のとき，

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \longrightarrow N(0,1)$$

となる。

## 5.6　推定量の望ましい性質

$\hat{\alpha}$, $\hat{\beta}$ の性質を求めるために

## 5.6.1　不偏性

ある母集団のある母数 $\theta$ に対して，$\theta$ の推定量として $\hat{\theta}$ を考える。このとき,

$$\mathrm{E}(\hat{\theta}) = \theta$$

となるとき, $\hat{\theta}$ は $\theta$ の不偏推定量であると言う。$\hat{\theta}$ は不偏性を持つと言う。$\mathrm{E}(\hat{\theta})-\theta$ は偏りと定義される。

$n$ 個の確率変数 $X_1$, $X_2$, $\cdots$, $X_n$ に関して，すべての $= 1, 2, \cdots, n$ について $\mathrm{E}(X_i) = \mu$ とするとき，標本平均 $\overline{X}$ は $\mu$ の不偏推定量である。

証明：

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

このように，$\mathrm{E}(\overline{X}) = \mu$ なので，標本平均 $\overline{X}$ は $\mu$ の不偏推定量となる。

## 5.6.2　有効性 (最小分散性)

ある母数 $\theta$ に対して，$\hat{\theta}_1$ と $\hat{\theta}_2$ の 2 つの不偏推定量を考える。このとき, $\mathrm{V}(\hat{\theta}_1) \leq \mathrm{V}(\hat{\theta}_2)$ が成り立つとき，$\hat{\theta}_1$ は $\hat{\theta}_2$ より有効であると言う。

ある母数 $\theta$ に対して，可能なすべての不偏推定量を考え，$\hat{\theta}$ が最も小さな分散を持つ不偏推定量であるとする。このとき，$\hat{\theta}$ を最小分散不偏推定量，または，最良不偏推定量と言う。

一般に，有効推定量が存在するとは限らない。代わりに，推定量 $\displaystyle\sum_{i=1}^{n} c_i X_i$（すなわち，線形推定量）の中で最も小さい分散を持つ推定量を求めることを考える。この推定量を最良線形不偏推定量と呼ぶ。

標本平均 $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ は不偏推定量の中で最も小さな分散を持つ推定量である。

証明：

期待値を取ると，

$$\mathrm{E}(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i \mathrm{E}(X_i) = \mu \sum_{i=1}^{n} c_i$$

となる。$\sum_{i=1}^{n} c_i X_i$ が不偏推定量になるためには $\sum_{i=1}^{n} c_i = 1$ が必要となる。分散は,

$$V(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} V(c_i X_i) = \sum_{i=1}^{n} c_i^2 V(X_i) = \sigma^2 \sum_{i=1}^{n} c_i^2$$

となる。

したがって，最良線形不偏推定量を得るためには，$\sum_{i=1}^{n} c_i = 1$ の条件のもとで，$\sum_{i=1}^{n} c_i^2$ を最小にする $c_1, c_2, \cdots, c_n$ を求めればよい。ラグランジェ未定乗数法を用いれば，$c_i = \frac{1}{n}$ が得られる。

### 5.6.3　一致性

ある母数 $\theta$ について推定量 $\hat{\theta}$ を考える。$n$ 個の標本から構成された推定量を $\hat{\theta}^{(n)}$ と定義する。数列 $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \cdots, \hat{\theta}^{(n)}, \cdots$ を考える。十分大きな $n$ について,

$\hat{\theta}^{(n)}$ が $\theta$ に確率的に収束するとき，$\hat{\theta}$ は $\theta$ の一致推定量であると言う。

$$\hat{\theta} \longrightarrow \theta, \quad \text{または，} \quad \operatorname{plim} \hat{\theta} = \theta,$$

と表現する。plim とは probability limit の略である。

$\mathrm{E}(\hat{\theta}) = \theta$ とする。$n \to \infty$ のとき $\mathrm{V}(\hat{\theta}) \to 0$ が成り立てば，$\hat{\theta}$ は $\theta$ の一致推定量である。

$\mu$ の推定量 $\overline{X}$ を調べる。

$$\mathrm{E}(\overline{X}) = \mu$$

である。

$$\mathrm{V}(\overline{X}) = \frac{\sigma^2}{n}$$

となる。$n \to \infty$ のとき，

$$\mathrm{V}(\overline{X}) = \frac{\sigma^2}{n} \longrightarrow 0$$

となるので，$\overline{X}$ は $\mu$ の一致推定量であると言える。

## 5.7 $\chi^2$ 分布

$m$ 個の確率変数 $Z_1, Z_2, \cdots, Z_m$ は，互いに独立な標準正規分布に従うものとする。このとき，$Y = \sum_{i=1}^{m} Z_i^2$ は，自由度 $m$ の $\chi^2$ 分布に従う。

$Y \sim \chi^2(m)$，または，$Y \sim \chi_m^2$ と表記する。

$\chi^2$ (カイ二乗) 分布表から確率を求める。

$Y \sim \chi^2(m)$ のとき，$\mathrm{E}(Y) = m$，$\mathrm{V}(Y) = 2m$ となる。(証明略)

1. 2 つの独立な $\chi^2$ 分布からの確率変数 $X, Y$ を考える。$X \sim \chi^2(n)$，$Y \sim \chi^2(m)$ とする。このとき，$Z = X + Y \sim \chi^2(n+m)$ となる。(証明略)

2. $n$ 個の独立な確率変数 $X_1, X_2, \cdots, X_n$ が同一の正規分布 $N(\mu, \sigma^2)$ に従うものとする。

3. $\dfrac{X_i - \mu}{\sigma} \sim N(0,1)$ なので，$\left(\dfrac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1)$ となる。

$\dfrac{X_1-\mu}{\sigma}, \dfrac{X_2-\mu}{\sigma}, \cdots, \dfrac{X_n-\mu}{\sigma}$ はそれぞれ独立なので,

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

となる。

4. $\mu$ を $\overline{X}$ に置き換えると,

$$\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

となる。(証明は後述)

さらに,

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

を定義すると,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

となる。$S^2$ は $\sigma^2$ の不偏推定量である (後述)。

5. すなわち,

$$\mathrm{E}\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \qquad \mathrm{V}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1),$$

となる。

## 5.8 $t$ 分布

**正規分布の重要な定理：** $n$ 個の独立な確率変数 $X_1, X_2, \cdots, X_n$ が同一の正規分布 $N(\mu, \sigma^2)$ に従うものとする。このとき,

$$\sum_{i=1}^{n} c_i X_i \sim N(\mu \sum_{i=1}^{n} c_i, \sigma^2 \sum_{i=1}^{n} c_i^2)$$

となる。ただし，$c_1, c_2, \cdots, c_n$ は定数とする。

**$t$ 分布：**　$Z$ を標準正規分布，$Y$ を自由度 $m$ の $\chi^2$ 分布に従い，両者は独立な確率変数とする。このとき，$U = \dfrac{Z}{\sqrt{Y/m}}$ は，自由度 $m$ の $t$ 分布に従う。

$U \sim t(m)$，または，$U \sim t_m$ と表記する。

$U \sim t(m)$ のとき，$m > 1$ について $\mathrm{E}(U) = 0$，$m > 2$ について $\mathrm{V}(U) = \dfrac{m}{m-2}$ となる。(証明略)

$t$ 分布表から確率を求める。(表 10.1.3 を見よ)

1. ゼロを中心に左右対称。($\mathrm{E}(U) = 0$)
2. $t$ 分布は，標準正規分布より裾野の広い分布 (なぜなら，$\mathrm{V}(U) = \dfrac{m}{m-2} > 1$)
3. $m \longrightarrow \infty$ のとき，$t(m) \longrightarrow N(0,1)$ となる。(期待値は $m > 1$ について $\mathrm{E}(U) = 0$，分散は $\mathrm{V}(U) = \dfrac{m}{m-2} \longrightarrow 1$)

## 5.9　標本平均 $\overline{X}$ の分布

$X_1, X_2, \cdots, X_n$ の $n$ 個の確率変数は，互いに独立で，平均 $\mu$，分散 $\sigma^2$ の正規分布に従うものとする。

1. $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})$ なので，$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ となる。
2. $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$ である。(証明は略)
3. $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ と $\dfrac{(n-1)S^2}{\sigma^2}$ は独立。(証明は略)

   すなわち，$\overline{X}$ と $S^2$ は独立。

4. したがって，

$$\frac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/ n-1}} = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

を得る。

重要な結果は，

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

ただし，$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ である。

$\sigma^2$ を $S^2$ に置き換えると，正規分布から $t$ 分布になる。

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \Longrightarrow \quad \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

# 5.10　区間推定 (信頼区間)

$\overline{X}$ の分布を利用して，$\mu$ の信頼区間を求める。

1. $\overline{X}$ の分布は以下の通り。

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

となる。

2. $t_{\alpha/2}(n-1)$，$t_{1-\alpha/2}(n-1)$ を自由度 $n-1$ の $t$ 分布の上から $100\times\frac{\alpha}{2}$ % 点，$100\times(1-\frac{\alpha}{2})$ % 点の値とする。このとき，

$$\mathrm{Prob}\Big(t_{1-\alpha/2}(n-1) < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\Big) = 1-\alpha$$

となる。ただし，自由度と $\alpha$ が決まれば，$t_{\alpha/2}(n-1)$，$t_{1-\alpha/2}(n-1)$ は $t$ 分布表から得られる。

3. $t$ 分布は左右対称なので，

$$t_{1-\alpha/2}(n-1) = -t_{\alpha/2}(n-1) \qquad t_{\alpha/2}(n-1) = |t_{1-\alpha/2}(n-1)|$$
$$t_{1-\alpha/2}(n-1) = -|t_{\alpha/2}(n-1)|$$

となる。

4. 書き直して，

$$\mathrm{Prob}\left(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

となる。

5. $\mu$ が区間 $(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}})$ にある確率は $1-\alpha$ である。

6. 推定量 $\overline{X}$，$S^2$ をその推定値 $\overline{x}$，$s^2$ で置き換える。ただし，$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$，$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ とする。

7. 区間 $(\overline{x} - t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}})$ を信頼係数 $1-\alpha$ の信頼区間といい，$\overline{x} - t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}}$ を信頼下限，$\overline{x} + t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}}$ を信頼上限と呼ぶ。

## 5.11 仮説検定

$\overline{X}$ の分布を利用して，$\mu$ の仮説検定を行う。

1. 帰無仮説 $H_0$： $\mu = \mu_0$ 対立仮説 $H_1$： $\mu \neq \mu_0$
2. 帰無仮説 $H_0$： $\mu = \mu_0$ が正しいもとでの分布は,

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

となる。

3. $\mathrm{Prob}\left(t_{1-\alpha/2}(n-1) < \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\right) = 1-\alpha$

$t_{\alpha/2}(n-1)$，$t_{1-\alpha/2}(n-1)$ をそれぞれ自由度 $n-1$ の $t$ 分布の上から $100 \times \dfrac{\alpha}{2}$ % 点，$100 \times \dfrac{1-\alpha}{2}$ % 点の値とする。

自由度と $\alpha$ が決まれば，$t_{\alpha/2}(n-1)$，$t_{1-\alpha/2}(n-1)$ は $t$ 分布表から得られる。

4. $\alpha$ を有意水準と呼ぶ。慣習的に $\alpha = 0.01, 0.05$ が使われる。

5. $-t_{\alpha/2}(n-1) > \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$，または，$\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_{\alpha/2}(n-1)$ ならば，帰無仮説 $H_0$：$\mu = \mu_0$ は，分布の端にあり，起こりにくいと考える。

   $\Longrightarrow$ 有意水準 $\alpha$ で帰無仮説 $H_0$：$\mu = \mu_0$ を棄却する。

6. 実際の検定手続：

   (a) $\overline{X}, S^2$ を実績値で置き換えて，

$$\frac{\overline{x}-\mu_0}{s/\sqrt{n}}$$

を得る。ただし，$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$，$s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2$ とする。

(b) $-t_{\alpha/2}(n-1) > \dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}$，または，$\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}} > t_{\alpha/2}(n-1)$ ならば，有意水準 $\alpha$ で帰無仮説 $H_0$：$\mu = \mu_0$ を棄却する。

# 第6章　統計学の回帰分析への応用

## 6.1　確率的モデル：単回帰モデル

再び，話を簡単にするために単回帰モデルを考えることにしよう。すなわち，$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ のように $n$ 組のデータがあり，$X_i$ と $Y_i$ との間に線型関係を想定する。

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

最小二乗法を用いて，データに直線のあてはめを行った。その結果，$\hat{\alpha}$，$\hat{\beta}$，$\hat{Y}_i$ を求めるための公式は,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2},$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X},$$

であった。

$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ とするとき，$Y_i$，$\hat{Y}_i$，$\hat{u}_i$，$\hat{\alpha}$，$\hat{\beta}$ の関係は以下の通りである。

$$Y_i = \hat{Y}_i + \hat{u}_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i,$$

残差 $\hat{u}_i$ が必ず含まれることから，回帰モデルを

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

として誤差項 (または，攪乱項) $u_i$ を含め，それを確率変数として考える。$u_i$ は平均 0，分散 $\sigma^2$ の正規分布が仮定されることが多い。ある確率密度分布（ここでは正規分布）があって，その分布に従い，データ（ここでは $Y_i$）が生成されるモデルのことを確率的モデルと呼ぶ。

$Y_i$： 被説明変数，従属変数

$X_i$： 説明変数，独立変数

$\alpha$，$\beta$： 未知母数 (未知パラメータ)

$\hat{\alpha}$，$\hat{\beta}$： 推定量 (特に，最小二乗推定量)，時には，推定値（最小二乗推定値）

1. 残差 $\hat{u}_i$ は $u_i$ の実現値としてみなすことができる。
2. $\hat{\alpha}$，$\hat{\beta}$ の性質を統計学的に考察可能となる。

## 6.2　回帰モデルの仮定

回帰モデル

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

の仮定：

1. $X_i$ は確率変数でないと仮定する (固定された値)。
2. すべての $i$ について，$\mathrm{E}(u_i) = 0$ とする。
3. すべての $i$ について，$\mathrm{V}(u_i) = \sigma^2$ とする。($\mathrm{V}(u_i) = \mathrm{E}(u_i^2) = \sigma^2$ に注意)
4. すべての $i \neq j$ について，$\mathrm{Cov}(u_i, u_j) = 0$ とする。($\mathrm{Cov}(u_i, u_j) = \mathrm{E}(u_i u_j) = 0$ に注意)
5. すべての $i$ について，$u_i \sim N(0, \sigma^2)$ とする（$u_i$ は正規分布を仮定する）。
6. $n \longrightarrow \infty$ のとき，$\sum_{i=1}^n (X_i - \overline{X})^2 \longrightarrow \infty$ とする。

攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に平均ゼロ，分散 $\sigma^2$ の正規分布する。

再度，まとめて，回帰モデル：

$$Y_i = \alpha + \beta X_i + u_i, \quad u_i \sim N(0, \sigma^2),$$

ただし，

$Y_i$： 被説明変数，従属変数

$X_i$： 説明変数，独立変数

$\alpha$，$\beta$，$\sigma^2$： 未知母数 (未知パラメータ)

$\hat{\alpha}$，$\hat{\beta}$： 推定量 (特に，最小二乗推定量)

特に，回帰直線は,

$$\mathrm{E}(Y_i) = \alpha + \beta X_i$$

として解釈される。

### 6.2.1 誤差項 (攪乱項) の経済学的意味

1. 経済理論自身が不完全： $X$ 以外にも他の説明変数が必要であるにもかかわらず，それを誤って除いている可能性がある。

2. モデルの定式化が不完全：$Y$ と $X$ との間の線形関係が誤りかもしれない。

3. 理論モデルとデータとの対応：理論モデルで考えられる変数と実際に用いたデータが適当でないかもしれない。例：所得のデータについては国民総生産，国民所得，可処分所得，労働所得 ⋯，金利では公定歩合，国債利回り，定期預金金利，全国銀行平均約定金利 ⋯

4. 測定上の誤差：経済データは一般的に推計されているため完全ではない。誤差を含む。

## 6.3　$\hat{\alpha}$，$\hat{\beta}$ の統計的性質

**準備：**　$\omega_i = \dfrac{(X_i - \overline{X})}{\sum_{j=1}^{n}(X_j - \overline{X})^2}$ とする。

このとき，

$$\sum_{i=1}^{n} \omega_i = 0$$

となる（$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ から $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ となるので，$\sum_{i=1}^{n} \omega_i = 0$ が得られる）。

また,

$$\sum_{i=1}^{n} \omega_i X_i = 1$$

となる（$\sum_{i=1}^{n} \omega_i X_i = \sum_{i=1}^{n} \omega_i X_i - \sum_{i=1}^{n} \omega_i \overline{X} = \sum_{i=1}^{n} \omega_i (X_i - \overline{X}) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sum_{j=1}^{n}(X_j - \overline{X})^2} = 1$ となるので）。

### 6.3.1 $\hat{\beta}$ について

$\beta$ の最小二乗推定量 $\hat{\beta}$ は,

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ &= \sum_{i=1}^{n} \omega_i Y_i\end{aligned}$$

となる。ただし，$\omega_i = \dfrac{(X_i - \overline{X})}{\sum_{j=1}^n (X_j - \overline{X})^2}$ である。二つ目の等式ついて，添え字を変更しても等式は成り立つので，分母の添え字を $j$ に変更している。最後の等式では，$\sum_{i=1}^n \omega_i = 0$ を利用している。

さらに，

$$\hat{\beta} = \sum_{i=1}^n \omega_i Y_i = \sum_{i=1}^n \omega_i(\alpha + \beta X_i + u_i) = \alpha \sum_{i=1}^n \omega_i + \beta \sum_{i=1}^n \omega_i X_i + \sum_{i=1}^n \omega_i u_i$$
$$= \beta + \sum_{i=1}^n \omega_i u_i$$

となる。

最後の等式では，$\sum_{i=1}^n \omega_i = 0$，$\sum_{i=1}^n \omega_i X_i = 1$ を利用している。

### 6.3.2　$\hat{\alpha}$ について

$\alpha$ の最小二乗推定量 $\hat{\alpha}$ については，

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = \alpha - (\hat{\beta} - \beta)\overline{X} + \overline{u}$$

ただし，$\overline{u} = \dfrac{1}{n}\sum_{i=1}^{n} u_i$ である。$\overline{Y} = \alpha + \beta\overline{X} + \overline{u}$ を途中で使う。

### 6.3.3　$\hat{\alpha}$，$\hat{\beta}$ の平均

$\hat{\beta}$ は次のように書き換えられた。

$$\hat{\beta} = \beta + \sum_{i=1}^{n} \omega_i u_i,$$

の両辺に期待値をとると，

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + \sum_{i=1}^{n} \omega_i u_i) = \beta + \sum_{i=1}^{n} \mathrm{E}(\omega_i u_i) = \beta + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta,$$

となり，$\hat{\beta}$ は $\beta$ の不偏推定量であると言える。

$\hat{\alpha}$ については，

$$\hat{\alpha} = \alpha - (\hat{\beta} - \beta)\overline{X} + \overline{u}$$

を利用して，辺々に期待値をとると，

$$\mathrm{E}(\hat{\alpha}) = \alpha - \mathrm{E}(\hat{\beta} - \beta)\overline{X} + \mathrm{E}(\overline{u}) = \alpha$$

となる。$\mathrm{E}(\hat{\beta}) = \beta$ なので，$\mathrm{E}(\hat{\beta} - \beta) = 0$ に注意。また，$\mathrm{E}(\overline{u})$ の計算は以下のとおり。

$$\mathrm{E}(\overline{u}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} u_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(u_i) = 0$$

$\hat{\alpha}$ は $\alpha$ の不偏推定量であると言える。

### $\hat{\alpha}$，$\hat{\beta}$ の分散

$\hat{\beta}$ の分散について，$\hat{\beta} = \beta + \sum_{i=1}^{n} \omega_i u_i$ を用いると，

$$\begin{aligned}\mathrm{V}(\hat{\beta}) &= \mathrm{V}(\beta + \sum_{i=1}^{n}\omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n}\omega_i u_i) = \sum_{i=1}^{n}\omega_i^2\mathrm{V}(u_i^2) = \sigma^2\sum_{i=1}^{n}\omega_i^2 \\ &= \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\end{aligned}$$