第 4 章　単回帰モデルによる分析

第２章：単回帰の記述的な説明
この章：単回帰の推測統計的な議論

1　古典的仮定

単回帰モデル　$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \cdots, n)$　(1)

α と β：それぞれ切片と傾きを示すパラメータ
ε_i ：**誤差項**（確率変数とみなされる）で，$E(\varepsilon_i) = 0$ とする．

(a) 説明変数の非確率性

統計的推測の理論を容易にするためであり，非現実的な場合がある．例えば，所得を所与として，その所得に応じた世帯の消費支出を計測する，ということは実際上不可能であろう．

(b) 誤差項の分散均一性と無相関性

$V(\varepsilon_i) = \sigma^2$ ：分散均一
$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j)$ ：無相関性

(c) 誤差項の正規性

各 ε_i は x_i 以外のさまざまな要因から成り立っていると考えられるので，この仮定は追加的な条件のもとで**中心極限定理**により正当化することが可能である．

単回帰モデル
$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \cdots, n)$$
においては，古典的仮定のもとで，
(i) $y_1, \cdots, y_n$ は互いに独立である．
(ii) 各 y_i は，平均 $\alpha + \beta x_i$，分散 σ^2 の正規分布に従う．すなわち，
$$y_i \sim N(\alpha + \beta x_i, \sigma^2) \tag{2}$$

2　最小 2 乗推定量と最尤推定量

LSE（Least Squares Estimator：最小 2 乗推定量）
サイズ n の標本にもとづく単回帰モデル (1) において，傾き β の LSE $\hat{\beta}$ は，

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} c_i y_i = \beta + \sum_{i=1}^{n} c_i \varepsilon_i \tag{3}$$

ここで，

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - x)^2} \tag{4}$$

以上のことから，$\hat{\beta}$ も正規分布に従うことになる．また，$E(\hat{\beta}) = \beta$ となるから，$\hat{\beta}$ は**不偏**である．

他方，分散は

$$V(\hat{\beta}) = \sigma^2 \sum_{i=1}^{n} c_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5}$$

分散の平方根，すなわち標準偏差は，推定の観点からは**標準誤差** と呼ばれる．標準誤差は，推定量とパラメータの真の値とのずれの程度を表すものである．$\hat{\beta}$ がとる値は平均的には β であるが，$\sqrt{V(\hat{\beta})}$ 程度のずれが見込まれることになる．

図 4-1 には $\hat{\beta}$ の分布が例示されている．想定した単回帰モデルは，

$$y_i = 10 + 0.8 \times i + \varepsilon_i \qquad (i = 1, \cdots, 10) \tag{6}$$

であり（$\alpha = 10$，$\beta = 0.8$，$x_i = i$，$n = 10$），$\sigma^2 = 4$ としたものである．

図の曲線：平均 0，分散 4/82.5 の正規分布

一致性（Consistency）： **図 4-1** で標本サイズが大きくなる状況をイメージすれば，$\hat{\beta}$ の分布は β のまわりにより集中して行くことが予想される．このとき，$\hat{\beta}$ は β の**一致推定量**である，あるいは単に，$\hat{\beta}$ は**一致性**をもつという．一致推定量のより厳密な定義は，次のように与えられる．

一致推定量の定義

パラメータ θ の推定量 $\hat{\theta}$ が θ の一致推定量であるとは，標本サイズ n が大きくなるとき，両者の差が任意の正数 δ 以下となる確率が 1（＝両者の差が任意の正数 δ を越える確率が 0）に収束すること，すなわち，次式が成り立つことである．

$$\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| \le \delta\right) = 1 \quad \Leftrightarrow \quad \lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| > \delta\right) = 0$$

$\hat{\theta}$ が θ の一致推定量であることを，

$$\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$$

と表すのが普通である．$p\lim$ は確率極限（Probability Limit）の略である．$\hat{\theta}$ が一致性をもつための十分条件は，

（i） $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ （ii） $\lim_{n\to\infty} V(\hat{\theta}) = 0$

となることである．(i) が成り立つとき，$\hat{\theta}$ は θ の**漸近的不偏推定量**であるという．

LSE $\hat{\beta}$ が一致性をもつための十分条件：

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\infty \tag{7}$$

次に，切片 α の LSE $\hat{\alpha}$ の性質を考えよう．

$$\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=\sum_{i=1}^{n}\left(\frac{1}{n}-c_i\bar{x}\right)y_i=\alpha+\sum_{i=1}^{n}\left(\frac{1}{n}-c_i\bar{x}\right)\varepsilon_i \tag{8}$$

$\hat{\alpha}$ は正規分布に従う．その平均は $E(\hat{\alpha})=\alpha$ （不偏），分散は

$$V(\hat{\alpha})=\sigma^2\sum_{i=1}^{n}\left(\frac{1}{n}-c_i\bar{x}\right)^2=\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \tag{9}$$

$\hat{\alpha}$ は (7) の条件のもとで一致性をもつ．

$\hat{\alpha}$ と $\hat{\beta}$ の同時分布は 2 次元正規分布であり，共分散は

$$Cov(\hat{\alpha},\hat{\beta})=Cov\left(\sum_{i=1}^{n}\left(\frac{1}{n}-c_i\bar{x}\right)\varepsilon_i,\sum_{i=1}^{n}c_i\varepsilon_i\right)=-\frac{\sigma^2\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{10}$$

誤差項の分散 σ^2 の推定量 $\hat{\sigma}^2$ は，

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{\alpha}-\hat{\beta}x_i)^2=\frac{1}{n-2}\sum_{i=1}^{n}e_i^2 \tag{11}$$

ここで，$e_i=y_i-\hat{\alpha}-\hat{\beta}x_i$ は残差であり，誤差 $\varepsilon_i=y_i-\alpha-\beta x_i$ の推定量である．$\hat{\sigma}^2$ は残差 2 乗和を，n ではなく $n-2$ で割って定義することにより，$\hat{\sigma}^2$ の不偏性，すなわち，$E(\hat{\sigma}^2)=\sigma^2$ が得られる．

このとき，

$$t_{\hat{\beta}}=\frac{\hat{\beta}-\beta}{\hat{\sigma}\Big/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\frac{\hat{\beta}-\beta}{\sqrt{\hat{V}(\hat{\beta})}}\quad\sim\quad t(n-2) \tag{12}$$

ただし，$t(n-2)$ は自由度 $n-2$ の t 分布を表し，$\hat{V}(\hat{\beta})$ は，(5) の $V(\hat{\beta})$ に含まれる未知のパラメータ σ^2 を (11) で定義された推定量 $\hat{\sigma}^2$ でおきかえたものである．同様にして，

$$t_{\hat{\alpha}}=\frac{\hat{\alpha}-\alpha}{\hat{\sigma}\Big/\sqrt{\frac{1}{n}+\bar{x}^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\frac{\hat{\alpha}-\alpha}{\sqrt{\hat{V}(\hat{\alpha})}}\quad\sim\quad t(n-2) \tag{13}$$

MLE

(1) の単回帰モデルに対する尤度関数は，α，β，σ^2 の関数となるので $L(\alpha,\beta,\sigma^2)$ と表すことにする．それは，$y_1,\cdots,y_n$ の同時密度関数であるから，

$y_1,\cdots,y_n$ が互いに独立で，各 y_i が (2) にあるように $N(\alpha+\beta x_i,\sigma^2)$ に従うことを使うと，

$$L(\alpha,\beta,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left(y_i-\alpha-\beta x_i\right)^2\right\}$$
$$=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2\right\} \qquad (14)$$

のように表現される．

α，β，σ^2 の MLE を，それぞれ $\tilde{\alpha}$，$\tilde{\beta}$，$\tilde{\sigma}^2$ とすると，これらは (14) の尤度関数を最大にする値である．今の場合，尤度関数の最大化よりも，**対数尤度関数**

$$l(\alpha,\beta,\sigma^2)=\log L(\alpha,\beta,\sigma^2)=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \qquad (15)$$

の最大化を考える方が容易である．このとき，対数尤度関数の形から，$\tilde{\alpha}$ と $\tilde{\beta}$ は，それぞれ LSE $\hat{\alpha}$ と $\hat{\beta}$ に一致することが了解されよう．

他方，$\tilde{\sigma}^2$ を求めるには，

$$\frac{\partial l(\alpha,\beta,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(y_i-\alpha-\beta x_i\right)^2=0$$

を σ^2 に関して解き，α と β には MLE を代入すればよい．したがって，

$$\tilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\tilde{\alpha}-\tilde{\beta}x_i\right)^2=\frac{n-2}{n}\hat{\sigma}^2$$

を得る．$\tilde{\sigma}^2$ は (11) で定義した不偏推定量 $\hat{\sigma}^2$ よりも小さめであるから，期待値は σ^2 よりも小さくなり，下方へのバイアスをもつ推定量となる．しかし，漸近的不偏である．

以上，古典的仮定のもとでは，α と β の LSE と MLE は一致し，σ^2 の LSE と MLE も漸近的には同等であることをみた．しかし，古典的仮定をゆるめた一般のモデルにおいては，両者は本質的に異なり，漸近的には MLE の方が優れていることがわかっている．これらの最適性については，さらに，本章の 6 節で考えることにする．

3　信頼区間

傾き β に関する信頼区間の作り方

(i) 指定された**信頼係数** 1 - γ に対して，確率表現

$$P\left(\frac{\left|\hat{\beta}-\beta\right|}{\sqrt{\hat{V}(\hat{\beta})}}\le t_{\gamma/2}^{(n-2)}\right)=1-\gamma$$

をみたす $t(n-2)$ 分布の上側 γ/2 点 $t_{\gamma/2}^{(n-2)}$ を t 分布表から求める．

(ii) 確率表現のカッコ内の不等式を β に関して次のように解く．

$$\hat{\beta}-t_{\gamma/2}^{(n-2)}\times\sqrt{\hat{V}(\hat{\beta})}\le\beta\le\hat{\beta}+t_{\gamma/2}^{(n-2)}\times\sqrt{\hat{V}(\hat{\beta})} \qquad (16)$$

〔**例題 4.1**〕次のデータは単回帰モデル（1）から得られたものである．このとき，信頼係数 95% で α と β の信頼区間を求めよ．

x	1	2	3	4	5	6	7	8	9	10
y	10.6	10.9	14.6	8.2	16.7	16.1	14.6	14.3	19.3	17.4

（解）まず，$n=10$，$\gamma=0.05$ であるから，$t_{\gamma/2}^{(n-2)}=2.306$ となる．次に，α と β の LSE を計算すると，$\hat{\alpha}=9.7$，$\hat{\beta}=0.83$ となる．また，誤差項の分散 σ^2 の推定値として，$\hat{\sigma}^2=6.16$ を得る．このとき，

$$\hat{V}(\hat{\alpha})=\hat{\sigma}^2\left(1/n+\bar{x}^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)=6.16\times\left(1/10+5.5^2/82.5\right)=2.875$$

$$\hat{V}(\hat{\beta})=\hat{\sigma}^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2=6.16/82.5=0.0747$$

となるから，標準誤差は，それぞれ 1.696，0.273 である．以上より，α の 95% 信頼区間として，

[9.7 - 2.306×1.696，9.7 + 2.306×1.696] = [5.79，13.61]

を得る．また，β の 95% 信頼区間は

[0.83 - 2.306×0.273，0.83 + 2.306×0.273] = [0.20，1.46]

となる．

4　t 検定

単回帰モデルのパラメータに関する仮説，例えば，傾き β がある値 β_0 とみなすことができるかどうかは，検定問題

$$H_0\colon \beta=\beta_0 \qquad \text{vs} \qquad H_1\colon \beta\neq\beta_0 \tag{17}$$

として定式化することができる．検定統計量として

$$t_{\hat{\beta}}=\frac{\hat{\beta}-\beta_0}{\hat{\sigma}\Big/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\frac{\hat{\beta}-\beta_0}{\sqrt{\hat{V}(\hat{\beta})}} \tag{18}$$

を使うことができる．H_0 のもとでの分布（帰無分布）は自由度 $n-2$ の t 分布である．このことから，(18) の統計量を使った検定は **t 検定**と呼ばれる．検定の手続きは以下の通りである．

検定問題 (17) に対する t 検定の手続き

（i）指定した**有意水準** γ に対して，自由度 $n-2$ の t 分布の上側 γ/2 点 $t_{\gamma/2}^{(n-2)}$ を t 分布表から求める．

（ii）実際のデータから，(18) の統計量 $t_{\hat{\beta}}$ を計算して，

$\left|t_{\hat{\beta}}\right|>t_{\gamma/2}^{(n-2)}$ ならば H_0 を棄却し，

$\left|t_{\hat{\beta}}\right|\leq t_{\gamma/2}^{(n-2)}$ ならば H_0 を受容する．

両側検定の受容域

$$\left|t_{\hat\beta}\right| \le t_{\gamma/2}^{(n-2)} \quad \Leftrightarrow \quad \hat\beta - t_{\gamma/2}^{(n-2)}\sqrt{\hat V(\hat\beta)} \le \beta_0 \le \hat\beta + t_{\gamma/2}^{(n-2)}\sqrt{\hat V(\hat\beta)}$$

⇔ 信頼係数 1 - γ の β に関する信頼区間に β_0 が含まれる．

片側検定の場合

(a) $H_1{:}\beta < \beta_0$ のとき

$t_{\hat\beta} < -t_{\gamma}^{(n-2)}$ ならば H_0 を棄却し，さもなければ受容する．

(b) $H_1{:}\beta > \beta_0$ のとき

$t_{\hat\beta} > t_{\gamma}^{(n-2)}$ ならば H_0 を棄却し，さもなければ受容する．

〔**例題 4.2**〕例題 4.1 で扱った単回帰モデル (1) からのデータに基づいて，
有意水準 10% および 5% で次の検定を行え．
(i) $H_0{:}\alpha = 13$ vs $H_1{:}\alpha \ne 13$
(ii) $H_0{:}\beta = 1$ vs $H_1{:}\beta \ne 1$

（解）例題 4.1 の結果から，検定統計量の値は，それぞれ

$$t_{\hat\alpha} = (\hat\alpha - \alpha_0)\Big/\sqrt{\hat V(\hat\alpha)} = (9.7-13)/1.696 = -1.95$$

$$t_{\hat\beta} = (\hat\beta - \beta_0)\Big/\sqrt{\hat V(\hat\beta)} = (0.83-1)/0.273 = -0.62$$

となる．有意点は，有意水準が 10% の場合は，$n=10$，$\gamma = 0.1$ として $t_{\gamma/2}^{(n-2)} = 1.86$ となり，5% の場合は，$n=10$，$\gamma = 0.05$ として $t_{\gamma/2}^{(n-2)} = 2.306$ となる．したがって，(i) は有意水準 10% で棄却，5% で受容，(ii) はいずれの有意水準でも受容となる．有意水準 5% の場合には，例題 4.1 で得られた信頼係数 95% の信頼区間の中に $\alpha = 13$ も $\beta = 1$ も含まれていることから，受容されることは明らかである．

t 値：(18) の t 検定統計量において，特に $\beta_0 = 0$ としたときの値
（β の推定量 $\hat\beta$ を標準誤差で割った値）を $\hat\beta$ の **t 値**という．
t 値の絶対値が大きい場合は，β の値が 0 と有意に離れていると解釈できるので，β を係数にもつ説明変数 x は説明力があるものとみなされる．

回帰の結果を報告する場合には，種々の統計値に加えて t 値も推定値の下にカッコ付きで記すのが普通である．例えば，例題の場合には，

$$\underset{(5.72)}{y = 9.7} + \underset{(3.04)}{0.83}\,x \qquad R^2 = 0.536,\ \overline{R}^2 = 0.478,\ \hat\sigma = 2.482$$

のようにする．ここで，R^2 は決定係数，$\overline{R}^2$ は自由度修正済決定係数，$\hat\sigma$ は誤差項の標準偏差の推定値である．なお，t 値の代わりに標準誤差が報告される場合もある．その場合は，t 値は推定値を標準誤差で割れば求められる．

５ 予測

予測：確率変数の実現値に関して推論すること．**点予測**と**区間予測**がある．

点予測：

説明変数の値が x_0 のとき，対応する被説明変数 y_0 の実現値を一点で推論することを**点予測**という．点予測するための統計量を**予測量**と呼ぶ．(1) の単回帰モデルにおいては，$y_0=\alpha+\beta x_0+\varepsilon_0$ の自然な予測量は，

$$\hat{y}_0=\hat{\alpha}+\hat{\beta}x_0 \tag{19}$$

で与えられる．ここで，$\hat{\alpha}$ と $\hat{\beta}$ はサイズ n の標本から得られる α と β の LSE である．予測する y_0 は，この標本とは独立とする．

予測誤差： $y_0-\hat{y}_0=\alpha+\beta x_0+\varepsilon_0-(\hat{\alpha}+\hat{\beta}x_0)=\varepsilon_0-(\hat{\alpha}-\alpha)-(\hat{\beta}-\beta)x_0$

不偏予測： 予測誤差の期待値＝0 となる予測量…(19) は不偏である．

予測量の分布：平均 0 の正規分布で，分散は ε_0 が $\hat{\alpha}$ および $\hat{\beta}$ と独立であることとなどを使って

$$\begin{aligned}V(y_0-\hat{y}_0)&=V(\varepsilon_0)+V(\hat{\alpha})+x_0^2V(\hat{\beta})+2x_0Cov(\hat{\alpha},\hat{\beta})\\&=\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)\end{aligned} \tag{20}$$

・$x_0=\overline{x}$ のときに最小　・$\overline{x}$ から離れるにしたがって大きくなる

・決して 0 にはならない．

→ 説明変数がとりうる値からはずれた所での予測は精度が悪くなることを示唆

→ 標本サイズが大きくなるとき σ^2 に収束し，決して 0 にはならない．

（かりにデータが無数に存在しても完全な予測はありえない）

この点が予測と推定の本質的な違いである．

区間予測

確率変数の実現値を区間の形で推論することを**区間予測**という．

予測の信頼区間の構成：予測誤差の次の性質

$$\frac{y_0-\hat{y}_0}{\sqrt{\hat{V}(y_0-\hat{y}_0)}}\quad\sim\quad t(n-2) \tag{21}$$

が成り立つ事実を利用する．ここで，$\hat{V}(y_0-\hat{y}_0)$ は (18) の $V(y_0-\hat{y}_0)$ の表現において，σ^2 を不偏推定量 $\hat{\sigma}^2$ でおきかえたものである．

このとき，

$$P\left(\frac{|y_0-\hat{y}_0|}{\sqrt{\hat{V}(y_0-\hat{y}_0)}}\le t_{\gamma/2}^{(n-2)}\right)=1-\gamma$$

が成立するので，確率表現の中の不等式を y_0 に関して解いて

$$\hat{y}_0 - t_{\gamma/2}^{(n-2)} \times \sqrt{\hat{V}(y_0 - \hat{y}_0)} \le y_0 \le \hat{y}_0 - t_{\gamma/2}^{(n-2)} \times \sqrt{\hat{V}(y_0 - \hat{y}_0)} \quad (22)$$

を得る．この区間の下限と上限に実際のデータから計算される値を代入することにより，信頼係数 1 - γ の y_0 に関する信頼区間が構成される．

〔**例題** 4.2〕例題 4.1 で扱った単回帰モデル（1）からのデータに基づいて，$1 \le x_0 \le 10$ の範囲にある各 x_0 に対応する y_0 の点予測と，信頼係数 95% の区間予測を行え．

（解）点予測は，例題 4.1 の結果を使って，$\hat{y}_0 = 9.7 + 0.83x_0$ となる．また，信頼係数95% の信頼区間は，（21）式に $\hat{y}_0 = 9.7 + 0.83x_0$，$t_{\gamma/2}^{(n-2)} = 2.306$ （$n = 10$，$\gamma = 0.05$），および

$$\hat{V}(y_0 - \hat{y}_0) = 6.16 \times \left(1 + 1/10 + (x_0 - 5.5)^2 / 82.5\right)$$

を代入することにより求められる．**図 4-2** には，例題 4.1 のデータ（＋）と，各 x_0 に対応する y_0 の予測値すなわち推定回帰直線，および信頼区間の下限と上限（破線）が示されている．下限，上限それぞれの線は双曲線である．

6　ガウス・マルコフの定理

古典的仮定のもとで，α と β の LSE（= MLE）は不偏推定量の中で最小の分散を達成しているという意味で，**最小分散不偏推定量**であることを示すことができる．

以下では，古典的仮定の中で正規性の仮定だけをはずした場合を考えることにしよう．その場合には，MLE は必ずしも適用できないが，LSE は定義可能である．このとき，β の LSE について，正規性を仮定した場合よりも弱めの次のような最適性をもつことが示される．

〔**例題** 4.3〕単回帰モデル（1）において，説明変数が非確率的，誤差項が無相関で均一分散をもつとき，β の LSE $\hat{\beta}$ は，$y_1, \cdots, y_n$ の一次結合で表される不偏推定量の中で最小の分散をもつことを示せ．

（解）$y_1, \cdots, y_n$ の一次結合となっている β の任意の推定量 β^* は，

$$\beta^* = \sum_{i=1}^{n} d_i y_i$$

と表すことができる．ここで，$d_1, \cdots, d_n$ は定数であり，β^* が不偏であることから，

$$E(\beta^*) = \sum_{i=1}^{n} d_i E(y_i) = \alpha \sum_{i=1}^{n} d_i + \beta \sum_{i=1}^{n} d_i x_i = \beta$$

が常に成立する．このことから，$d_1, \cdots, d_n$ は

$$\sum_{i=1}^{n} d_i = 0, \qquad \sum_{i=1}^{n} d_i x_i = 1$$

をみたさなければならない．このとき，（4）で定義された LSE の ウェイト $c_1, \cdots, c_n$ との間には，

$$\sum_{i=1}^{n} c_i d_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})d_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} c_i^2$$

の関係がある．したがって，

$$V(\beta^*) = \sigma^2 \sum_{i=1}^{n} d_i^2 = \sigma^2 \sum_{i=1}^{n} \left\{(d_i - c_i) + c_i\right\}^2$$

$$= \sigma^2 \sum_{i=1}^{n} (d_i - c_i)^2 + \sigma^2 \sum_{i=1}^{n} c_i^2 \geq \sigma^2 \sum_{i=1}^{n} c_i^2 = V(\hat{\beta})$$

となることがわかり，題意は示されたことになる．

ガウス・マルコフの定理：$y_1, \cdots, y_n$ の一次結合で表される不偏推定量の中で分散が最小となる特性．このとき，その推定量は**最良線形不偏推定量**（BLUE：Best Linear Unbiased Estimator）と呼ばれる．α と β の LSE は BLUE である．

予測量に関してもガウス・マルコフの定理を考えることができる．

〔**例題 4.4**〕単回帰モデル（1）において，説明変数が非確率的，誤差項が無相関で均一分散をもつとき，（19）で定義された予測量は，$y_1, \cdots, y_n$ の一次結合で表される不偏予測量の中で予測誤差分散が最小となることを示せ．

（解）$y_1, \cdots, y_n$ の一次結合となっている任意の予測量 y_0^* を

$$y_0^* = \sum_{i=1}^{n} q_i y_i$$

と表すことにする．特に，（19）の予測量 $\hat{y}_0$ は

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0 = \sum_{i=1}^{n} \left(\frac{1}{n} + (x_i - \bar{x})(x_0 - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) y_i = \sum_{i=1}^{n} p_i y_i$$

と表すことができる．ここで，

$$\sum_{i=1}^{n} p_i^2 = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - x)^2}$$

が成り立つ．ところで，不偏性から，

$$E(y_0 - y_o^*) = E\left(\alpha + \beta x_0 + \varepsilon_0 - \sum_{i=1}^{n} q_i y_i \right)$$

$$= \alpha \left(1 - \sum_{i=1}^{n} q_i \right) + \beta \left(x_0 - \sum_{i=1}^{n} q_i x_i \right) = 0$$

が常に成立しなければならないから，

$$\sum_{i=1}^{n} q_i = 1, \qquad \sum_{i=1}^{n} q_i x_i = x_0$$

となる．このとき，

$$\sum_{i=1}^{n} p_i q_i = \sum_{i=1}^{n} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) q_i = \sum_{i=1}^{n} p_i^2$$

を得る．したがって，

$$\begin{aligned} V(y_0 - y_o^*) &= V\left(\alpha + \beta x_0 + \varepsilon_0 - \sum_{i=1}^{n} q_i y_i \right) \\ &= \sigma^2 \left(1 + \sum_{i=1}^{n} q_i^2 \right) = \sigma^2 \left(1 + \sum_{i=1}^{n} \left((q_i - p_i) + p_i \right)^2 \right) \\ &= \sigma^2 \left(1 + \sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} \left(q_i - p_i \right)^2 \right) \end{aligned}$$

となり，この最小値は $q_i = p_i$（$i = 1, \cdots, n$）のとき，すなわち，$y_0^* = \hat{y}_0$ のときに達成されることがわかる．よって，証明された．

この例題で示された特性を，予測量に関するガウス・マルコフの定理という．このとき，(19) の予測量は**最良線形不偏予測量**（BLUP：Best Linear Unbiased Predictor）と呼ばれる．