第二章　単回帰（続き）[1]

劉慶豊[2]

小樽商科大学

November 20, 2009

[1]第二章からの資料は森棟公夫先生著「基礎コース　計量経済学」をもとに作成したものである。

[2]E-mail:qliu@res.otaru-uc.ac.jp, URL:http://www.otaru-uc.ac.jp/~qliu/

検定の根拠

一定の条件のもとでは

$$z = \left(\widehat{\beta} - \beta\right) / \sqrt{V(\widehat{\beta})} = \frac{\widehat{\beta} - \beta}{\sqrt{\sigma^2 / \{(n-1)s_{xx}\}}} \tag{1}$$

が標準正規分布に従う。

$$V(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1,n}(x_i - \overline{x})^2} = \frac{\sigma^2}{(n-1)s_{xx}} \tag{2}$$

$V(\widehat{\beta})$ を計算するために σ^2 の値が必要、しかし、σ^2 は母集団に関するパラメーターで未知である代わりに以前説明した残差分散 $s^2 = \frac{1}{n-1}\sum_{i=1}^{2} \hat{u}^2$ を利用する。そうすることで $V(\hat{\beta})$ が $s_{\hat{\beta}}^2$ となる。

s_β は $\hat{\beta}$ の標準誤差

$$s_\beta = \sqrt{s^2/\{(n-1)s_{xx}\}}$$

t 値

$$t_\beta = \frac{\widehat{\beta} - \beta}{s_\beta}$$

t_β は自由度 $n-2$ の t 分布に従う。

回帰係数に関するt検定

真のβがβ_0であるかどうかを調べる。ただし、β_0が定数である。β_0を0にする場合が多い。
片側検定$H_0 : \beta = \beta_0, H_1 : \beta > \beta_0$と$H_0 : \beta = \beta_0, H_1 : \beta < \beta_0$二種類。
両側検定$H_0 : \beta = \beta_0, H_1 : \beta \neq \beta_0$。

$$t_\beta = \frac{\widehat{\beta} - \beta_0}{\sqrt{s^2 / \{(n-1)s_{xx}\}}} \tag{3}$$

帰無仮説のもとでt_βは自由度$n-2$のt分布に従う。t_βと自由度$n-2$のt分布の有意水準点と比較する。

区間推定

信頼区間 $P\{c_1 \leq \beta \leq c_2\} = 0.95$ となるような c_1 と c_2 を求め、β が 95% の確率で入る空間が 95% 信頼区間。

信頼空間の求め方 95% の信頼空間を例に。有意水準 95% の t 値を t_β とする。$P\{t_{2.5} \leq t_\beta \leq t_{97.5}\} = 0.95$, ただし、$t_{2.5}$ と $t_{97.5}$ はそれぞれ t 分布の 2.5% と 97.5% 分位点である。

$$P\{t_{2.5} \leq t_\beta \leq t_{97.5}\} = 0.95$$
$$P\{-t_{97.5} \leq t_\beta \leq t_{97.5}\} = 0.95$$
$$P\{-t_{97.5} \leq \frac{\widehat{\beta} - \beta}{s_\beta} \leq t_{97.5}\} = 0.95$$

$$P\left\{\{\widehat{\beta} - t_{97.5} \times s_\beta \leq \beta \leq \widehat{\beta} + t_{97.5} \times s_\beta\right\} = 0.95$$

95% の信頼区間は $\left(\widehat{\beta} - t_{97.5} \times s_\beta, \widehat{\beta} + t_{97.5} \times s_\beta\right)$ である。

信頼区間と検定

帰無仮説の β_0 が信頼区間に入れば、帰無仮説は棄却できない、入らなければ、棄却する。

例題

例2.8 **車のスピードと停止距離の例の結果の続き。スピードが停止距離に影響を与えないかどうかを調べる。**

- $H_0 : \beta = 0, H_1 : \beta \geq 0$。**有意水準を**5%**とする。（　）の中は** t **値である（慣例として帰無仮説が係数** $= 0$ **のときの** t **統計量を示す）。**

$$\widehat{y} = -1.8294(-0.527) + 0.1643(2.881)x$$

$$t_\beta = \frac{(\hat{\beta} - \beta_0)}{s_\beta} = \frac{(0.1643 - 0)}{s_\beta} = 2.881$$

- **自由度** $n - 2 = 7 - 2 = 5$ **の** t **分布の右側**5%**有意水準点は**2.02。$t_\beta > 2.02$ **なので** H_0 **を棄却する。**

練習問題 **車のスピードと停止距離の例に関して、テキストの表** 2 − 2, 2 − 3 **の結果を用いて、**α, β **の推定量、**s^2, s_{xx} **を求めなさい。そして** t **値を計算しなさい。**

例題

例2.8の続き βの90%信頼区間を求める。

信頼空間は

$$\widehat{\beta} - t_{97.5} \times s_\beta \le \beta \le \widehat{\beta} + t_{97.5} \times s_\beta$$

- 自由度$n-2=7-2=5$のt分布の右側5%有意水準点（95%分位点）は2.02.
- $\hat{\beta}$の標準誤差$s_\beta = \sqrt{s^2/\{(n-1)s_{xx}\}} = 0.057$.
- 90%信頼区間は$(0.1643 - 2.02 \times 0.057, 0.1643 + 2.02 \times 0.057)$.

例題

ジップ（Ziph）の法則 都市人口 × 都市人口順位 = 定数

データ

表 2.4 京都12市の人口とその順位

順位	1	2	3	4	5	6	7	8	9	10	11	12
人口（千人）	1390	186	94	94	85	77	73	67	53	53	40	25

モデル c を定数として、

$$P_i \times i = c$$

$$\log(P_i) = c - \log(i)$$

$$\log(P_i) = c + \beta \log(i) + u_i, i = 1, 2, \cdots, n$$

推定結果　（　）の中は帰無仮説が $\beta = 0$ の t 値である。

$$\widehat{\log(P_i)} = 6.53\,(23.7) - 1.235\,(-8.1)\log(i)$$

検定 $H_0 : \beta = -1, H_1 : \beta \neq -1$（両側検定）。有意水準を 10% とする。帰無仮説が $\beta = 0$ の t 値の公式から逆算して標準誤差を求めて、そして t 値を求める。

$$s_\beta = -1.235 \times \frac{1}{-8.1} = 0.152$$

$$t = \frac{-1.235 - (-1)}{0.152} = -1.55.$$

検定結果 自由度 10 の右側 5% 有意水準点（95% 分位点）$t_{95} = 1.81$。$|t| < t_{95}$、H_0 は棄却できない。$\beta = -1$ であろうと認識する。

CAPM式の推定

株価の収益率

$$r = \frac{x_t - x_{t-1}}{x_{t-1}}$$

$$r = \log(\frac{x_t}{x_{t-1}})$$

市場ポートフォリオ 市場全体を一つのポートフォリオと見なす。その収益率が r_M.

安全資産の収益率 r_f

CAPM（資産価格決定モデル）

$$r - r_f = \beta(r_M - r_f)$$

CAPMの計量モデル

$$r - r_f = \alpha + \beta(r_M - r_f) + u$$

NTTに関するCAPM式の推定例

設定 日経平均の収益率を市場ポートフォリオの収益率とする（r_M）とする。郵便貯金の利息を r_f とする。$n = 239.$

推定結果 （ ）の中は t 値である。

$$\hat{r} - r_f = -0.00053(-0.44) + 0.874(10.8)(r_M - r_f)$$

α の検定 自由度が $n - 2 = 237, |-0.44| < t_{95}$ なので $\alpha = 0$ の帰無仮説が棄却できない。NTTの期待（平均）収益率は0に近いと認識する。

NTTに関するCAPM式の推定例

βの検定 帰無仮説が$\beta = 0$のt値の公式から逆算して標準誤差を求める

$$s_\beta = 0.874 \times \frac{1}{10.8} = 0.081$$

$H_0 : \beta = 1, H_1 : \beta < 1$の検定を有意水準5%で行う。$t$値が

$$t = \frac{0.874 - 1}{0.081} = -1.56$$

右側5%有意水準点が1.65となり、$|t| < 1.65$なので、$H_0 : \beta = 1$を棄却できない。NTTの安定さは日経平均と殆ど変わらないと結論を付ける。

最小２乗推定量の導出(1)

$\Phi = RSS = \sum_{i=1}^{n}\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)^2$

最小化の為の１次条件 $\frac{\partial\Phi}{\partial\widehat{\alpha}} = 0, \frac{\partial\Phi}{\partial\widehat{\beta}} = 0.$

$$\begin{aligned}\frac{\partial\Phi}{\partial\widehat{\alpha}} &= \sum_{i=1}^{n}\frac{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)^2}{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)}\frac{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)}{\partial\widehat{\alpha}} \\ &= \sum_{i=1}^{n}\left\{(-2)\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)\right\} = 0\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{\partial\Phi}{\partial\widehat{\beta}} &= \sum_{i=1}^{n}\frac{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)^2}{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)}\frac{\partial\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)}{\partial\widehat{\beta}} \\ &= \sum_{i=1}^{n}\left\{(-2)\left(y_i - \widehat{\alpha} - \widehat{\beta}x_i\right)x_i\right\} = 0\end{aligned} \tag{5}$$

最小２乗推定量の導出(2)

正規方程式

$$\sum_{i=1}^{n}\left(y_i-\widehat{\alpha}-\widehat{\beta}x_i\right)=0$$

$$\sum_{i=1}^{n}\left(y_i-\widehat{\alpha}-\widehat{\beta}x_i\right)x_i=0$$

最小化の為の２次条件

$$\frac{\partial^2\Phi}{(\partial\widehat{\beta})^2}=2\sum_{i=1}^{n}(x_i)^2$$

最小２乗推定量の導出(3)

正規方程式を解いて最小二乗（OLS）推定量を求める

$$\widehat{\beta} = \frac{s_{xy}}{s_{xx}} = r_{xy}\frac{s_y}{s_x} \tag{6}$$

$$= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{7}$$

$$= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{8}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{9}$$

最小２乗法の性質（証明はテキストを参照）

性質1. 残差和は0である

$$\sum_{i=1}^{n} \widehat{u}_i = 0. \tag{10}$$

性質2. 残差と x_i は直交する

$$\sum_{i=1}^{n} \widehat{u}_i x_i = 0.$$

性質3. 観測値の和は，回帰値の和に等しい

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \widehat{y}_i. \tag{11}$$

性質4. 観測値の平均は，回帰値の平均に等しい

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i. \tag{12}$$

性質5. 残差と回帰値は直交する

$$\sum_{i=1}^{n} \widehat{u}_i \widehat{y}_i = 0. \tag{13}$$

性質6. 座標$(\bar{x},\bar{y})$は，回帰直線上に位置する

$$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}. \tag{14}$$

性質7. 観測値と回帰値の積和は，回帰値の平方和に等しい

$$\sum_{i=1}^{n} y_i \hat{y}_i = \sum_{i=1}^{n} \hat{y}_i^2. \tag{15}$$

性質8. R^2は，回帰値$\hat{y}_i$と観測値y_iの相関係数の２乗に等しい **回帰値$\hat{y}_i$と観測値y_iの相関係数は，重相関係数とよばれる．**

性質9. R^2は，y_iとx_iの相関係数の2乗に等しい. (単回帰だけの性質)

予測

予測の方法

$$\widehat{y}_i = \widehat{\alpha} + \widehat{\beta} x_i \tag{16}$$

$$y_{n+1} = \alpha + \beta x_{n+1} + u_{n+1} \tag{17}$$

ただし x_{n+1} **は将来値である。**

$$\widehat{y}_{n+1} = \widehat{\alpha} + \widehat{\beta} x_{n+1} \tag{18}$$

予測分散

$$V(y_{n+1} - \widehat{y}_{n+1}) = \sigma^2 \{1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}\}$$

予測値の信頼空間（95%）

$$\widehat{y}_{n+1} - t_{97.5} \sqrt{\widehat{V}} < y_{n+1} < \widehat{y}_{n+1} + t_{97.5} \sqrt{\widehat{V}}$$

Excelで単回帰分析

ツール→分析ツール→回帰分析→OK→入力Y範囲に非説明変数（分析したいまたは予測したい変数）の範囲、入力X範囲に説明変数（非説明変数の変化を説明できる変数）の範囲を指定→OK。

出力結果の見方

概要

回帰統計	
重相関 R	0.641785
重決定 R2	0.411887
補正 R2	0.390883
標準誤差	20.70401
観測数	30

分散分析表

	自由度	変動	分散	された分散	有意 F
回帰	1	8405.914	8405.914	19.60993	0.000132
残差	28	12002.37	428.656		
合計	29	20408.28			

	係数	標準誤差	t	P−値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	94.39179	171.995	0.548805	0.587489	−257.924	446.7081	−257.924	446.7081
X 値 1	20.47776	4.624284	4.42831	0.000132	11.00534	29.95019	11.00534	29.95019

黄色の部分の値は求めたい係数です。

出力結果の見方

- **重相関：決定係数の平方根** R。**重決定：決定係数** R^2
- **上段にある標準誤差：** s
- **回帰変動：**ESS
- **係数：一つ目は** $\hat{\alpha}$, **二番目は** $\hat{\beta}$
- **下段にある標準誤差：一つ目は** $\hat{\alpha}$ **の標準誤差** s_a、**二番目は** $\hat{\beta}$ **の標準誤差** s_β
- **残差変動：残差二乗和**RSS。**合計：**ESS+RSS=TSS
- P**値：**t **値が対応する点より外側の** t **分布の両裾の確率（面積）**

演習課題5（提出締め切り：12月11日） DATA03をダウンロードし、まずデータの基本統計量と共分散および相関係数をExcelで計算して下さい。ビールの消費量と平均気温の間はどのような関係があるのかを計算の結果を持って説明しよう。そして、さらに、その関係を厳密に分析するため、Excelで分析ツールで単回帰分析を行って、推定式を求めなさい。仮に2007年の5月に生産計画を立てることになっているとする。2007年7月の平均気温の予測値は37度として、2007年にこのブランドのビールをどのぐらい生産すればいいでしょう？（メールのタイトルを学籍番号, 5の形式（例えば、190000,5）でqliu@res.otaru-uc.ac.jpまで提出してください。）