# 回帰係数の推定量の検証

川又 茜　　前田 梨紗
千野 香織　　八木澤 美穂子

# 【概要】

①線形回帰モデルとは何か
②回帰係数の推定法とは何か
③売上データを使って回帰係数の最尤推定値を求める
④乱数を用いて回帰係数の推定量の分布の検証を行う

# 回帰分析とは

目的変数(従属変数)と説明変数(独立変数)の関係を式で表し、目的変数が説明変数によってどの程度説明できるかの分析.

目的変数と説明変数の関係式を、回帰モデルという.

一般的な線形回帰モデル

$$Y_t = X_t'\beta + \varepsilon_t \ , \quad \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, 2, \ldots, T) \quad \sigma > 0$$

$X_t$ ; K個の変数からなる説明変数で既知.

$$X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{Kt} \end{bmatrix}$$ でK次列ベクトル

$X_t'$ ; $X_t$を転置したもの. すなわち $X_t' = [X_{1t} \ X_{2t} \ \ldots X_{Kt}]$ でK次行ベクトル

$\beta$ ; 回帰係数(時刻tの説明変数$X_t$の係数)で未知.

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$ でK次列ベクトル

$Y_t$ ; 時刻tの説明変数$X_t$に対する目的変数

$\varepsilon_t \ (t = 1, 2, \ldots, T)$ ; 誤差項 $N(0, \sigma^2)$

平均0、未知の分散$\sigma^2$の正規分布に独立に従う.

# 最小二乗法による推定

$Y_t$と$X'_t\beta$の違いがなるべく小さくなるような$\beta$を求める。

$E=\sum_{t=1}^{T}(Y_t-X'_t\beta)^2$とおく.

これは二次式となり最小値は一つなので一階の条件を求める.

$$\beta=\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_K\end{bmatrix}$$

$$\frac{\partial E}{\partial \beta}=\begin{bmatrix}\frac{\partial E}{\partial \beta_1}\\ \frac{\partial E}{\partial \beta_2}\\ \vdots\\ \frac{\partial E}{\partial \beta_K}\end{bmatrix}=\begin{bmatrix}\frac{\partial}{\partial \beta_1}\sum_{t=1}^{T}(Y_t-X'_t\beta)^2\\ \frac{\partial}{\partial \beta_2}\sum_{t=1}^{T}(Y_t-X'_t\beta)^2\\ \vdots\\ \frac{\partial}{\partial \beta_K}\sum_{t=1}^{T}(Y_t-X'_t\beta)^2\end{bmatrix}=0$$

# 最尤法による推定

誤差項が正規分布に従うことにより尤度を求める。

$$L(\beta,\sigma^2)=\prod_{t=1}^{T}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(Y_t-X'_t\beta)^2\right)$$

対数尤度は

$$l(\beta,\sigma^2)=\sum_{t=1}^{T}\left(-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma^2-\frac{1}{2\sigma^2}(Y_t-X'_t\beta)^2\right)$$

それぞれの成分毎に計算していくと、

$$\frac{\partial E}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{t=1}^{T} (Y_t - X_t' \beta)^2$$

$$= -2 \sum_{t=1}^{T} X_t (Y_t - X_t' \beta) = 0$$

$\hat{\beta}$：回帰係数 $\beta$ の最小二乗法の推定量

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

$\sum_{t=1}^{T} X_t X_t'$ が逆行列をもつとき

$$\hat{\beta} = \left( \sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t Y_t$$

対数尤度を最大化するための
1階条件は

$$\begin{cases} \dfrac{\partial l(\beta,\sigma^2)}{\partial\beta} = \displaystyle\sum_{t=1}^{T} \frac{1}{\sigma^2} X_t (Y_t - X_t'\beta) = 0 \\ \dfrac{\partial l(\beta,\sigma^2)}{\partial\sigma^2} = \displaystyle\sum_{t=1}^{T} \left( -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(Y_t - X_t'\beta)^2 \right) = 0 \end{cases}$$

これより、$\beta$と$\sigma^2$の最尤推定量は

$$\hat\beta = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t Y_t$$

$$\hat\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}(Y_t - X_t'\hat\beta)^2$$

# 実際にデータを用いて$\boldsymbol{\beta}$と$\boldsymbol{\sigma^2}$の最尤法で推定する．

売上（対数をとったもの）

```
> Y
 [1] 4.9272537 3.8712010 4.9904326 4.0073332 1.3862944 3.0445224 4.6913479 4.1431347 4.7273878 4.4426513
[11] 3.4011974 2.5649494 1.7917595 0.6931472 2.7080502 4.9199809 4.7004804 2.5649494 1.6094379 2.3025851
```

価格（対数をとったもの）

```
> X2
 [1] 5.720312 5.986452 5.690359 5.697093 6.013715 6.001415 5.726848 5.765191 5.710427 5.986452
[11] 5.869297 6.098074 6.098074 6.098074 6.001415 5.707110 5.697093 5.986452 6.008813 6.098074
```

特別陳列の有無

```
> X3
 [1] 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 1 1 0 0 0
```

```
X<-cbind(rep(1,20),X2,X3)
```

$Y_t = X'_t \beta + \varepsilon_t \quad (t=1,2,\cdots,20)\ \varepsilon_t \sim N(0,\sigma^2)$

$Y_1 = 4.93,\ Y_2 = 3.87,\ \cdots,\ Y_{20} = 2.30$

$X_1 = \begin{pmatrix} 1 \\ 5.72 \\ 1 \end{pmatrix},\ X_2 = \begin{pmatrix} 1 \\ 3.87 \\ 1 \end{pmatrix},\ \cdots,\ X_{20} = \begin{pmatrix} 1 \\ 6.10 \\ 0 \end{pmatrix}$

となる。

$\beta$と$\sigma^2$の最尤推定量は

$$\hat{\beta}=\left(\sum_{t=1}^{20} X_t X_t'\right)^{-1}\sum_{t=1}^{20} X_t Y_t$$

$$\hat{\sigma}^2=\frac{1}{20}\sum_{t=1}^{20}\left(Y_t - X_t'\hat{\beta}\right)^2$$

であったので

$\beta$の推定値

$\hat{\beta}_1 = 28.079899$

$\hat{\beta}_2 = -4.294855$

$\hat{\beta}_3 = 1.251444$

$\sigma^2$の推定値

$\hat{\sigma}^2 = 0.2843377$

$\sigma$の推定値

$\hat{\sigma} = 0.5332333$

# 最尤法を用いて推定量$\hat{\beta}$の検証

①右記コマンドで$\varepsilon_t$の正規乱数を10,000個生成。

$\varepsilon_t \sim N(0, \sigma^2)$ $(t=1, 2, \dots, 20)$

```
for(k in 1:10000){
 Yk<-X%*%hbeta+rnorm(20,0,se)
 hbetak<-solve(crossprod(X))%*%crossprod(X,Yk)
 sigma2k<-1/20*sum((Yk-X%*%hbetak)^2)
 beta[,k]<-hbetak
 sig[k]<-sigma2k}
```

(※ここでseは$\sigma$である。)

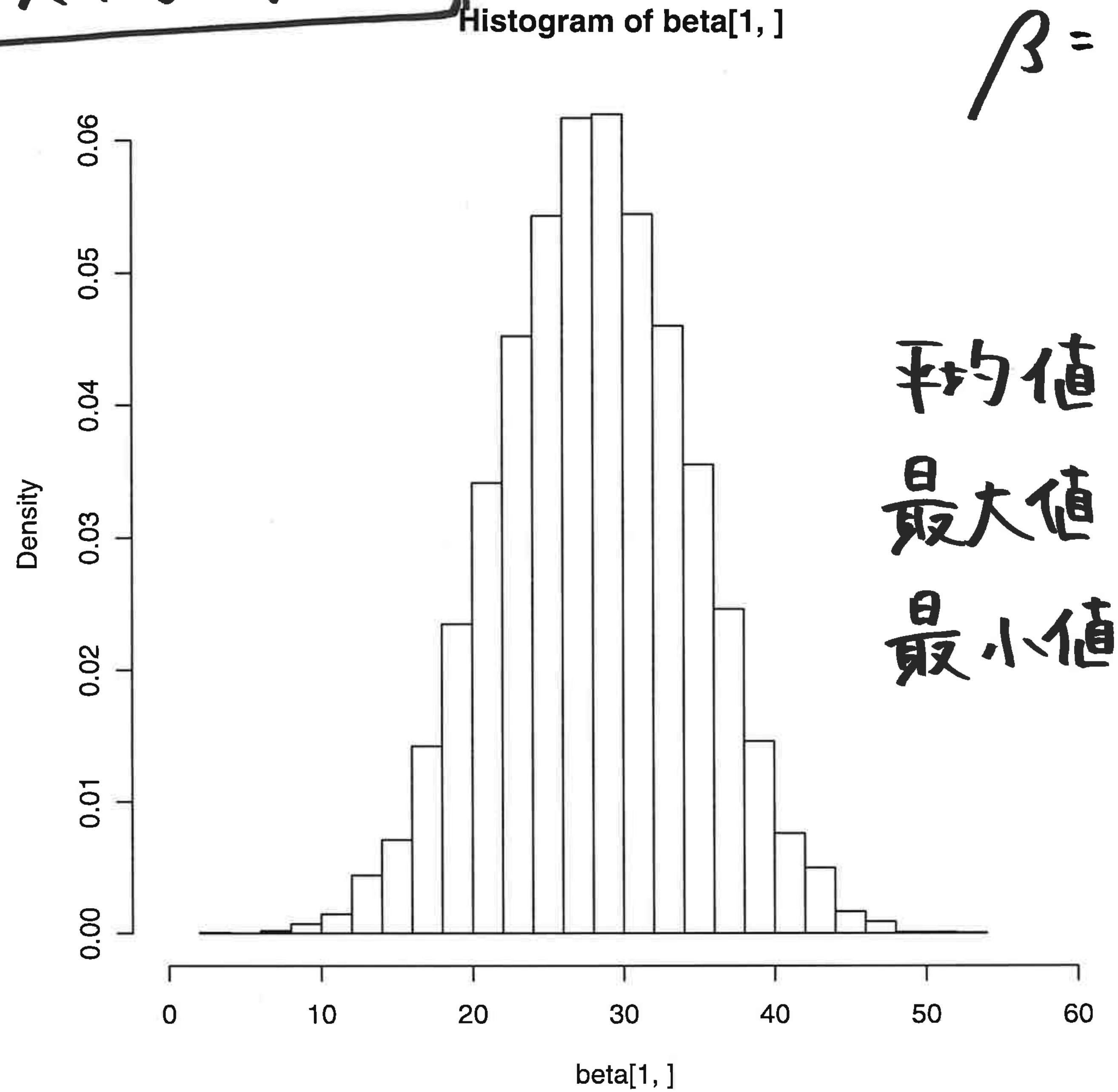
($\sigma^2$と$\beta$は推定値を便宜的に使っている。)

②$Y_t = X_t\beta + \varepsilon_t$に従い$Y_t$を10,000個生成。$(t=1, 2, \dots, 20)$

但し、価格(対数をとったもの)、特別陳列の値は固定。

これより、$\beta$の実現値$\hat{\beta}$が10,000個生成される。

$\hat{\beta}_1$ のヒストグラム




$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

平均値：(28.10729)

最大値：(53.7656)

最小値：(3.363367)

理論的 $\hat{\beta}$ がどのような分布に従うかを求める

$$\hat{\beta} = \left(\sum_{t=1}^{20} X_t X_t'\right)^{-1} \sum_{t=1}^{20} X_t X_t' = \hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3]'$$

であったので、

正規分布の性質を利用して

の期待値ベクトル $E[\hat{\beta}]$

の分散・共分散行列 $V[\hat{\beta}]$ を求めればよい

$$E[\hat{\beta}] = \begin{bmatrix} E[\hat{\beta}_1] \\ E[\hat{\beta}_2] \\ E[\hat{\beta}_3] \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \beta$$

$$V[\hat{\beta}] = \begin{bmatrix} COV(\hat{\beta}_1, \hat{\beta}_1) & COV(\hat{\beta}_1, \hat{\beta}_2) & COV(\hat{\beta}_1, \hat{\beta}_3) \\ COV(\hat{\beta}_2, \hat{\beta}_1) & COV(\hat{\beta}_2, \hat{\beta}_2) & COV(\hat{\beta}_2, \hat{\beta}_3) \\ COV(\hat{\beta}_3, \hat{\beta}_1) & COV(\hat{\beta}_3, \hat{\beta}_2) & COV(\hat{\beta}_3, \hat{\beta}_3) \end{bmatrix}$$

$$= \sigma^2 \left(\sum_{t=1}^{20} X_t X_t'\right)^{-1}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad E[\hat{\beta}] = \begin{bmatrix} E[\hat{\beta}_1] \\ E[\hat{\beta}_2] \\ E[\hat{\beta}_3] \end{bmatrix}$$

$$V[\beta] = \sigma^2 \left( \sum_{t=1}^{20} X_t X_t' \right)^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$ とおくと

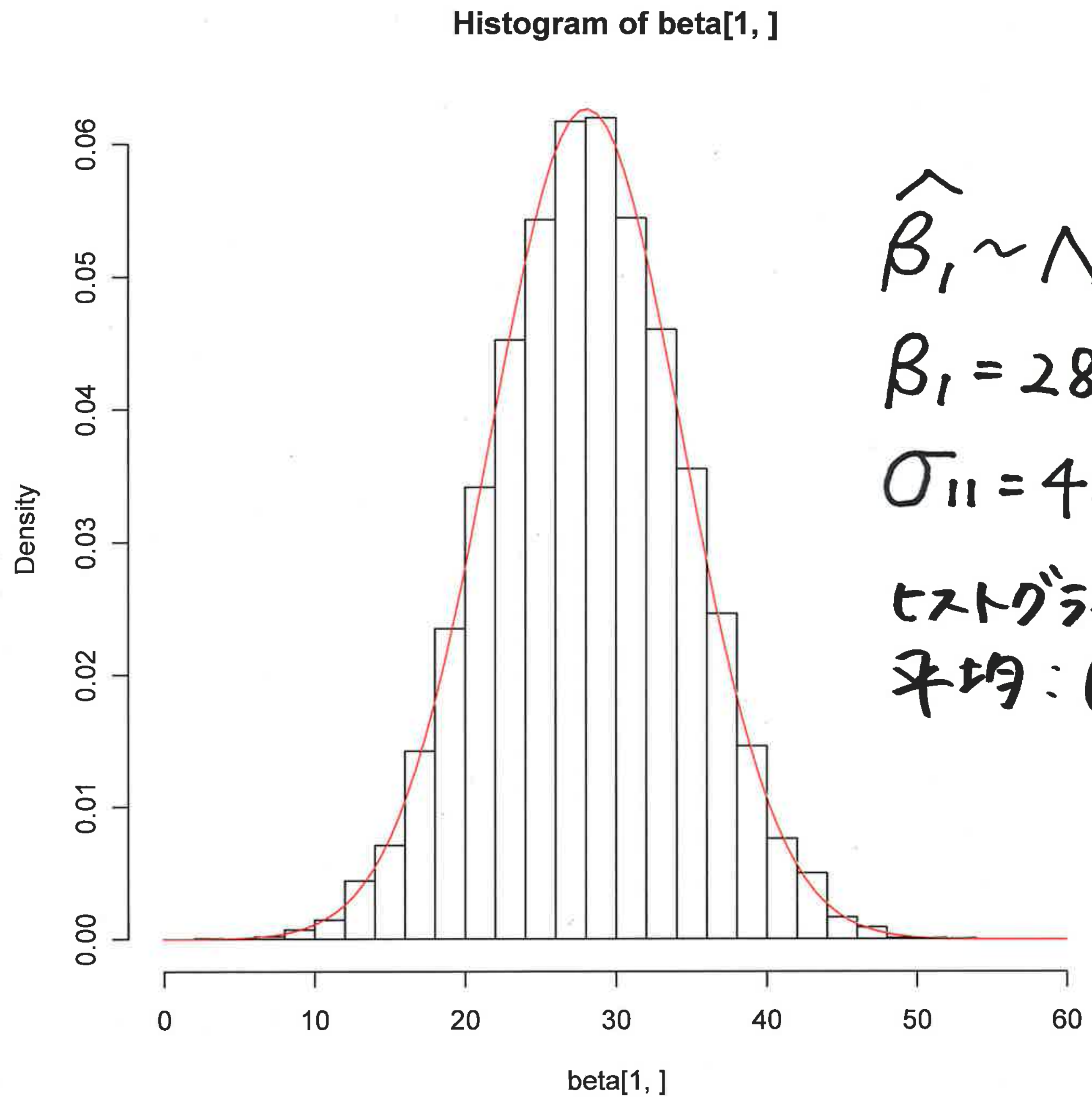
$$\hat{\beta} \sim N\left(\beta, \sigma^2 \left( \sum_{t=1}^{20} X_t X_t' \right)^{-1} \right)$$ となる

これより、

$\hat{\beta}_1$ は $N(\beta_1, \sigma_{11})$

$\hat{\beta}_2$ は $N(\beta_2, \sigma_{22})$

$\hat{\beta}_3$ は $N(\beta_3, \sigma_{33})$ となる。




$\hat{\beta}_1 \sim N(\beta_1, \sigma_{11})$

$\beta_1 = 28.07$

$\sigma_{11} = 40.43420$

ヒストグラムの
平均：(28.10729)