データ

In[1]:= $X = \begin{pmatrix} 1 & 48 & 28 & 304 & 0 \\ 1 & 52 & 24 & 286 & 1 \\ 1 & 61 & 33 & 273 & 1 \\ 1 & 58 & 30 & 295 & 0 \\ 1 & 56 & 21 & 290 & 0 \\ 1 & 46 & 26 & 262 & 0 \\ 1 & 55 & 30 & 318 & 1 \\ 1 & 63 & 37 & 298 & 0 \\ 1 & 52 & 25 & 299 & 1 \\ 1 & 40 & 22 & 313 & 0 \\ 1 & 59 & 34 & 285 & 1 \\ 1 & 57 & 28 & 306 & 0 \\ 1 & 45 & 30 & 291 & 1 \\ 1 & 64 & 21 & 274 & 0 \\ 1 & 53 & 29 & 283 & 0 \end{pmatrix};$

```
y = {3087, 3229, 3204, 3346, 3579,
     2325, 3159, 3589, 2969, 2819,
     3191, 3346, 2444, 3662, 3241};
```

In[3]:= `Dimensions[X]`

Out[3]= {15, 5}

In[4]:=
```
n = 15;
p = 4;
```

1) 偏回帰係数の推定

行列公式で偏回帰係数を求める (表示：有効数字7桁)

In[6]:= $\hat{\beta}$ = Inverse$\left[X^{T}.X\right].X^{T}.y$;
N$\left[\hat{\beta}, 7\right]$

Out[7]= {-1675.193, 54.89659, -23.47004, 8.809766, -140.5539}

2) 目的変数の予測値

まず予測値を求める (表示：有効数字7桁)

In[8]:= $\hat{y}$ = X.$\hat{\beta}$;
N[$\hat{y}$, 7]

Out[9]= {2980.852, 2995.188, 3163.500, 3403.590, 3460.978, 2547.988, 3300.970, 3540.212, 3086.245, 2761.787, 3135.954, 3492.540, 2514.141, 3759.194, 3046.859}

3) 残差変動と標準誤差

まずyの平均値を計算し$\bar{y}$とする

In[10]:= $\bar{y}$ = Mean[y]

Out[10]= 3146

残差変動 $\sum_{i=1}^{n} (y_i - \hat{y}_2)^2$ を計算する

In[11]:= $z_{残差変動}$ = Plus @@ $\left((y - \hat{y})^2\right)$;

残差変動を自由度で割って、誤差分散の不偏推定を計算する (表示：有効数字7桁)

In[12]:= σ2 = N[$z_{残差変動}$ / (n - p - 1), 7]

Out[12]= 25066.20

その平方根を計算して標準誤差を求める

In[13]:= $\sqrt{\sigma 2}$

Out[13]= 158.32309

4) 全変動および回帰変動、決定係数、自由度調整済み決定係数、重相関

全変動 $\sum_{i=1}^{n}(y_i-\bar{y})^2$ および回帰変動 $\sum_{i=1}^{n}(\hat{y}_2-\bar{y})^2$ を計算する

In[14]:= $z_{全変動}$ = Plus @@ $((\bar{y}-y)^2)$;

In[15]:= $z_{回帰変動}$ = Plus @@ $((\bar{y}-\hat{y})^2)$;

回帰変動と全変動の比を計算し、決定係数を求める (表示：有効数字7桁)

In[16]:= R2 = N$\left[\frac{z_{回帰変動}}{z_{全変動}}, 7\right]$

Out[16]= 0.8783224

この量は次にも等しい

In[17]:= R2 = N$\left[1-\frac{z_{残差変動}}{z_{全変動}}, 7\right]$

Out[17]= 0.8783224

その平方根を計算して重相関係数を求める

In[18]:= $\sqrt{R2}$

Out[18]= 0.9371885

これは y とその予測値 ŷ との相関係数に等しい

In[19]:= N[Correlation[ŷ, y], 7]

Out[19]= 0.9371885

相関係数を定義から計算すると以下となる

In[20]:= yn1 = (ŷ - Mean[ŷ]);
yn2 = (y - Mean[y]);
N$\left[\frac{yn1.yn2}{\sqrt{yn1.yn1}\sqrt{yn2.yn2}}, 7\right]$

Out[22]= 0.9371885

自由度で割ったものが自由度調整済み決定係数

In[23]:= $\tilde{R2}$ = N$\left[1-\frac{z_{残差変動}/(n-p-1)}{z_{全変動}/(n-1)}, 7\right]$

Out[23]= 0.8296513

5) 分散共分散行列、標準誤差、t比

行列公式から分散共分散行列を計算する

In[24]:= Σ = σ2 Inverse$[X^{\mathsf{T}}.X]$;
MatrixForm[Σ]

Out[25]//MatrixForm=

$$\begin{pmatrix} 917623.8 & -3212.778 & -154.5763 & -2528.740 & -1207.392 \\ -3212.778 & 46.33289 & -26.57072 & 4.904253 & 58.16780 \\ -154.5763 & -26.57072 & 99.97998 & -3.777969 & -240.1839 \\ -2528.740 & 4.904253 & -3.777969 & 8.112140 & 5.986072 \\ -1207.392 & 58.16780 & -240.1839 & 5.986072 & 7541.492 \end{pmatrix}$$

その対角成分は誤差分散を表す

In[26]:= Diagonal[Σ]

Out[26]= {917623.8, 46.33289, 99.97998, 8.112140, 7541.492}

その対角成分の各々の平方根をとれば標準誤差

In[27]:= err = $\sqrt{\text{Diagonal}[\Sigma]}$

Out[27]= {957.9268, 6.806826, 9.998999, 2.848182, 86.84176}

偏回帰係数と標準誤差の比がｔ比となる

```
In[28]:= t = β̂ / √Diagonal[Σ]
Out[28]= {-1.7487689, 8.064932, -2.347239, 3.093119, -1.6185058}
```

6) p値

自由度ｎ－ｐ－１のｔ分布の裾確率を計算する手続き

```
In[29]:= TailProb = Function[x, 1 - CDF[StudentTDistribution[n - p - 1], Abs[x]]]
Out[29]= Function[x, 1 - CDF[StudentTDistribution[n - p - 1], Abs[x]]]
```

これをｔ比に適用して２倍する（両側検定）

```
In[30]:= 2 (TailProb /@ (β̂ / √Diagonal[Σ]))
Out[30]= {0.1109016, 0.0000109693, 0.04083246, 0.01138331, 0.1366236}
```

7) 信頼区間

両側のｐ値が０．０５になるｔの値を計算する

```
In[31]:= a = Quantile[StudentTDistribution[n - p - 1], 0.05/2]
Out[31]= -2.23
```

９５％信頼区間の上限

```
In[32]:= β̂ + Abs[a] err
Out[32]= {459., 70.1, -1.19, 15.2, 52.9}
```

９５％信頼区間の下限

```
In[33]:= β̂ - Abs[a] err
Out[33]= {-3810., 39.7, -45.7, 2.46, -334.}
```