7　決定係数

7.1　決定係数

全変動 TSS，回帰変動 ESS，残差変動 RSS を次のように定める[*i]。各変動量をデータの大きさ n で割ったものは，観測値，回帰値，残差の分散を表す。

$$\mathrm{TSS} = \sum(y_i - \bar{y})^2, \quad \mathrm{ESS} = \sum(\hat{y}_i - \bar{y})^2, \quad \mathrm{RSS} = \sum(y_i - \hat{y}_i)^2$$

観測値の偏差 $y_i - \bar{y}$ を，回帰部分 $\hat{y}_i - \bar{y}$ と残差部分 $y_i - \hat{y}_i$ に分解する。

$$\begin{aligned}
\mathrm{TSS} &= \sum(y_i - \bar{y})^2 \\
&= \sum(\hat{y}_i - \bar{y} + y_i - \hat{y}_i)^2 \\
&= \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 \\
&\quad + 2\sum(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \qquad (*)
\end{aligned}$$

$(*)$ の項は 0 なので[*ii]，全変動は回帰変動と残差変動の和になる。

$$\begin{aligned}
\mathrm{TSS} &= \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 \\
\mathrm{TSS} &= \mathrm{ESS} + \mathrm{RSS} \qquad (1)
\end{aligned}$$

決定係数（coefficient of determination）　全変動 $\mathrm{TSS} = \mathrm{ESS} + \mathrm{RSS}$ のうち，回帰変動 ESS の占める割合のことを決定係数または寄与率といい，R^2 と表す。

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{ESS} + \mathrm{RSS}} = \frac{\mathrm{ESS}}{\mathrm{TSS}} = \frac{s_{\hat{y}}^2}{s_y^2}, \qquad (0 \leqq R^2 \leqq 1) \qquad (2)$$

決定係数 R^2 は回帰直線の当てはまりの良さを表している。$R^2 = 1$ のとき，全変動 TSS はすべて回帰変動 ESS により説明され，すべての観測点 (x_i, y_i) は回帰直線上にある。$R^2 = 0$ のとき，全変動 TSS はすべて残差変動 RSS によるもので，回帰直線は役立っていない。

決定係数と相関係数　$\hat{y} = a + bx$ のとき，分散の性質 $s_{\hat{y}}^2 = s_{a+bx}^2 = b^2 s_x^2$ から，

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{b^2 \times s_x^2}{s_y^2} = \left(\frac{s_{xy}}{s_x^2}\right)^2 \frac{s_x^2}{s_y^2} = \left(\frac{s_{xy}}{s_x s_y}\right)^2 = (r_{xy})^2$$

単回帰（$\hat{y} = a + bx$）のとき，決定係数は x と y の相関係数の 2 乗に等しい。

$$R^2 = (r_{xy})^2 \qquad (3)$$

[*i] Total sum of squares; Explained sum of squares; Residual sum of squares

[*ii] 正規方程式から，$(*) = 2(a - \bar{y})\sum(y_i - a - bx_i) + 2b\sum x_i(y_i - a - bx_i) = 0$

7.2 決定係数の求め方

決定係数の求め方（1）

変数	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	$\hat{y}$	$(\hat{y}-\bar{y})^2$
a.	**21**	**11**	−17	−12	289	144	204	**10.25**	162.56
b.	**26**	**12**	−12	−11	144	121	132	**14.00**	81.00
c.	**29**	**12**	−9	−11	81	121	99	**16.25**	45.56
d.	**30**	**17**	−8	−6	64	36	48	**17.00**	36.00
e.	**32**	**19**	−6	−4	36	16	24	**18.50**	20.25
f.	**34**	**26**	−4	3	16	9	−12	**20.00**	9.00
g.	**40**	**28**	2	5	4	25	10	**24.50**	2.25
h.	**42**	**23**	4	0	16	0	0	**26.00**	9.00
i.	**52**	**30**	14	7	196	49	98	**33.50**	110.25
j.	**55**	**39**	17	16	289	256	272	**35.75**	162.56
k.	**57**	**36**	19	13	361	169	247	**37.25**	203.06
合計	418	253	0	0	1496	946	1122	253	841.5
平均	38	23	0	0	136	86	102	23	76.5

平均，分散，共分散は，

$$\bar{x}=38, \qquad \bar{y}=23, \qquad s_x^2=136, \qquad s_y^2=86, \qquad s_{xy}=102$$

回帰係数 b，回帰定数 a は，

$$b=\frac{s_{xy}}{s_x^2}=\frac{102}{136}=0.75, \qquad a=\bar{y}-b\bar{x}=23-\frac{102}{136}\cdot 38=-5.5$$

なので，回帰式 $\hat{y}=a+bx$ は次のとおり。または $\hat{y}-\bar{y}=b(x-\bar{x})$ でもよい。

$$\hat{y}=-4.5+0.75x, \qquad \hat{y}-23=0.75(x-38)$$

$x_1=21$ に対応する回帰値は $\hat{y}_1=10.25$ である（他の回帰値も同様にする）。

$$\hat{y}_1=-5.5+0.75\cdot 21=10.25$$

回帰値 $\hat{y}_i$ の平均，分散は，

$$\bar{\hat{y}}=23, \qquad s_{\hat{y}}^2=76.5$$

よって決定係数は，$R^2=s_{\hat{y}}^2/s_y^2$ または $R^2={r_{xy}}^2$ により，$R^2=0.890$ となる。

$$R^2=\frac{76.5}{86}=0.890, \qquad R^2=\left(\frac{102}{\sqrt{136}\sqrt{86}}\right)^2=0.890$$

決定係数の求め方（2）　決定係数は次のように求めることもできる。

変数	x	y	x^2	y^2	xy	$\hat{y}$	$\hat{y}^2$
a.	**21**	**11**	441	121	231	**10.25**	105.06
b.	**26**	**12**	676	144	312	**14.00**	196.00
c.	**29**	**12**	841	144	348	**16.25**	264.06
d.	**30**	**17**	900	289	510	**17.00**	289.00
e.	**32**	**19**	1024	361	608	**18.50**	342.25
f.	**34**	**26**	1156	676	884	**20.00**	400.00
g.	**40**	**28**	1600	784	1120	**24.50**	600.25
h.	**42**	**23**	1764	529	966	**26.00**	676.00
i.	**52**	**30**	2704	900	1560	**33.50**	1122.25
j.	**55**	**39**	3025	1521	2145	**35.75**	1278.06
k.	**57**	**36**	3249	1296	2052	**37.25**	1387.56
合計	418	253	17380	6765	10736	253	6660.5
平均	38	23	1580	615	976	23	605.5

平均は $\bar{x} = 38$，$\bar{y} = 23$ である。分散，共分散は，

$$s_x^2 = 1580 - 38^2 = 136, \qquad s_y^2 = 615 - 23^2 = 86$$
$$s_{xy} = 976 - 38 \cdot 23 = 102$$

回帰係数 b，回帰定数 a は，

$$b = \frac{s_{xy}}{s_x^2} = \frac{102}{136} = 0.75, \qquad a = \bar{y} - b\bar{x} = 23 - \frac{102}{136} \cdot 38 = -5.5$$

なので，回帰式 $\hat{y} = a + bx$ は次のとおり。または $\hat{y} - \bar{y} = b(x - \bar{x})$ でもよい。

$$\hat{y} = -5.5 + 0.75x, \qquad \hat{y} - 23 = 0.75(x - 38)$$

回帰値 $\hat{y}_i$ の平均，分散は，

$$\bar{\hat{y}} = 23, \qquad s_{\hat{y}}^2 = 605.5 - 23^2 = 76.5$$

よって決定係数は，$R^2 = s_{\hat{y}}^2 / s_y^2$ または $R^2 = {r_{xy}}^2$ により，$R^2 = 0.890$ となる。

$$R^2 = \frac{76.5}{86} = 0.890, \qquad R^2 = \left(\frac{102}{\sqrt{136}\sqrt{86}} \right)^2 = 0.890$$

決定係数の求め方（3）　「回帰直線の求め方（2）」の表から，次のように求めることもできる。平均は $\bar{x} = 418/11 = 38$，$\bar{y} = 253/11 = 23$ である。分散，共分散は，

$$s_x^2 = \frac{11 \cdot 17380 - 418^2}{11^2} = \frac{16456}{11^2}, \quad s_y^2 = \frac{11 \cdot 6765 - 253^2}{11^2} = \frac{10406}{11^2}$$
$$s_{xy} = \frac{11 \cdot 10736 - 418 \cdot 253}{11^2} = \frac{12342}{11^2}$$

回帰係数 b，回帰定数 a は，

$$b = \frac{n^2 \cdot s_{xy}}{n^2 \cdot s_x^2} = \frac{12342}{16456} = 0.75, \qquad a = \bar{y} - b\bar{x} = 23 - \frac{12342}{16456} \cdot 23 = -5.5$$

なので，回帰式 $\hat{y} = a + bx$ は次のとおり。または $\hat{y} - \bar{y} = b(x - \bar{x})$ でもよい。

$$\hat{y} = -5.5 + 0.75x, \qquad \hat{y} - 23 = 0.75(x - 38)$$

回帰値 $\hat{y}_i$ の平均，分散は，

$$\bar{\hat{y}} = 23, \qquad s_{\hat{y}}^2 = \frac{11 \cdot 6660.5 - 253^2}{11^2} = \frac{9256.5}{11^2}$$

よって決定係数は，$R^2 = s_{\hat{y}}^2 / s_y^2$ または $R^2 = {r_{xy}}^2$ により，$R^2 = 0.890$ となる。

$$R^2 = \frac{9256.5}{10406} = 0.890, \qquad R^2 = \left(\frac{12342}{\sqrt{16456}\sqrt{10406}}\right)^2 = 0.890$$

本稿の参考文献

統計学入門（基礎統計学）
東京大学教養学部統計学教室（編）東京大学出版会 978-4-13-042065-5

統計学
久保川 達也 （著） 東京大学出版会 978-4-13-062921-8

経済・経営系のためのよくわかる統計学
前川 功一 （編著） 朝倉書店 978-4-254-12197-1

共立講座２１世紀の数学 第２版 １４ 統計
竹村 彰通 （著） 共立出版 978-4-320-01851-8

経済・経営系のための統計入門 （事例でわかる統計シリーズ）
景山 三平 （監修） 実教出版 978-4-407-33711-2

統計学序論 改訂版
山本 義郎 （著） 東海大学出版部 978-4-486-02133-9

統計学基礎 日本統計学会公式認定統計検定２級対応 改訂版
日本統計学会 （編） 東京図書 978-4-489-02227-2

基本統計学 第４版
宮川 公男 （著） 有斐閣 978-4-641-16455-0

www5e.biglobe.ne.jp/~emm386/statistics/