

【第3章補足資料】

表 3.5 四半期国内総支出：季節調整値(表書籍版 p.74 参照)

実質国内総支出		(単位：億円)			
年 \ 四半期	I	II	III	IV	
1982		609,281 (2,558,437)	640,910 (2,574,580)	714,022 (2,588,606)	
1983	616,900 (2,603,555)	625,948 (658,319)	663,123 (2,666,293)	738,699 (2,678,277)	
1984	645,979 (2,723,116)	2,762,617 (689,857)	694,175 (2,790,946)	778,250 (2,825,146)	
1985	677,873 (2,852,625)	2,894,108 (704,540)	726,254 (2,920,192)	813,648 (2,954,325)	
1986	698,974 (2,935,762)	2,967,011 (726,130)	744,947 (2,995,432)	830,711 (3,014,178)	
1987	726,134 (3,047,800)	3,063,143	776,419 (3,119,316)	876,440 (3,177,288)	
1988	776,321 (3,253,924)	-	-	-	

(出典：1989年度『国民経済計算年報』経済企画庁経済研究所)

【第3章補足資料】

図 3.36 の原データ (書籍版 p.68 参照)

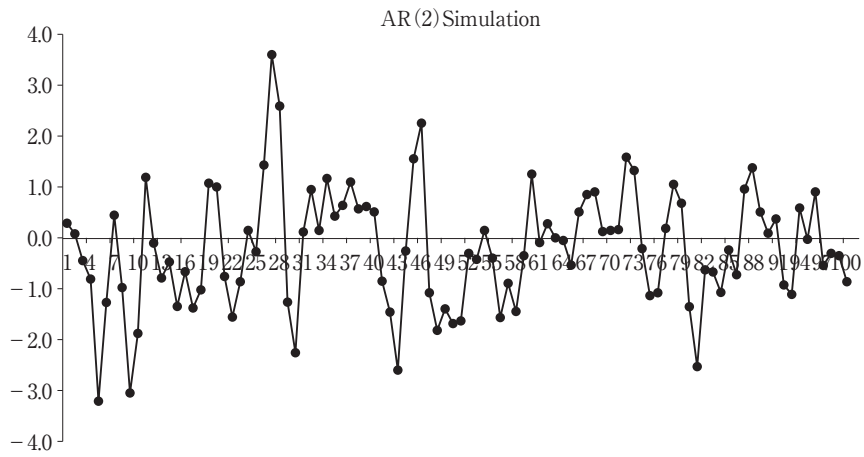


図 3.9 自己回帰モデル AR(2)のシミュレーションより (書籍版 p.75 関連)

【第3章補足資料】

年	1	2	3	4	5	6	7	8	9	10	11	12	年平均
1958	0.00	0.00	315.58	317.18	317.43	0.00	315.71	314.71	313.04	0.00	313.07	314.34	0.00
1959	315.16	315.97	316.37	317.40	317.96	317.82	316.23	314.54	313.60	313.03	314.57	315.32	315.66
1960	316.10	316.68	317.37	318.79	319.63	319.29	317.86	315.55	313.85	313.64	314.61	315.81	316.59
1961	316.54	317.34	318.12	319.06	320.20	319.44	318.24	316.52	314.57	315.13	315.75	316.73	317.30
1962	317.70	318.29	319.37	320.25	320.84	320.43	319.35	317.13	316.01	315.19	316.42	317.47	318.20
1963	318.45	318.82	319.72	321.06	321.87	321.22	319.44	317.48	315.89	315.83	316.72	317.98	318.71
1964	319.17	0.00	0.00	0.00	322.08	321.92	320.42	318.58	316.68	316.65	317.60	318.49	0.00
1965	319.32	320.36	320.82	322.06	322.17	321.95	321.20	318.81	317.82	317.37	318.93	319.09	319.99
1966	319.94	320.98	321.81	323.03	323.36	323.11	321.65	319.64	317.86	317.25	319.06	320.26	320.66
1967	321.65	321.81	322.36	323.67	324.17	323.39	321.93	320.29	318.58	318.60	319.98	321.25	321.47
1968	321.88	322.47	323.17	324.23	324.88	324.75	323.47	321.34	319.56	319.45	320.45	321.92	322.30
1969	323.40	324.21	325.33	326.31	327.01	326.24	325.37	323.12	321.85	321.31	322.31	323.72	324.18
1970	324.60	325.57	326.55	327.80	327.80	327.54	326.28	324.63	323.12	323.11	323.99	325.09	325.51
1971	326.12	326.61	327.16	327.92	329.14	328.80	327.52	325.62	323.61	323.80	325.10	326.25	326.47
1972	326.93	327.83	327.95	329.91	330.22	329.25	328.11	326.39	324.97	325.32	326.54	327.71	327.59
1973	328.73	329.69	330.47	331.69	332.65	332.24	331.03	329.36	327.60	327.29	328.28	328.79	329.82
1974	329.45	330.89	331.63	332.85	333.28	332.47	331.34	329.53	327.57	327.57	328.53	329.69	330.40
1975	330.45	330.97	331.64	332.87	333.61	333.55	331.90	330.05	328.58	328.31	329.41	330.63	331.00
1976	331.63	332.46	333.36	334.45	334.82	334.32	333.05	330.87	329.24	328.87	330.18	331.50	332.06
1977	332.81	333.23	334.55	335.82	336.44	335.99	334.65	332.41	331.32	330.73	332.05	333.53	333.63
1978	334.66	335.07	336.33	337.39	337.65	337.57	336.25	334.39	332.44	332.25	333.59	334.76	335.20
1979	335.89	336.44	337.63	338.54	339.06	338.95	337.41	335.71	333.68	333.69	335.05	336.53	336.55
1980	337.81	338.16	339.88	340.57	341.19	340.87	339.25	337.19	335.49	335.51	336.63	337.74	338.36

【第3章補足資料】

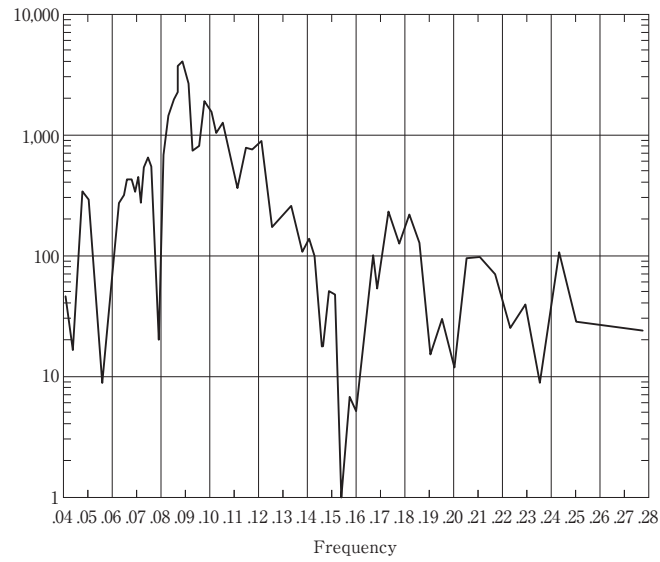


図 3.10 黒点数データのスペクトル推定(書籍版 p.78 参照)

【第3章補足資料】

<主成分分析の計算方法の解説：第3章表3.2(書籍版 p.60)についての補足>
数学と国語のテストの例題を用いて、主成分分析の計算方法の詳細を示す。
あわせて、初学者のために「固有値問題」の早わかり解説を用意した。

1. 1次元指標の形成—分散共分散行列の固有値問題

数学と国語のテストの例題では、これから要約の記述統計として、 x , y の標準偏差, 分散, 相関係数, 共分散は

$$s_x=20, s_y=15, s_x^2=400, s_y^2=225, r_{xy}=0.5, s_{xy}=20 \cdot 15 \cdot 0.5=150$$

となっている。いま、数学、国語の成績 x , y をもとに総合点数(指標とよぶ)

$$f=l_1x+l_2y$$

を作り、学力が f でうまく総合判断しやすいよう f が大きく広がることを目的とする。 (l_1, l_2) の符号組み合わせで f の意味が異なってくるから、2通りの f が予想される)そのためには、分散の公式から

$$s_f^2=l_1^2s_x^2+2l_1l_2s_{xy}+l_2^2s_y^2$$

であるが、 l_1, l_2 の値を制約しておかないと最大化には意味がないから

$$\text{条件} \quad l_1^2+l_2^2=1 \quad (\text{ベクトルの長さ}=1)$$

のもとで、

$$s_f^2=l_1^2s_x^2+2l_1l_2s_{xy}+l_2^2s_y^2$$

の最大値を求める。条件付き最適化の手続きに従い λ をラグランジュの未定乗数として、関数

$$\mathcal{L}(l_1, l_2, \lambda)=l_1^2s_x^2+2l_1l_2s_{xy}+l_2^2s_y^2-\lambda(l_1^2+l_2^2-1)$$

の(無条件)最大化の条件を求めよう。すなわち l_1, l_2 の連立方程式が得られる

$$\partial \mathcal{L} / \partial l_1 = 2l_1s_x^2 + 2l_2s_{xy} - 2\lambda l_1 = 0,$$

$$\partial \mathcal{L} / \partial l_2 = 2l_1s_{xy} + 2l_2s_y^2 - 2\lambda l_2 = 0$$

行列形式の方が見やすく、次の分散共分散行列の「固有値問題」(Eigenvalue problem)となる：

【第3章補足資料】

$$\begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \lambda \begin{pmatrix} l_1 \\ l_2 \end{pmatrix}$$



ここで λ を「固有値」、ベクトル (l_1, l_2) を「固有ベクトル」という。この解き方の定跡は、まず

$$\begin{pmatrix} s_x^2 - \lambda & s_{xy} \\ s_{xy} & s_y^2 - \lambda \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

となるが、右辺が0であるため l_1, l_2 がともに0はもちろん解である。これは無意味であり、有意な(「自明でない」non-trivial)解が(一つでも、ただし実際は無限に)あるための必要十分条件は、係数の行列式

$$\begin{vmatrix} s_x^2 - \lambda & s_{xy} \\ s_{xy} & s_y^2 - \lambda \end{vmatrix} = 0$$

である(よく知られているので証明略)。これを展開して結局

$$\lambda^2 - (s_x^2 + s_y^2)\lambda - (s_x^2 s_y^2 - s_{xy}^2) = 0 \quad (\text{固有方程式})$$

を得る。ここでできたこの2根が固有値 λ_1, λ_2 、それに対応する l_1, l_2 がそれぞれの固有ベクトルである。実際、 λ_1, λ_2 をそれぞれ代入して l_1, l_2 を解く際、当然不定(解は一つに定まらない)となるので、ここでは $l_1^2 + l_2^2 = 1$ のように長さ1に規格化(normalize)する。

例 この具体例では、固有値問題

$$\ast \quad \begin{pmatrix} 400 & 150 \\ 150 & 225 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \lambda \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \quad (\text{連立方程式})$$



を解く固有方程式は

$$\lambda^2 - 625\lambda + 67500 = 0$$

となる。ここで625, 67500が重要だが後で触れる。解の公式より

$$\lambda = \frac{625 \pm \sqrt{625^2 - 4 \cdot 67500}}{2} = \frac{625 \pm \sqrt{120625}}{2}$$

【第3章補足資料】

より、2つの固有値は大きさの順に、

$$\lambda_1=486.15, \lambda_2=138.85$$

次に固有ベクトル(l_1, l_2)の一例は、 $\lambda_1=486.15, \lambda_2=138.85$ それぞれに対し、元の連立方程式※に戻って(1, 0.5743), (0.5743, -1)などが得られる。このベクトルは長さ1にはではない。長さ1に規格化した固有ベクトルは

$$\left(\frac{1}{\sqrt{1^2+0.5743^2}}, \frac{0.5743}{\sqrt{1^2+0.5743^2}} \right) = (0.8672, 0.4980)$$

同様に、

$$\left(\frac{0.5743}{\sqrt{0.5743^2+(-1)^2}}, \frac{-1}{\sqrt{0.5743^2+(-1)^2}} \right) = (0.498, -0.8672)$$

のように求められる。予想通り f は2通り得られて

$$f_1=0.8672x+0.4980y, \quad f_2=0.4980x-0.8672y$$

ここで、今後結果をみるために次の数学的事項を指摘しておこう。固有値は2次方程式の2解だから、高校来の知識を使い、

- ii) 固有値は実数、しかも正(>0)
- iii) 固有値の和 = 対角項の和すなわちトレース： $s_x^2+s_y^2=625$
- iv) 固有値の積 = 行列式の値： $s_x^2s_y^2-s_{xy}^2=67500$
- v) 異なった固有値に対する固有ベクトルは直交



2. 主成分を抜き出す

以上、 x, y データから得られた2通りの指標

$$f_1=0.8672x+0.4980y, \quad f_2=0.4980x-0.8672y$$

を「第1主成分」1st principal component, PC1, 「第2主成分」2nd principal component, PC2 という。各係数の符号に注意しておこう。主成分の性質は以下に述べる通り。うまくできている。

1) PC1の分散は

$$s_{f_1}^2=0.8672^2 \cdot 400 + 2 \cdot 0.8672 \cdot 0.4980 \cdot 150 + 0.4980^2 \cdot 225$$

【第3章補足資料】

$$\begin{aligned}
 &= 300.8 + 129.6 + 55.8 \\
 &= 496.2 \\
 &= \lambda_1
 \end{aligned}$$

PC2 の分散は同様に $s_{f_2}^2 = 138.9 = \lambda_2$ で、第1主成分のそれより小さい。

すなわち、各主成分 PC1, 2 の分散は対応する固有値に等しい。実際 i) から必ず > 0 となる

$$s_{f_1}^2 + s_{f_2}^2 = \lambda_1 + \lambda_2 = 625$$

したがって、PC1, PC2 の重要度 (寄与率 contribution) の割合は、それぞれ

$$486.15/625 = 0.778 (77.8\%), \quad 138.85/625 = 0.222 (22.2\%)$$

たしかに PC1 は、文字通り重要であり、PC2 はこれを補っている。

2) 第1主成分 f_1 と第2主成分 f_2 の共分散の計算は ($\bar{x} = \bar{y} = 0$ に注意)

$$\begin{aligned}
 &\Sigma(0.8672x + 0.4980y)(0.4980x - 0.8672y)/10 \quad (\Sigma: \text{サンプル上での和}) \\
 &= 0.8672 \cdot 0.4980 \cdot 400 + (0.4980^2 - 0.8672^2) \cdot 150 - 0.4970 \cdot 0.8672 \cdot 225 \\
 &= 172.8 - 75.6 - 92.2 \\
 &= 0
 \end{aligned}$$

で相関関係はない (無相関)。これは固有ベクトルの直交性からも導かれる。

(x, y) サンプル中の分散は互いに無相関のこれらの2つの主成分 f_1, f_2 の要因に完全に分解される。

以上の結果、次のように要約される (実際の計算はコンピュータが行う)。

主成分結果 (要約表)

	第1主成分 (PC1)	第2主成分 (PC2)	計
数学 x	0.8672	0.4980	
国語 y	0.4980	-0.8672	
分散 (固有値)	486.15	138.85	625.00
寄与率 (%)	77.8	22.2	100.0

【第3章補足資料】

③ 主成分の解釈とネーミング

x , y には実体的な意味があるが,

$$f_1=0.8672x+0.4980y, \quad f_2=0.4980x-0.8672y,$$

は数理上の操作の結果で固有の意味は与えられていないため、各主成分は何をあらわすかを解釈し、命名するのがふつうである。これはコンピュータにはできず人間(のみ)が行う。機械学習への応用でも変わらない。

要約表にある係数(主成分への「負荷量」という)をみると、第1主成分には数学 x 、国語 y の学力がともに+で入っており、第1主成分は「総合的能力」(ある意味での「実力」)の要素を表すことができる。第2主成分には、数学 x のみが+で、国語 y のような文系的傾向は-となっているから、文系-理系を対照的に表す軸の上で、「理系的能力」の要素をあらわすと解釈できる。

もとより、このような「解釈」や「ネーミング」は一通りではなく、なにがしかの主観やその分野のさまざまな知識にもとづくが、これは複雑多様な現象の反映である。これを容認してこそ主成分分析による現象のモデル化が可能であり、力を発揮する。