

主成分分析演習問題解答

2002年10月31日

作成：森 一将

1 対象データ

対象となるデータは(表1)の水質データです。

2 相関行列を使った主成分分析

このデータを使って主成分分析を行います。
まず、(表1)のデータの相関行列を求めます。(表2)

この行列を C とします。 C は以下の行列式で表すことができます。

$$C = \begin{pmatrix} 1 & 0.649544189 & 0.515332317 \\ 0.649544189 & 1 & 0.415381519 \\ 0.515332317 & 0.415381519 & 1 \end{pmatrix}$$

したがって、以下の固有方程式を解くことで、固有値、固有ベクトルが求められます。

$$|C - \lambda I| = \begin{vmatrix} 1 - \lambda & 0.649544189 & 0.515332317 \\ 0.649544189 & 1 - \lambda & 0.415381519 \\ 0.515332317 & 0.415381519 & 1 - \lambda \end{vmatrix}$$

この計算は非常に複雑ですので、QMSS サイトに掲載してある固有値計算用の EXCEL シートを使い、計算します。
すると、結果が(表3)のようになります。

相関行列を用いた場合、主成分得点は標準化されたデータに対して求められます。つまり、以下の変換を元のデータに対して行います。

$$X = \frac{x - \bar{x}}{\sigma}$$

観測所	BOD	NO3-N	NH3-N	観測所	BOD	NO3-N	NH3-N
1	2.27	1.97	0.11	20	3.19	8.74	0.88
2	4.41	12.83	0.61	21	3.22	9.80	0.95
3	4.03	11.11	0.53	22	3.17	9.64	0.93
4	3.75	9.86	0.47	23	3.13	9.49	0.90
5	3.37	9.54	0.62	24	3.08	9.34	0.88
6	3.23	8.85	0.56	25	3.04	9.20	0.86
7	3.18	8.02	0.64	26	3.00	9.06	0.84
8	4.08	8.94	1.14	27	2.96	8.03	0.82
9	4.00	8.76	1.11	28	2.93	8.81	0.80
10	3.92	8.59	1.07	29	2.89	8.69	0.78
11	3.83	8.43	1.04	30	2.86	8.57	0.76
12	3.74	8.27	1.00	31	2.82	8.45	0.74
13	3.66	8.13	0.97	32	2.79	8.35	0.73
14	3.58	7.99	0.94	33	2.76	8.24	0.71
15	3.16	6.72	0.83	34	2.73	8.14	0.70
16	3.43	9.23	0.94	35	2.70	8.04	0.68
17	3.36	9.10	0.93	36	2.51	6.54	0.48
18	3.30	8.97	0.91	37	2.49	6.51	0.47
19	3.24	8.85	0.89	38	2.46	6.46	0.46

平均	3.218	8.533	0.781
標準偏差	0.496	1.58	0.213
変動係数	15.6 %	18.5 %	27.3 %

表 1: 水質データ

	BOD	NO3-N	NH3-N
BOD	1	0.649544	0.515332
NO3-N	0.649544	1	0.415382
NH3-N	0.515332	0.415382	1

表 2: 水質データの相関行列

固有値	第 1 主成分	第 2 主成分	第 3 主成分
	2.059476	0.604986	0.335538
固有ベクトル	第 1 主成分	第 2 主成分	第 3 主成分
	0.615508	-0.20522	0.760942
	0.584594	-0.52866	-0.61544
	0.528583	0.823651	-0.20542
寄与率 (%)	第 1 主成分	第 2 主成分	第 3 主成分
	68.64919	20.1662	11.18461

表 3: 水質データの固有値計算の結果

標準化したデータに対して, 上記で求めた固有ベクトル a_i を使い, 以下の式でそれぞれの主成分得点を求めます.

$$Z = a_1x_1 + a_2x_2 + a_3x_3$$

この計算の結果を (表 4) に示します.

観測所	主成分			観測所	主成分		
	第 1 主成分	第 2 主成分	第 3 主成分		第 1 主成分	第 2 主成分	第 3 主成分
1	-5.2738	-0.00347	0.454295	20	0.287741	0.324788	-0.02825
2	2.646675	-2.59206	-0.00829	21	0.891318	0.22796	-0.32745
3	1.339218	-2.16824	0.001185	22	0.720309	0.224918	-0.36113
4	0.379646	-1.86583	0.000392	23	0.540587	0.175756	-0.39299
5	0.161765	-1.02206	-0.31371	24	0.37328	0.169366	-0.43057
6	-0.4166	-0.96501	-0.31768	25	0.222096	0.155494	-0.45669
7	-0.58732	-0.35734	0.005674	26	0.070912	0.141623	-0.48282
8	2.112842	0.894584	1.50964	27	-0.4098	0.425733	-0.16252
9	1.872335	0.871999	1.428088	28	-0.20793	0.099703	-0.5314
10	1.610694	0.807427	1.333019	29	-0.35171	0.079135	-0.56532
11	1.365173	0.782279	1.22834	30	-0.48307	0.054435	-0.58388
12	1.094815	0.718492	1.114036	31	-0.62685	0.033867	-0.61779
13	0.869118	0.682515	1.016915	32	-0.72597	0.041109	-0.63452
14	0.643421	0.646537	0.919793	33	-0.85362	0.01306	-0.65698
15	-0.62161	0.820311	0.663865	34	-0.95274	0.020302	-0.67371
16	0.916244	0.293368	0.206994	35	-1.0767	-0.01109	-0.70007
17	0.75634	0.327187	0.140571	36	-2.36476	-0.20312	-0.6002
18	0.58402	0.318233	0.079867	37	-2.42554	-0.22345	-0.62884
19	0.415402	0.305932	0.015271	38	-2.50615	-0.23295	-0.66503

表 4: 主成分得点の計算結果 (相関行列を使用)