

以秩和为因子的一种多元判别法

吴绍敏 彭 沛

(应用数学系)

摘 要

本文提出一种以秩和为因子及以绝对距离为判别函数的判别法。使用该方法,可不受模型的限制,能充分地利用大量因子的信息,计算量较小,并可作趋势和定值预报。

一、引 言

本文的基本思想:(1)为了摆脱模型的限制,扩大方法的应用范围,直接从实测数据出发,从中提取统计信息的规律,然后按这种规律进行预测。(2)因不考虑预测对象的变量模型,为保证方法有较好的预测效率和可靠性,应用计算机寻找大量的预报因子,并采用秩和这个统计量去集中部份因子的信息,尔后把这些秩和作为新的因子参加预测,这样就能充分地利用大量因子的信息。(3)因判别函数的不同,就可得不同的判别预测方法。为了简化计算,本文采用绝对距离来建立判别函数。

现结合具体实例,把方法介绍如下。

二、方 法 步 骤

根据龙溪地区气象台1955~1984年的观测资料,对漳州地区1984年4月降雨量的预测,来具体说明该方法的各步骤。

1. 确定预报对象与初选因子

(1)设预报对象 Y 为漳州4月降雨量,其历史数据为 y_1, y_2, \dots, y_N ,列入表(2)(此例 $N=28$)。对 Y 作秩变换: $R(y_t) = r_t, t=1, 2, \dots, N$,其中 r_t 为 y_1, y_2, \dots, y_N 按从小到大的顺序排列, y_t 所对应的次序,亦称 r_t 为 y_t 的秩数,并将 Y 之秩数 $R(Y): r_1, r_2, \dots, r_N$ 列表(2)。

(2)利用秩相关系数检验法,在信度 $\alpha=0.10$ 下,此例从287个因素中选取35个与 Y 秩相关较大的作为初选因子。这里我们只列出最后参加预报的25个因子,并记:

本文1985年7月13日收到。

X_1 —漳州5月下旬绝对湿度; $R_{1y} = -0.6853$

X_3 —漳州1月上旬绝对湿度; $R_{3y} = 0.5315$

X_5 —漳州3月中旬最低气温; $R_{5y} = 0.5030$

X_7 —漳州5月下旬极端最高气温; $R_{7y} = -0.4449$

X_9 —漳州5月下旬最低气温; $R_{9y} = -0.4273$

X_{11} —漳州5月上旬最低气温; $R_{11y} = -0.4221$

X_{13} —漳州10月下旬极端最高气温; $R_{13y} = 0.4247$

X_{15} —漳州9月上旬平均气压; $R_{15y} = 0.4131$

X_{17} —漳州8月中旬绝对湿度; $R_{17y} = 0.4002$

X_{19} —漳州5月下旬极端最高气压; $R_{19y} = 0.3830$

X_{21} —漳州7月中旬绝对湿度; $R_{21y} = -0.3863$

X_{23} —漳州3月中旬平均气压; $R_{23y} = -0.3612$

X_{25} —漳州7月下旬最低气温; $R_{25y} = -0.3700$

X_2 —漳州1月上旬平均气温; $R_{2y} = 0.6230$

X_4 —漳州5月下旬平均气温; $R_{4y} = -0.5257$

X_6 —漳州1月上旬最低气温; $R_{6y} = 0.4647$

X_8 —漳州7月中旬极端最高气温; $R_{8y} = -0.4522$

X_{10} —漳州10月中旬平均气温; $R_{10y} = 0.4267$

X_{12} —漳州5月中旬最低气压; $R_{12y} = -0.4319$

X_{14} —漳州8月上旬最低气温; $R_{14y} = 0.4143$

X_{16} —漳州10月上旬雨量; $R_{16y} = -0.4023$

X_{18} —漳州8月中旬最低气压; $R_{18y} = -0.3937$

X_{20} —漳州10月上旬最低气温; $R_{20y} = 0.3766$

X_{22} —漳州3月上旬极端最高气温; $R_{22y} = 0.3608$

X_{24} —漳州3月中旬绝对湿度; $R_{24y} = 0.3626$

为减少篇幅,其原始数据略。

2. 构造预报因子

先对35个初选因子作秩变换,即列出它们的秩数(若与 Y 正相关,则按从小到大的顺序;若与 Y 负相关,则按从大到小的顺序)。因子当前值的秩数,取与其 N 个历史数据中最接近的那个数值的秩数。在表(1)中,我们只列出因子 X_1, X_2, \dots, X_{25} 的秩数(包括当前值的秩数)。尔后以 M ($M < 35$)个因子的秩数之和构成一个秩和因子,并计算它与 $R(Y)$ 的相关系数,较大者(至少在信度 $\alpha = 0.001$ 下通过检验)选作为预报因子。

此例我们取 $M = 5$,为减少计算量,显然不便在 C_{35}^5 个秩和因子中,挑选与 $R(Y)$ 最相关者。这里我们按初选因子与 Y 的秩相关系数的大小,先在秩相关系数最大的10个因子中任取5个组合,而得 C_{10}^5 个秩和因子,从中选出与 $R(Y)$ 的相关系数最大者作为预报因子,记为 R_1 ,其数值为 $\nu_{11}, \nu_{12}, \dots, \nu_{1N}$;然后在其余的25个初选因子中取出秩相关系数最高的5个因子与上面剩余的5个因子再进行任意组合,从 C_{10}^5 个中再挑出与 $R(Y)$ 最相关者,记为 R_2 ,其数值为 $\nu_{21}, \nu_{22}, \dots, \nu_{2N}$;重复上述方法,我们共取了5个秩和因子 R_1, R_2, \dots, R_5 参加预报。经计算得知:

$$R_1 = X_1 + X_2 + X_5 + X_6 + X_8$$

$$R_2 = X_4 + X_7 + X_{10} + X_{11} + X_{13}$$

$$R_3 = X_3 + X_9 + X_{12} + X_{18} + X_{19}$$

$$R_4 = X_{15} + X_{16} + X_{17} + X_{21} + X_{23}$$

$$R_5 = X_{14} + X_{20} + X_{22} + X_{24} + X_{25}$$

它们的历史数值及当前值为:

$$\nu_{1j} = r_{1j} + r_{2j} + r_{5j} + r_{6j} + r_{8j},$$

$$\nu_{2j} = r_{4j} + r_{7j} + r_{10j} + r_{11j} + r_{13j},$$

$$\nu_{3j} = r_{3j} + r_{9j} + r_{12j} + r_{18j} + r_{19j},$$

$$\nu_{4j} = r_{15j} + r_{16j} + r_{17j} + r_{21j} + r_{23j},$$

$$\nu_{5j} = r_{14j} + r_{20j} + r_{22j} + r_{24j} + r_{25j},$$

其中 $j=0, 1, \dots, 28$, $v_{10}, v_{20}, \dots, v_{50}$ 为当前值, 具体数值列入表(2), 其与 $R(Y)$ 的相关系数分别为: 0.8228、0.7742、0.7579、0.7842、0.6777.

3. 分级

(1) 将预报对象 Y 之历史数值 y_1, y_2, \dots, y_N , 适当地分成 m 级, 记第 k 级

$Y^{(k)} = \{y_1^{(k)}, y_2^{(k)}, \dots, y_N^{(k)}\}$, $k=1, 2, \dots, m$; $\sum_{k=1}^m N_k = N$. 对应于 Y 的级别秩和因子

$R_i (i=1, 2, \dots, 5)$ 也相应地分成 m 级, 并记各级数值为: $v_{i1}^{(k)}, v_{i2}^{(k)}, \dots, v_{iN_k}^{(k)}$, $k=1, 2, \dots, m$. 此例取 $m=3$, 即将 Y 分成 3 级, 第一级 $y^{(1)}$: $Y \leq 100_{mm}$; 第二级 $y^{(2)}$, $100_{mm} < Y \leq 150_{mm}$; 第三级 $y^{(3)}$: $Y > 150_{mm}$. 各秩和因子也相应地分级 3 级, 分级后的数值列入表(3).

(2) 计算秩和因子各级均值. 记因子 R_i 第 k 级均值为:

$$a_{ik} = \frac{1}{N_k} \sum_{j=1}^{N_k} v_{ij}^{(k)}$$

此例各级均值如表 4.

表 4

级 别	R_1	R_2	R_3	R_4	R_5
1	47.20	51.15	53.80	55.55	54.00
2	70.69	70.75	70.81	67.81	67.19
3	99.25	95.25	92.55	93.20	95.25

4. 建立判别函数如 Y 的趋势预报

记判别函数为:

$$u(k) = \sum_{i=1}^5 |v_{i0} - a_{ik}|, \quad k=1, 2, \dots, m$$

若 $U(l) = \min_{1 \leq k \leq m} U(k)$, 则说明 $(v_{10}, v_{20}, \dots, v_{50})$ 与 $(a_{1l}, a_{2l}, \dots, a_{5l})$ 最接近, 即意味着秩和因子的当前值是属于第 l 级, 因此我们可预报 Y 的未来是属于第 l 级. 此例经计算得:

$$U(1) = 158.80, U(2) = 79.63, U(3) = 62.10$$

因 $U(3) = \min_{1 \leq k \leq 3} U(k)$, 所以预报 Y 的未来是属于第三级的, 即漳州 1984 年 4 月降雨量大于 150_{mm} .

5. 趋势预报的效果检验

本方法我们用回报来检验趋势预报的效果. 即将各组历史数值 $(v_{1j}, v_{2j}, \dots, v_{5j})$ ($j=1, 2, \dots, N$) 代入判别函数, 替换当前值 $(v_{10}, v_{20}, \dots, v_{50})$ 而回报 Y 的所属级别. 若回报准确率较高 (一般得在 80.00% 以上), 则说明趋势预报的效果较可靠, 接着可进行下一步的定值预报; 否则, 适当地调整预报因子或作其它处理.

此例经回报计算得知, 回报准确率 $p = \frac{24}{28} = 85.71\%$.

6. 用绝对距离相似法作 Y 的定值预报

此例经计算已判得 Y 的未来是属于第三级, 则计算因子之当前值 ($v_{10}, v_{20}, \dots, v_{50}$) 与因子属于第三级的各组数值 ($v_{1j}^{(3)}, v_{2j}^{(3)}, \dots, v_{5j}^{(3)}$), $j=1, 2, \dots, N_3$ 的绝对距离:

$$D(j) = \sum_{i=1}^5 |v_{i0} - v_{ij}^{(3)}|, \quad j=1, 2, \dots, N_3$$

其中 $N_3=10$, 计算结果列表 5.

表 5

j	1	2	3	4	5	6	7	8	9	10
D(j)	71.0	89.0	99.5	85.0	67.0	90.0	142.5	72.5	99.5	62.0

从表(5)中可知, 因子的当前值与第三级中的第五组和第十组的数值最接近, 而它们所对应的年号为1982和1973. 所以预报 $\hat{y}_{85} \approx \frac{y_{82} + y_{73}}{2} = \frac{186.0 + 329.2}{2} = 257.6(mm)$, 即预报漳州1984年4月的降雨量约为257.6mm. 实况为237.8mm, 基本相符.

参 考 文 献

- [1] 吴绍敏, 几种相似预报法, 福建师大学报(自然科学版), 1(1979).
- [2] 福建师大数学系统计数学教研组编, 统计预测(讲义), (1977).

A Method of Multivariate Distinction with the Rank Sum as the Factors

Wu Shaomin Peng Pei

Abstract

A method of distinction both with the rank sum as the factors and with the absolute distance as the distinctive function is presented in this paper. By using this method, it may be cast off the restriction of the model and can fully use the number less predictive information of factors and the computation will be simplified. It may be used both as tendency prediction and numerical prediction.

表1

年份	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅
1956	11.5	13.5	9	21.5	15	3	14.5	27.5	9	1	9	12	11	20.5	19	25.5	7.5	4	16	19	2	25	7	15	12
1957	4	6	12.5	8	1	2	7	6	4.5	3	7.5	3	16	6	2	9	9.5	7	21.5	9.5	4	2	15	23	25
1958	10	27	24.5	21.5	4	23	28	19	18.5	4	18.5	3	1	8.5	26	13	11	12	12.5	15	12	15	4	10	1
1959	20.5	20.5	22	10.5	27	5	14.5	20.5	14	26	10.5	7	18	15	7.5	15	12.5	1	23	8	23.5	21	28	28	13
1960	22	13.5	12.5	23.5	3	20	16	4.5	21.5	6	13	6	8	18.5	9.5	25.5	27	10	26	24.5	4	22.5	1	5	14
1961	27	15	5	28	23	6	24	17.5	27	19	20	5	10	28	27	20.5	19	16	28	6.5	28	20	22	17.5	
1962	13	22	24.5	16.5	16	22	9.5	27.5	21.5	20.5	13	28	19	20.5	15	25.5	22.5	8	20	27	18	26	8.5	20	8
1963	14	5	6	12	25	7	5	8	18.5	17.5	3	8.5	3	10	11	2	4.5	2.5	17	14	4	13	22	21	20.5
1964	3	1	1	1	14	1	1	3	1	7.5	2	8.5	7	4	9.5	8	28	5	18	21	20.5	19	6	7	6
1965	8	28	28	7	22	26	23	4.5	7	28	18.5	27	13	13.5	24	12	20.5	18	10.5	13	18	9	2	17	20.5
1966	24	12	7	10.5	10	8	21.5	25	25	5	25	3	9	7	21.5	5	2	14.5	12.5	3	12	8	8.5	4	10.5
1967	28	19	18.5	25	24	28	13	23	28	16	26	25	22	24.5	6	25.5	12.5	24	27	16	12	22.5	24.5	24	23
1968	1	4	4	3	12	9	3	1	2	15	1	10	12	5	5	17	17.5	9	2	23	1	5	12	19	2
1969	16.5	10.5	11	19	18	10	18	10	10	9.5	4.5	19	2	2	4	1	9.5	23	3	2	6.5	11	27	15	20.5
1970	20.5	7	16.5	14.5	8.5	15	12	12.5	20	9.5	6	24	5	13.5	12	11	7.5	23.5	7	11	12	1	10	6	4.5
1971	5	18	14.5	5.5	8.5	16.5	2	11	16.5	27	15.5	14.5	26	12	1	16	1	13	19	4.5	12	3	16	2.5	8
1972	8	2	2	4	2	27	9.5	12.5	8	2	23.5	12	16	18.5	13	18	3	11	8	17	25	4	5	2.5	28
1973	19	23	20.5	14.5	7	21	5	26	11.5	20.5	17	26	27	1	28	22	6	28	15	20	26.5	12	17	15	25
1974	11.5	16.5	18.5	18	20.5	13.5	19	23	13	12	21.5	23	6	26	23	4	15	19.5	25	9.5	16	14	14	9	10.5
1975	8	3	3	5.5	11	4	9.5	14	4.5	17.5	10.5	1	14	16	3	19	4.5	26	5	4.5	15	24	3	1	15.5
1976	25	25	27	27	6	25	26.5	17.5	16.5	23	13	19	20	24.5	19	3	20.5	25	10.5	18	26.5	10	18	8	25
1977	2	8.5	9	2	5	11	5	8	6	13	27	21	24	8.5	19	14	15	17	1	6.5	22	17	21	18	8
1978	6	10.5	20.5	9	19	18.5	20	8	3	7.5	4.5	19	23	3	7.5	6	25	22	14	1	18	27	13	25.5	4.5
1979	16.5	8.5	9	20	13	12	17	16	24	14	7.5	12	28	23	16	10	15	19.5	4	12	23.5	7	19	11	3
1980	23	26	26	26	28	24	26.5	2	11.5	11	21.5	16	16	17	25	25.5	22.5	27	6	26	8.5	17	24.5	13	15.5
1981	16.5	24	23	16.5	17	16.5	9.5	20.5	6	23	25	17	4	27	21.5	25.5	24	6	9	22	8.5	20	11	12	20.5
1982	16.5	16.5	14.5	23.5	20.5	13.5	25	15	15	22	15.5	22	21	22	14	7	26	14.5	21.5	28	20.5	6	23	25.5	27
1983	26	20.5	16.5	13	26	18.5	21.5	23	26	24	23.5	14.5	25	11	17	20.5	17.5	21	24	24.5	6.5	17	26	27	17.5
当前值	4	25	27	8	6.5	24	16	8	7	28	23.5	25.5	23	16	25.5	25.5	12.5	17	14	28	15	4	15	7	15.5

表3

年份	Y	R ₁	R ₂	R ₃	R ₄	R ₅	回报所 属级别
1968	47.5	27.0	34.0	27.0	52.5	54.0	1
1963	64.8	59.0	40.5	52.5	43.5	78.5	1
1957	69.2	19.0	41.5	48.5	39.5	65.5	1
1977	79.8	34.5	71.0	54.0	91.0	58.0	1
1958	83.8	83.0	73.0	70.5	66.0	49.5	2
1971	90.1	59.0	76.0	76.5	46.0	30.0	2
1975	90.7	40.0	57.0	39.5	44.5	61.0	1
1964	91.7	22.0	18.5	33.5	72.0	57.0	1
1969	92.6	65.0	53.0	66.0	48.0	50.5	1
1970	94.1	63.5	47.0	70.0	52.5	36.0	1
1966	107.5	79.0	71.0	63.0	49.0	32.5	2
1956	108.0	70.5	57.0	50.0	61.0	91.5	2
1972	108.7	51.5	55.0	41.0	64.0	70.0	1
1974	114.3	85.0	76.5	99.0	72.0	69.0	2
1960	115.6	63.0	66.5	76.0	67.0	84.5	2
1979	129.8	66.0	86.5	68.5	83.5	56.0	2
1965	136.2	88.5	89.5	90.5	76.5	73.0	3
1978	144.2	62.0	64.0	78.5	69.5	61.0	2
1976	152.4	98.5	109.5	98.0	87.0	85.5	3
1981	156.7	94.5	83.0	78.0	90.5	101.5	3
1962	156.8	100.5	78.5	102.0	89.5	101.5	3
1961	167.0	88.5	101.0	81.0	114.5	102.0	3
1982	186.0	82.0	107.0	87.5	90.5	108.5	3
1959	194.9	93.5	79.5	67.0	86.5	85.5	3
1967	222.6	122.0	102.0	122.5	80.5	110.0	3
1980	257.1	103.0	101.0	86.5	106.0	88.5	3
1981	156.7	94.5	83.0	78.0	90.5	101.5	20
1982	186.0	82.0	107.0	87.5	90.5	108.5	23
1983	274.1	114.0	107.0	102.0	87.5	97.0	3
1982	329.2	96.0	84.0	101.0	99.5	73.0	3
当前值		67.5	98.5	90.5	93.5	70.5	

表2

年份	Y	R ₁	R ₂	R ₃	R ₄	R ₅	R(Y)
1956	108.0	70.5	57.0	50.0	61.0	91.5	12
1957	69.2	19.0	41.5	48.5	39.5	65.5	3
1958	83.8	83.0	73.0	70.5	66.0	49.5	5
1959	194.9	93.5	79.5	67.0	86.5	85.0	24
1960	115.6	63.0	66.5	76.0	67.0	84.5	15
1961	167.0	88.5	101.5	81.0	114.5	102.0	22
1962	156.8	100.5	78.5	102.0	89.5	101.5	21
1963	64.8	59.0	40.5	52.5	43.5	78.5	2
1964	91.7	22.0	18.5	33.5	72.0	57.0	8
1965	136.2	88.5	89.5	90.5	76.5	73.0	17
1966	107.5	79.0	71.0	63.0	49.0	32.5	11
1967	222.6	122.0	102.0	122.5	80.5	110.0	25
1968	47.5	27.0	34.0	27.0	52.5	54.0	1
1969	92.6	65.0	53.0	66.0	48.0	50.5	9
1970	94.1	63.5	47.0	70.0	52.5	36.0	10
1971	90.1	59.0	76.0	76.5	46.0	30.0	6
1972	108.7	51.5	55.0	41.0	64.0	70.0	13
1973	329.2	96.0	84.0	101.0	99.5	73.0	28
1974	114.3	85.0	76.5	99.0	72.0	69.0	14
1975	90.7	40.0	57.0	39.5	44.5	61.0	7
1976	152.4	98.5	109.5	98.0	87.0	85.5	19
1977	79.8	34.5	71.0	54.0	91.0	58.0	4
1978	144.2	62.0	64.0	78.5	69.5	61.0	18
1979	129.8	66.0	86.5	68.5	83.5	56.0	16
1980	257.1	103.0	101.0	86.5	106.0	88.5	26
1981	156.7	94.5	83.0	78.0	90.5	101.5	20
1982	186.0	82.0	107.0	87.5	90.5	108.5	23
1983	274.1	114.0	107.0	102.0	87.5	97.0	27
当前值		67.5	98.5	90.5	93.5	70.5	