应用气象统计学实验

任传友 王艳华 编写

沈阳农业大学应用气象教研室

2008 年 7 月

目　　录

实验一 平均值的假设检验

一．实验目的

掌握平均值的假设检验方法。

二．实验内容

根据沈阳市和本溪市的年平均气温序列，分析两地的年平均气温是否有显著差异。

三．原理

气候分析中经常遇到比较两地气候状况是否有显著差异的问题，这时首先考虑的当然是各种气象要素的平均值是否相等，这可以通过平均值的假设检验方法解决。

设 ξ_1，ξ_2 为二独立正态总体，数学期望分别为 m_1，m_2；方差分别为 σ_1^2，σ_2^2。今从 ξ_1，ξ_2 中抽取容量各为 n_1，n_2 的样本，要检验原假设 H_0："$m_1 = m_2$"。

我们把两个样本的平均值分别记为 $\bar{X}_1$，$\bar{X}_2$，由于它们均为正态分布，且相互独立，因此，$Y = \bar{X}_1 - \bar{X}_2$ 也是正态分布，其数学期望为

$$E(Y) = E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = m_1 - m_2 \tag{1.1}$$

方差为

$$D(Y) = D(\bar{X}_1 - \bar{X}_2) = D(\bar{X}_1) + D(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{1.2}$$

因此，若 H_0 成立，则统计量

$$u = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{1.3}$$

是服从 $N(0, 1)$的变量，于是可用于检验平均值相等的假设。

这样，当指定 α 时，由正态分布表可查到 u_α 值。然后将 $\bar{X}_1$、$\bar{X}_2$、σ_1、σ_2、n_1、n_2 代入(1.3)式，可算得一个 $u_{实}$，按实际推断原理，若 $|u_{实}| < u_\alpha$，则接受原假设；若 $|u_{实}| \geqslant u_\alpha$，则拒绝原假设。

四．讨论

讨论应用上述方法进行平均值的假设检验时应注意什么问题？

五．附资料(单位：℃)

年	1951	1952	1953	1954	1955	1956	1957	1958	1959
沈阳	7.71	7.19	7.37	6.82	7.83	6.22	6.47	7.94	8.54
本溪						5.99	6.77	8.22	8.64
年	1960	1961	1962	1963	1964	1965	1966	1967	1968
沈阳	7.98	8.83	7.9	8.05	7.67	7.91	7.63	7.95	8.12
本溪	8.16	8.91	8.49	8.41	8.03	8.11	7.83	7.96	8.05
年	1969	1970	1971	1972	1973	1974	1975	1976	1977
沈阳	6.88	7.68	7.87	7.84	8.48	7.86	9.03	7.61	8.03
本溪	7.01	7.92	7.79	7.78	8.07	7.32	8.24	7.19	7.64
年	1978	1979	1980	1981	1982	1983	1984	1985	1986
沈阳	8.19	8.66	7.74	8.31	9.16	9.13	8.22	8.06	8.19
本溪	7.61	8.23	6.87	7.32	8.29	8.19	7.21	6.97	6.95
年	1987	1988	1989	1990	1991	1992	1993	1994	1995
沈阳	8.44	9.14	8.75	8.54	7.99	8.12	8.05	9.07	8.65
本溪	7.57	7.82	8.71	8.47	7.87	8.05	7.83	8.67	8.13
年	1996	1997	1998	1999	2000	2001			
沈阳	8.12	8.95	9.77	9.05	8.35	8.45			
本溪	7.8	8.43	9.33	8.49	7.63	7.73			

实验二 方差和相关系数的假设检验

一．实验目的

掌握方差和相关系数的假设检验方法。

二．实验内容

根据沈阳市和本溪市的年平均气温序列，分析：

1. 两地的年平均气温的方差是否有显著差异。
2. 两地从 1956～2001 年的年平均气温是否显著相关。

三．原理

1．方差的假设检验

设 s_1^2 及 s_2^2（或 s_1^{*2}，s_2^{*2}）为分别抽自相互独立的正态总体 $N(m_1,\sigma_1)$ 及 $N(m_2,\sigma_2)$ 的样本方差，当 $\sigma_1=\sigma_2=\sigma$ 时，统计量

$$F=\frac{s_1^{*2}}{s_2^{*2}}=\frac{\dfrac{n_1}{n_1-1}s_1^2}{\dfrac{n_2}{n_2-1}s_2^2} \tag{2.1}$$

服从自由度 $v_1=n_1-1$，$v_2=n_2-1$ 的 F 分布。因此，可用它来检验假设 $\sigma_1=\sigma_2=\sigma$。这一方法称为 F 检验法。

当用(2.1)式检验假设 $\sigma_1=\sigma_2=\sigma$ 时，显然，若假设成立，则 s_1^{*2} 及 s_2^{*2} 为从方差相等的总体中抽出的样本方差，它们当然不应差别很大，所以式中的 F 不应太大，也不应太小。因此，对于指定的信度 α，接受或拒绝假设的临界值应由下式确定。

$$\begin{cases} P\left(F\le f_{\alpha_1}\right)=\dfrac{\alpha}{2} \\ P\left(F\ge f_{\alpha_2}\right)=\dfrac{\alpha}{2} \end{cases} \tag{2.2}$$

图 2.1 即为 F 检验法的接受区和拒绝区图示。图中（f_{α_1}，f_{α_2}）为接受区，（0，f_{α_1}）及（f_{α_2}，∞）为拒绝区。f_{α_1} 及 f_{α_2} 可由 F 分布表查得。由于 F 分布是按关系式

$$P\left(F\ge f_P\right)=\int_{f_P}^{\infty}g\left(f\right)df=P \tag{2.3}$$

制定的，从表上只能查到(2.2)式中的 f_{α_2}，而不能查到 f_{α_1}，但由于

$$P\left(F \le f_{\alpha_1}\right) = P\left(\frac{1}{F} \ge \frac{1}{f_{\alpha_1}}\right) = \frac{\alpha}{2} \tag{2.4}$$

因此，可先以 $v_1 = n_1 - 1$， $v_2 = n_2 - 1$ 查到 f_{α_2}， 再以 $v_1 = n_2 - 1$，$v_2 = n_1 - 1$ 查到 f'_{α_1}，则由 $f'_{\alpha_1} = \dfrac{1}{f_{\alpha_1}}$ 可得 $f_{\alpha_1} = \dfrac{1}{f'_{\alpha_1}}$。

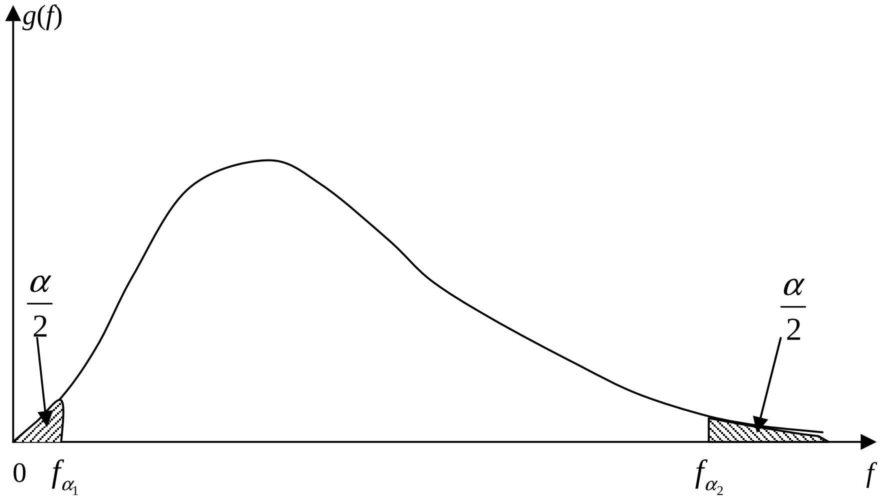

图 2.1 *F* 检验法的接受区和拒绝区图示

2．相关系数的假设检验

从不相关正态总体中抽取的样本相关系数 R 具有密度

$$f(r) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left(1 - r^2\right)^{\frac{n-4}{2}} \tag{2.5}$$

引进下列变换

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2.6}$$

则(2.6)式中 t 是自由度 $v = n-2$ 的 t 分布变量，因此，可用 t 检验法检验 $\rho = 0$ 的假设。

四．讨论

论述方差和相关系数的假设检验在气象分析中的重要作用。

五．附资料(单位：℃)

见实验一资料。

实验三 多元线性回归分析

一．实验目的

学会应用给定资料建立多元线性回归方程，能够定性评价回归方程的适用性。

二．实验内容

根据辽宁省 54 站的多年平均温度(T)资料，建立辽宁省春季(以 4 月份为代表)、夏季(以 7 月份为代表)、秋季(以 10 月份为代表)、冬季(以 1 月份为代表)或年平均温度依经度(λ)、纬度(φ)和海拔高度(h)的三元线性回归方程。从以上五个时段的温度预报中任选一个完成实验。

三．原理

设 x_1，x_2，x_3，…，x_m 为 m 个非随机的自变量，因变量η为随机变量，则

$$\mathrm{E}\left(\eta \mid x_1, x_2, ..., x_m\right) = g\left(x_1, x_2, ..., x_m\right) \tag{3.1}$$

而η得可能取值与 x_1，x_2，x_3，…，x_m 的关系表示为：

$$y = g\left(x_1, x_2, ..., x_m\right) + \varepsilon \tag{3.2}$$

对任意 x_1，x_2，x_3，…，x_m，式中ε为相互独立且数学期望为 0，方差为$\mathrm{D}(\eta)$的正态随机变量，而

$$y = g\left(x_1, x_2, ..., x_m\right) \tag{3.3}$$

称作η依 x_1，x_2，x_3，…，x_m 的多元回归方程，并用它作为 x_1，x_2，x_3，…，x_m 取各种特定值时η的可能取值 y 的估计值。作为多元回归的一种特例，若有，

$$y = g\left(x_1, x_2, ..., x_m\right) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m \tag{3.4}$$

则称(3.4)式为η依 x_1，x_2，x_3，…，x_m 的多元线性回归方程，它表示多维空间中的一个平面(m=2)或超平面(m≥3)。对于多元回归方程有

$$\mathrm{E}\left\{\left[\eta - g\left(x_1, x_2, ..., x_m\right)\right]^2\right\} = \min \tag{3.5}$$

另外，在气候分析中还常常需要分析多个自变量和因变量都是

随机变量的统计联系，例如同一测站或不同测站几个气象要素之间的相互联系就属于这一类。

设已知随机变量η和$\xi_1,\xi_2,...,\xi_m$的几组一一对应的观测值，求η和$\xi_1,\xi_2,...,\xi_m$之间统计联系的数学形式

$$\eta = g\left(\xi_1,\xi_2,...,\xi_m\right) \tag{3.6}$$

这类问题称作相关分析。

由于η和$\xi_1,\xi_2,...,\xi_m$都是随机变量，很自然地，可以要求函数$g\left(\xi_1,\xi_2,...,\xi_m\right)$满足

$$\mathrm{E}\left\{\left[\eta - g\left(\xi_1,\xi_2,...,\xi_m\right)\right]^2\right\} = \min \tag{3.7}$$

可以证明，当

$$g\left(x_1,x_2,...,x_m\right) = \mathrm{E}\left(\eta \mid \xi_1 = x_1, \xi_2 = x_2, ..., \xi_m = x_m\right) \tag{3.8}$$

时，(3.7)式即能满足。我们把(3.8)式叫做η依$\xi_1,\xi_2,...,\xi_m$的回归方程，为了和自变量为非随机变量的回归方程区别起见，把(3.4) 式叫做第一类回归方程，而把(3.8)式叫做第二类回归方程。

数理统计中证明，若$\left(\eta,\xi_1,\xi_2,...,\xi_m\right)$服从联合正态分布，则$g\left(\xi_1,\xi_2,...,\xi_m\right)$具有线性形式，即

$$y = g\left(\xi_1,\xi_2,...,\xi_m\right) = b_0 + b_1\xi_1 + b_2\xi_2 + \cdots + b_m\xi_m \tag{3.9}$$

许多实际问题中的变量都是近似地满足联合正态分布的，因此线性回归的应用非常广泛。即使在$\left(\eta,\xi_1,\xi_2,...,\xi_m\right)$的分布未知或不是正态分布时，由于线性回归方程计算简单，也常常以线性模型作为近似。

在第一类回归和第二类回归中，回归曲面满足的关系式(3.5)和 (3.7)具有完全相同的形式，因此，尽管在理论上回归分析和相关分析解决问题的性质不同，但在实际工作中并不去仔细地区分它们，而是把所有研究变量之间统计联系的数学形式问题笼统地称为回归分析。把因变量和自变量之间的关系统一写成下式的形式

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_m x_m \tag{3.10}$$

待定系数$\hat{b}_1$，$\hat{b}_2$，…，$\hat{b}_m$可通过回归方程的正规方程组或标准化正规方程组通过高斯-约当消去法或求解求逆同时变化的方法进行求解。

回归方程的正规方程组为

$$\begin{cases} \hat{b}_1 s_{11} + \hat{b}_2 s_{12} + \cdots + \hat{b}_m s_{1m} = s_{1y} \\ \hat{b}_1 s_{21} + \hat{b}_2 s_{22} + \cdots + \hat{b}_m s_{2m} = s_{2y} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{b}_1 s_{m1} + \hat{b}_2 s_{m2} + \cdots + \hat{b}_m s_{mm} = s_{my} \end{cases} \tag{3.11}$$

标准化正规方程组为

$$\begin{cases} \hat{b}'_1 r_{11} + \hat{b}'_2 r_{12} + \cdots + \hat{b}'_m r_{1m} = r_{1y} \\ \hat{b}'_1 r_{21} + \hat{b}'_2 r_{22} + \cdots + \hat{b}'_m r_{2m} = r_{2y} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{b}'_1 r_{m1} + \hat{b}'_2 r_{m2} + \cdots + \hat{b}'_m r_{mm} = r_{my} \end{cases} \tag{3.12}$$

通过下式对回归方程的回归效果作假设检验

$$F = \frac{\frac{U}{m}}{\frac{Q}{n-m-1}} \tag{3.13}$$

U 和 Q 分别为回归方程的回归平方和和残差平方和。给定信度α，假设$E(b_1)=E(b_2)=\cdots=E(b_m)=0$，查表求出满足

$$P(F \geq F_\alpha) = \alpha \tag{3.14}$$

时 F 的置信限F_α，即可拒绝原假设，认为自变量全体与因变量的线性联系是显著的。

四．讨论

讨论如何才能使所建立的回归方程更稳定。

五．附资料

站名	海拔高度	纬度	经度	1 月	4 月	7 月	10 月	年
	(m)	(°)			(0.1℃)			
沈阳	42.8	41.73	123.45	-115	98	245	95	81

辽中	12.2	41.5	122.72	-111	93	243	98	81
新民	30.7	41.98	122.83	-117	91	242	93	77
康平	118.6	42.75	123.33	-129	87	237	87	70
法库	97.8	42.5	123.4	-131	86	236	84	68
大连	91.5	38.9	121.63	-46	95	230	138	104
复县	118.5	39.63	122.02	-77	97	239	116	94
皮口	43.2	39.42	122.37	-72	86	228	121	91
长海	35.5	39.27	122.58	-48	80	220	143	99
庄河	34.8	39.72	122.95	-79	88	230	114	89
旅顺	66.7	38.82	121.23	-44	95	234	134	104
丹东	15.1	40.47	124.33	-80	86	229	111	87
宽甸	260.1	40.72	124.78	-128	79	223	88	67
凤城	72.6	40.05	124.07	-106	89	232	99	79
东沟	4.2	39.83	124.27	-84	85	229	110	86
辽阳	24.4	41.23	123.17	-111	103	246	101	85
锦州	65.9	41.13	121.12	-85	101	241	111	92
黑山	37.5	41.68	122.08	-106	92	239	98	81
义县	39.3	41.68	121.23	-108	95	238	94	79
北镇	68	41.58	121.77	-102	96	238	100	83
锦县	58.2	41.32	121.37	-93	92	240	106	86
锦西	23.9	40.77	120.83	-87	98	240	111	90
建昌	365.5	40.8	119.82	-97	104	233	96	82
绥中	15.3	40.35	120.35	-80	99	240	113	93
兴城	8.8	40.58	120.7	-84	93	237	110	89
抚顺	118.5	41.92	124.08	-141	87	235	82	66
清原	234.1	42.1	124.92	-155	77	228	69	54
新宾	328.4	41.73	125.05	-166	72	222	64	48
阜新	46.8	42.08	121.72	-112	94	242	90	78
彰武	79.4	42.42	122.53	-123	87	240	87	72
铁岭	85.4	42.3	123.87	-126	95	243	94	76
开原	98.2	42.53	124.05	-141	87	236	83	66

昌图	165.1	42.78	124.12	-135	85	235	86	67
西丰	196.1	42.73	124.75	-172	76	229	66	50
营口	3.3	40.67	122.27	-91	99	248	112	91
熊岳	20.4	40.17	122.15	-90	104	244	110	91
盖县	24.8	40.42	122.35	-81	110	248	114	97
营县	3	40.67	122.48	-96	100	244	105	88
朝阳	169.9	41.55	120.45	-105	111	245	98	86
建平	661.8	41.87	119.63	-140	78	221	68	55
北票	176.3	41.82	120.75	-105	106	245	96	84
叶寿	421.7	41.38	119.7	-104	104	239	92	81
凌源	417.4	41.23	119.35	-105	106	239	93	81
喀左	296.8	41.08	119.17	-104	110	241	96	84
鞍山	77.3	41.08	123	-96	105	247	108	91
台安	7.4	41.38	122.43	-105	93	243	101	83
岫岩	79.8	40.28	123.28	-106	86	230	94	76
海城	25.3	40.88	122.72	-109	101	245	102	85
本溪	185.27	41.32	123.78	-119	97	240	94	78
溪县	209.9	41.3	124.12	-139	85	231	83	65
桓仁	240.3	41.27	125.35	-135	86	228	83	66
草河口	233.4	40.88	123.9	-123	73	220	78	62
盘山	3.8	41.18	122.02	-100	92	243	104	85
大洼	3.9	40.98	122.07	-97	93	245	108	87

实验四 逐步回归分析

一．实验目的

了解回归方程中各自变量对回归方程的方差贡献的评价方法，掌握建立最优回归方程的方法。

二．实验内容

根据辽宁省54站的多年春季(以4月份为代表)平均温度资料，建立4月份平均温度依经度(λ)、纬度(φ)和海拔高度(h)的逐步回归方程。

三．逐步回归的基本思路

在多元回归分析中，我们将选定的若干个自变量全部引进方程，然后通过适当的假设检验，确定各个自变量对依变量的方差贡献是否显著，把对依变量方差贡献不显著的自变量剔除，重新建立回归方程。由于各个自变量之间彼此相关，经剔除以后保留下来的各自变量的回归系数还得重算，计算工作相当繁重。为了尽可能简便地从若干个自变量中选取一些对依变量方差贡献显著的变量建立回归方程，常常采用一种求解正规方程与检验自变量回归效果显著性同时进行的逐步回归分析方法。

逐步回归的基本思路是，根据各个自变量方差贡献的大小，每次引入一个在所有尚未进入方程的自变量中方差贡献最大而且达到一定显著水平的自变量建立回归方程；同时计算引进新变量后在原方程中的自变量对依变量的方差贡献，把那些由于新变量的进入而对依变量的方差贡献变得不显著的变量剔除掉，建立新的回归方程，这样逐步引入新的方差贡献显著的自变量，逐步剔除不显著的自变量，从而保证方程中始终只保留对依变量方差贡献显著的自变量。这种“筛选”过程一直进行到所有可供选择的变量中再也没有对依变量方差贡献显著的变量可以引进，也没有对方差贡献不显著的变量需要剔除为止。显然，逐步回归建立的回归方程就是一般的回归方程，不过，它的自变量都已经过检验。因此，逐步回归分析建立的回归方程被称为“最优”回归方程。

四．逐步回归的计算步骤

（1）计算自变量和依变量之间、自变量相互之间的相关系数矩阵

$$R^{(0)}=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} & r_{1,m+1} \\ r_{21} & r_{22} & \cdots & r_{2m} & r_{2,m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} & r_{m,m+1} \\ r_{m+1,1} & r_{m+2,2} & \cdots & r_{m+1,m} & r_{m+1,m+1} \end{pmatrix} \tag{4.1}$$

（2）规定引进变量的 F 检验置信限 F_1^*，在所有对依变量方差贡献达显著标准的自变量中，选择对依变量方差贡献最大者引进方程。其具体计算步骤为：

（a）计算各自变量 x_i 在一元回归中的方差贡献

$$V_i^{(0)}=\left(r_{i,m+1}^{(0)}\right)^2/r_{ii}^{(0)} \tag{4.2}$$

找出方差贡献最大者

$$V_{\max}^{(0)}=V_{k_1}^{(0)} \tag{4.3}$$

式中 k_1 为方差贡献最大的自变量序号。

（b）计算统计量

$$F_{1,k_1}=\frac{V_{k_1}^{(0)}\left(n-2\right)}{r_{m+1,m+1}^{(0)}-V_{k_1}^{(0)}} \tag{4.4}$$

检验 x_{k_1} 的方差贡献是否显著。若 $F_{1,k_1}<F_1^*$，则没有变量可引进方程。

（c）若 $F_{1,k_1}\ge F_1^*$， x_{k_1} 可引进方程，则对相关矩阵(4.1)进行变换

$$r_{ij}^{(1)}=\begin{cases} 1/r_{k_1k_1}^{(0)} & i=j=k_1 \\ r_{k_1j}^{(0)}/r_{k_1k_1}^{(0)} & i=k_1,\ j\ne k_1 \\ -r_{ik_1}^{(0)}/r_{k_1k_1}^{(0)} & i\ne k_1,\ j=k_1 \\ r_{ij}^{(0)}-r_{ik_1}^{(0)}r_{k_1j}^{(0)}/r_{k_1k_1}^{(0)} & i,j\ne k_1 \end{cases} \tag{4.5}$$

求出第一步矩阵 $R^{(1)}=\left(r_{ij}^{(1)}\right)$。

（3）引进一个变量后，在余下的变量中继续选择对依变量方

差贡献最大，并达到显著标准的引进方程。

（a）计算除 x_{k_1} 外，其它自变量在二元回归中的方差贡献

$$V_i^{(1)} = \left(r_{i,m+1}^{(1)}\right)^2 / r_{ii}^{(1)} \tag{4.6}$$

找出方差贡献最大者

$$V_{\max}^{(1)} = V_{k_2}^{(1)} \tag{4.7}$$

（b）计算统计量

$$F_{1,k_2} = \frac{V_{k_2}^{(1)}(n-3)}{r_{m+1,m+1}^{(1)} - V_{k_2}^{(1)}} \tag{4.8}$$

检验 x_{k_2} 的方差贡献是否显著。若 $F_{1,k_2} < F_1^*$，则建立一元线性回归方程。

（c）若 $F_{1,k_2} \ge F_1^*$， x_{k_2} 可引进方程，则对相关矩阵 $R^{(1)}$ 进行变换， $R^{(1)} \to R^{(2)}$ 。

（4）重复步骤（3），继续引进第三个变量 x_{k_3}， $R^{(2)} \to R^{(3)}$ 。

（5）检验方程中引进 x_{k_3} 后 x_{k_1} 的方差贡献是否仍旧显著，如果 x_{k_1} 的方差贡献已不显著，则对矩阵作剔除 x_{k_1} 的变换。反之，则继续在余下的因子中选择对依变量方差贡献最大并达到显著标准的变量引入方程。具体计算步骤为：

（a）计算方程中引进 x_{k_3} 后 x_{k_1} 的方差贡献

$$V_{k_1}^{(3)} = \left(r_{k_1,m+1}^{(3)}\right)^2 / r_{k_1k_1}^{(3)} \tag{4.9}$$

并计算统计量

$$F_{2,k_1} = \frac{V_{k_1}^{(3)}(n-4)}{r_{m+1,m+1}^{(3)}} \tag{4.10}$$

指定剔除变量的显著性检验标准 F_2^*，检验 x_{k_1} 的方差贡献是否仍旧显著。

（b）若 $F_{2,k_1} < F_2^*$，则对相关矩阵 $R^{(3)}$ 作剔除 x_{k_1} 的变换，求出第四步矩阵 $R^{(4)} = \left(r_{ij}^{(4)}\right)$。

（c）若 $F_{2,k_1} \geq F_2^*$，则 x_{k_1} 在三元回归方程中的方差贡献仍旧显著，无需剔除，可继续引进除 x_{k_1}、x_{k_2}、x_{k_3} 以外的其它变量。

（6）一般地，若已进行 p 步矩阵变换，在回归方程中引进了 l 个自变量，则 p＋1 步运算可分三种情况不同处理。

（a）检验有无方差不显著的自变量留在方程中，如果有的话把它从方程中剔除。

（b）若无需从 l 个自变量中剔除任何变量，则继续选择方差贡献最大并达显著标准的变量引进方程。

（c）若所有已引进方程的自变量的方差贡献都是显著的，而所有未引进方程的变量都不在能够引进方程，则逐步回归分析到此为止，得到一个 l 元回归方程。若即引进方程的自变量为 $k_j\left(j=1,2,\cdots,l\right)$，$l$ 元回归方程的系数从 p 步矩阵元素中求出

$$\hat{b}_i = \frac{\sqrt{s_{yy}}}{\sqrt{s_{ii}}} r_{i,m+1}^{(p)} \qquad \left(r = k_j,\ j = 1,2,\cdots,l\right) \tag{4.11}$$

由此可得回归方程

$$\hat{y} = \overline{y} + \sum_{j=1}^{l} \frac{\sqrt{s_{yy}}}{\sqrt{s_{k_j k_j}}} r_{k_j,m+1}^{(p)} \left(x_{k_j} - \overline{x}_{k_j}\right) \tag{4.12}$$

复相关系数和剩余方差也可从 p 步矩阵元素中得到

$$R = \sqrt{1 - r_{m+1,m+1}^{(p)}} \tag{4.13}$$

$$s_r^2 = \frac{s_{yy} r_{m+1,m+1}^{(p)}}{n-l-1} \tag{4.14}$$

五．讨论

阐述在进行回归分析时应注意一些什么问题。

六．附资料（见实验三资料）

实验五 Fisher 判别分析

一．实验目的

掌握 Fisher 二级判别分析方法。

二．实验内容

根据华北地区和长江中下游降水年变化的不同特点，根据给定资料，建立新增测站分属何种降水类型的判别方程。并判别青岛、兖州、临沂、徐州、阜阳等中间地带的测站应分属于何种降水类型。

三．原理

为了使判别尽可能少出错误，应要求判别对象 y 的条件分布集中位置差别尽可能大，离散程度尽可能小，这样的函数叫做判别函数。最简单的判别函数是 $x_1, x_2, \ldots\ldots, x_p$ 的诸判别因子的线性组合，即

$$y = c_1x_1 + c_2x_2 + \ldots\ldots + c_px_p \tag{5.1}$$

衡量条件分布集中位置的差异大小可以用集中位置的距离表示，而衡量集中程度的量是方差。因此，类间的方差与类内的方差的比值为最大可作为判别方程建立的原则，这就是 Fisher 判别准则。即要求下面的函数最大

$$\lambda = \frac{\left(\overline{y_1} - \overline{y_2}\right)^2}{\sum_{i=1}^{n_1}\left(y_{1i} - \overline{y_1}\right)^2 + \sum_{i=1}^{n_2}\left(y_{2i} - \overline{y_2}\right)^2} \to \max \tag{5.2}$$

式中 $\overline{y_1}$ 与 $\overline{y_2}$ 为判别对象在两类中的平均值。式中判别函数的平方和具有方差的意义。

求判别系数的标准方程组为

$$\begin{cases} w_{11}c_1 + w_{12}c_2 + \cdots + w_{1p}c_p = d_1 \\ w_{21}c_1 + w_{22}c_2 + \cdots + w_{2p}c_p = d_2 \\ \cdots\cdots\cdots\cdots\cdots \\ w_{p1}c_1 + w_{p2}c_2 + \cdots + w_{pp}c_p = d_p \end{cases} \tag{5.3}$$

式中 $w_{kl}=\sum_{i=1}^{n_1}\left(x_{k1i}-\overline{x}_{k1}\right)\left(x_{l1i}-\overline{x}_{l1}\right)+\sum_{i=1}^{n_2}\left(x_{k2i}-\overline{x}_{k2}\right)\left(x_{l2i}-\overline{x}_{l2}\right)$，为不同因子 k 与 l 在两类内离差交叉积和；$d_k=\overline{x}_{k1}-\overline{x}_{k2}$，为第 k 个因子在不同类别内的平均值之差。

判别时，临界 y_{c} 值可由两类的重心得到，即

$$y_c=\frac{1}{n_1+n_2}\left(n_1\overline{y}_1+n_2\overline{y}_2\right)=\frac{1}{n_1+n_2}\sum_{k=1}^{2}\left(n_1c_k\overline{x}_{k1}+n_2c_k\overline{x}_{k2}\right) \qquad (5.4)$$

如果 $\overline{y}_1>\overline{y}_2$，当 $y_0>y_c$ 时，预报 y_0 属于 1 类，$y_0<y_c$ 时报 2 类；反之，如果 $\overline{y}_1<\overline{y}_2$，当 $y_0>y_c$ 时，预报 y_0 属于 2 类，$y_0<y_c$ 时报 1 类。

四．讨论

在讨论多个变量之间的统计联系时，什么情况下会用到回归分析，什么情况下会用到判别分析？

五．附资料

地区	测站	x_1	x_2	x_3
华北地区	天津	2.45	8.9	12.1
	北京	3.46	9.7	14.3
	保定	3.26	9.0	12.5
	石家庄	3.39	8.5	13.0
	太原	2.13	10.6	13.3
	五台山	1.80	16.4	18.1
	兴县	3.01	10.1	13.3
	榆林	3.70	7.8	12.5
	张家口	1.82	11.4	12.7
	大同	2.05	11.6	12.7
长江中下游地区	上海	0.74	13.1	10.0
	东山	1.01	12.5	11.7
	南京	0.87	10.9	11.5
	合肥	1.18	10.3	10.1
	安庆	0.44	12.3	9.5
	九江	0.47	13.6	9.4
	汉口	0.61	11.7	8.5
	芜湖	0.76	10.5	10.9
	溧阳	0.75	11.3	12.2

	黄石	0.64	14.0	10.4
判别对象	青岛	1.68	13.7	11.6
	兖州	1.75	13.7	10.5
	临沂	1.65	10.0	12.0
	徐州	1.48	8.3	11.1
	阜阳	1.07	8.6	10.9

注：x_1 为八月降水量与六月降水量之比；x_2 为六月降水日数；x_3 为八月降水日数。

实验六 系统聚类分析

一．实验目的

掌握系统聚类分析方法。

二．实验内容

根据辽宁省 15 站多年一月平均气温序列，应用最短距离法对其进行系统聚类。

三．系统聚类的基本思路

系统聚类法是目前使用最多的一种聚类方法，其基本思路是先将 n 个样品各自成一类，然后规定样品之间距离的定义和类和类之间距离的定义，选择距离最近的一对并为一类，计算新类与其它类的距离，再将距离最近的两类合并。这样每次缩小一类，直至所有样品都成为一类为止。

类与类之间的距离有很多种定义方法。例如，可以定义为类与类之间的距离为属于两类样品之间的最近距离，也可以定义为属于两类样品之间的最长距离，或者定义为两类中心之间的距离等等。不同的类间距离定义，就产生了不同的系统聚类法。

四．附资料(单位：℃)

彰武	营口	岫岩	熊岳	沈阳
-12.25	-9.13	-10.52	-9.03	-11.77
清源	宽甸	开原	桓仁	复县
-15.39	-12.38	-14.18	-13.52	-7.48
阜新	抚顺	丹东	本溪	鞍山
-11.33	-13.82	-8.04	-11.89	-9.55

实验七 谐波分析

一．实验目的

掌握谐波分析的基本方法。

二．实验内容

根据沈阳市多年的月平均温度序列，应用谐波分析方法分析其中存在的周期波动。

三．谐波分析的基本思路

根据傅立叶级数理论，一个以基本周期 T 为区间的函数 $f(t)$，若满足一定的条件，总可以表示成一系列频率成倍增加的谐波之和，即

$$f\left(t\right)=c_0+\sum_{k=1}^{\infty}c_k\sin\left(\frac{2\pi kt}{T}+\varphi_k\right) \tag{7.1}$$

其中 c_0 为序列中各项的算术平均值；c_1，c_2，…，c_k…为各谐波的振幅；φ_1，φ_2，…，$\varphi_k\ldots$为各谐波的初位相。每个谐波的周期 $T_k=\dfrac{2\pi}{k\omega}$，当 k＝1 时，$T=\dfrac{2\pi}{\omega}$称为基本周期（或基波），而ω则称为基频。

将(7.1)式变换成下列形式

$$f\left(t\right)=a_0+\sum_{k=1}^{\infty}\left(a_k\cos k\omega t+b_k\sin k\omega t\right) \tag{7.2}$$

其中 a_0＝c_0，$a_k=c_k\sin\varphi_k$，$b_k=c_k\cos\varphi_k$称为傅立叶系数，显然

$$\left.\begin{aligned}c_k^2&=a_k^2+b_k^2\\ \varphi_k&=\arctan\frac{a_k}{b_k}\end{aligned}\right\}\quad k=1,2,\cdots \tag{7.3}$$

式中$\arctan\dfrac{a_k}{b_k}$为多值函数，必须根据系数 a_k 与 b_k 的符号对φ_k取值。

根据最小二乘法和三角级数的正交性，傅立叶系数的计算式为

$$\left.\begin{aligned} a_0 &= \frac{1}{T}\int_0^T f(t)dt \\ a_k &= \frac{2}{T}\int_0^T f(t)\cos k\omega t dt \\ b_k &= \frac{2}{T}\int_0^T f(t)\sin k\omega t dt \end{aligned}\right\} \tag{7.4}$$

对于离散时间序列 y_1，y_2，…，y_n，上式可近似为有限和形式。我们通常按照周期区间(0，T)上的 n 个等距时刻估计系数 a_k 和 b_k，即有

$$\left.\begin{aligned} a_0 &= \frac{1}{n}\sum_{t=1}^{n} y_t \\ a_k &= \frac{2}{n}\sum_{t=1}^{n} y_t \cos k\omega(t-1) \\ b_k &= \frac{2}{n}\sum_{t=1}^{n} y_t \sin k\omega(t-1) \end{aligned}\right\} \tag{7.5}$$

由于序列长度 n 有限，实际上只能取有限个正弦波来逼近序列$\{y_t\}$。一般说来，所取的谐波数最多只能为 $n/2$，因此(7.1)或(7.2)式等号右边只能为有限项，即

$$\hat{y}_t = a_0 + \sum_{k=1}^{K}\left(a_k \cos k\omega t + b_k \sin k\omega t\right) = a_0 + \sum_{k=1}^{K}\left(a_k \cos\frac{2\pi k}{n}t + b_k \sin\frac{2\pi k}{n}t\right) \tag{7.6}$$

其中 $\hat{y}_t$ 表示取 K 个谐波迭加来作为序列$\{y_t\}$的估计；而周期 T_k 与波数 k 以及序列长度 n 之间有如下关系

波数 k	1	2	…	K
周期 T_k	$n/1$	$n/2$	…	n/K
频率 ω_k	ω	2ω	…	$K\omega$

其中 $\omega_k = \dfrac{2\pi}{T_k} = \dfrac{2\pi k}{n}$， $K = \left[\dfrac{n}{2}\right] = \begin{cases} n/2 & n\text{为偶数} \\ (n-1)/2 & n\text{为奇数} \end{cases}$。

类似回归分析，利用正交性可以证明原序列$\{y_t\}$的总方差和各谐波分量的振幅 c_k 有如下关系

$$s_y^2 = \frac{1}{n}\sum_{t=1}^{n-1}\left(y_t - \bar{y}\right)^2 = \sum_{k=1}^{K}\frac{1}{2}\left(a_k^2 + b_k^2\right) = \sum_{k=1}^{K}\frac{1}{2}c_k^2 \tag{7.7}$$

若令 $s_k^2=\frac{1}{2}\left(a_k^2+b_k^2\right)$，则

$$s_y^2=\sum_{k=1}^{K}s_k^2 \tag{7.8}$$

可见，第 k 个谐波对原序列的方差贡献为 $c_k^2/2$，占总方差的比重为 $c_k^2/2s_y^2$。由于三角函数的正交性，各谐波分量彼此是不相关的，因此，根据各谐波的方差贡献大小，可以决定对第 k 个谐波的取舍。不难看出，这里方差贡献其实不过是振幅 c_k 的另一表达形式。

若记第 k 个谐波对原序列的方差贡献 $c_k^2/2$ 为 σ_k^2，那么它相对于时间序列总方差 σ^2 的贡献即为 σ_k^2/σ^2，就可以作为衡量第 k 个谐波重要性的度量，由回归分析中的方差分析知，类似地有统计量

$$F=\frac{\sigma_k^2/2}{\left(\sigma^2-\sigma_k^2\right)/(n-2-1)} \tag{7.9}$$

是服从第一自由度为 2，第二自由度为 n－2－1 的 F 分布。

如(7.9)式中的 σ^2 用样本方差进行估计，则式(7.9)又可写为

$$F=\frac{\frac{1}{2}\left(a_k^2+b_k^2\right)/2}{\left(s_y^2-\frac{1}{2}a_k^2-\frac{1}{2}b_k^2\right)/(n-2-1)} \tag{7.10}$$

四．讨论

论述谐波分析的优缺点。

五．附资料（单位：℃）

月份	1	2	3	4	5	6
气温	-11.77	-7.66	0.55	9.77	17.02	21.73
月份	7	8	9	10	11	12
气温	24.58	23.59	17.3	9.53	0.05	-8.05

实验八 谱分析

一．实验目的

掌握从频率域上研究序列的内部结构，分析时间序列的周期振动方法。

二．实验内容

根据沈阳市 1951～2000 年逐月的月平均温度距平序列，应用谱分析方法分析其中存在的周期波动。

三．谱分析的基本思路

相关函数的重要特性对于了解平稳过程的内部结构具有重要意义。相关函数随着时间间隔 τ 的增大具有波动性。就说明在随机过程的内部结构存在着周期性，由于随机干扰，有时这种周期性表现得很复杂。相关函数的周期性正是随机过程周期性的反应。因此，分析相关函数的结构特点就称为间接认识平稳过程内部结构的一种方法。

在连续频率域上，相关函数 $R_x(\tau)$ 与功率谱密度 $G_x(f)$ 是互为傅立叶变换的一个函数对

$$R_x(\tau)=2\int_0^{\infty}G_x(f)\cos 2\pi f\tau df \tag{8.1}$$

$$G_x(f)=2\int_0^{\infty}R_x(\tau)\cos 2\pi f\tau d\tau \tag{8.2}$$

对于气象时间序列来说，根据谱密度与相关函数互为傅立叶变换的重要性质，自然可通过对相关函数的傅立叶变换来估计谱密度。不过，由于离散序列所得到的相关函数 $R_x(\tau)$实际上是在最大时间间隔(设为 m)上的 m 个等间距点上的函数 R_τ(τ＝0，1，2，…，m)，将(8.2)式改写为

$$G_x(f)=2\int_0^{T}R_x(\tau)\cos 2\pi f\tau d\tau \tag{8.3}$$

借助于梯形求积公式，不难得到相应的有限和式

$$G_k=\frac{1}{m}\left[R_0+2\sum_{\tau=1}^{m-1}R_\tau\cos\frac{k\pi\tau}{m}+R_m\cos k\pi\right] \tag{8.4}$$

这就是通常的谱估计式。上式中 G_k 代表频率（序号）为 k 的谱估

计式，即在频率 $f=\dfrac{k}{2m}$ 邻域上的谱，R_τ 表示第 τ 个时间间隔上的相关函数，m 为最大时间间隔（又称最大时间后延）。实际计算时，还常写成如下形式

$$\left.\begin{aligned}
G_0&=\frac{1}{2m}\left(R_0+R_m\right)+\frac{1}{m}\sum_{\tau=1}^{m-1}R_\tau\\
G_k&=\frac{1}{m}\left[R_0+2\sum_{\tau=1}^{m-1}R_\tau\cos\frac{k\pi\tau}{m}+R_m\cos k\pi\right]\\
G_m&=\frac{1}{2m}\left[R_0+\left(-1\right)^m R_m\right]+\frac{1}{m}\sum_{\tau=1}^{m-1}\left(-1\right)^\tau R_\tau
\end{aligned}\right\}\tag{8.5}$$

这种由相关函数 $R_x(\tau)$的估计值经离散傅立叶变换得到谱估计值的，称为 Blackman-Turkey 方法。

(8.4)式为离散时间序列相关函数的傅立叶变换，且频率在$(0,m)$的有限区间上被截断，加之谱估计值随机抽样的波动性，因而用(8.4)式计算得到的谱是一种非平滑谱。为了减少误差，在实际计算时经常需要对谱估计值加以修正，常用的计算平滑谱的方法有以下两种：

（1）Hanning 平滑

$$\left.\begin{aligned}
\hat{G}_0&=0.5G_0+0.5G_1\\
\hat{G}_k&=0.25G_{k-1}+0.5G_k+0.25G_{k+1}\\
\hat{G}_m&=0.5G_{m-1}+0.5G_m
\end{aligned}\right\}\tag{8.6}$$

（2）Hamming 平滑

$$\left.\begin{aligned}
\hat{G}_0&=0.54G_0+0.46G_1\\
\hat{G}_k&=0.23G_{k-1}+0.54G_k+0.23G_{k+1}\\
\hat{G}_m&=0.46G_{m-1}+0.54G_m
\end{aligned}\right\}\tag{8.7}$$

(8.6)和(8.7)式中 k＝1，2，…，m－1。(8.6)和(8.7)式又称为谱估计值的频率平滑公式。

需要指出的是，在用 Blackman-Turkey 方法进行谱分析时，要涉及到最大后验 m 的选取问题，一般认为 $\dfrac{m}{n}$ 为 $\dfrac{1}{3}$～$\dfrac{1}{10}$，若 m 取的

较小，所能反应的波段少，谱分解就粗糙些，一些长波反应不出来，但谱值稳定；若 m 取的大些，所能反应的波段多，但谱值不稳定。n 很大时，m 可取 $\frac{n}{10}$，n 不太大时，m 取 $\frac{n}{6}$～$\frac{n}{3}$ 为宜。多数情况下，当 n 为偶数时，取 $m=\frac{n}{4}$ 左右。

为了排除样本随机性和有限性的影响，必须对谱进行检验。假设真实的总体谱分布完全均一，个别峰值仅仅是由于随机抽样造成的。也就是说，假设总体谱中的极大值纯由机会造成。在这一假设下，可以证明个别频率上谱估计值与谱的平均值（整个频域上）之比，遵从被其自由度 v 去除的 χ^2 分布，即

$$\frac{G_\alpha}{E(G)}=\frac{\chi_\alpha^2}{v} \tag{8.8}$$

$$v=\frac{2n-m/2}{m} \tag{8.9}$$

G_α 也称临界谱，$E(G)$可通过样本谱 $\bar{G}$ 进行估计，$\bar{G}=\frac{1}{m+1}\sum_{k=0}^{m}\hat{G}_k$，则有 $G_\alpha=\bar{G}\frac{\chi_\alpha^2}{v}$。若 $\hat{G}_k \geqslant G_\alpha$，则表明 k 波段的波动功率谱显著地与假设过程不同，即 k 波段所对应的周期波动是显著的；若 $\hat{G}_k < G_\alpha$，则接受 $H0$，认为是非周期的谱。

四．讨论

讨论如何识别谱图中可能存在的周期波动。

五．附资料（单位：℃）

月份	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
1	-0.19	0.92	-2.35	1.65	-2.18	-1.82	-2.6	-1.52	-0.44	-1.04
2	-1.5	-5.02	-1.58	0.24	0.13	-2.71	-5.85	-1.83	3.48	4.59
3	-2.47	-0.51	1.04	-2.94	-3	-2.34	-6.19	-1.39	3	0.42
4	-1.76	-1	-1.6	-1.45	-2.03	-1.32	-0.84	-0.42	-1.02	-1.36
5	0.44	0.24	-2.05	-2.3	-1.7	-1.33	-0.3	-0.46	1.23	-1.44
6	-1.39	-0.13	-1.12	-1.73	0.34	-0.94	-0.15	-0.11	-0.21	-1.3
7	-0.22	0.7	0.5	-1.81	-0.32	-0.25	-1.88	1.43	-0.54	0.59

8	0.11	-0.74	-1.53	0.34	1.24	-0.19	-1.17	-1.12	0.66	-0.24
9	-2.39	-1.02	-1.05	-0.72	1.33	0.69	-2.57	0.42	0.31	0.97
10	0.2	0.13	1.15	-1.07	-1.15	0.19	-0.39	-1.35	1.18	-0.03
11	0.45	-1.31	-1.72	0.89	0.56	-5.56	2.49	0.08	-3.35	-0.55
12	3.17	-3.43	0.93	-6.67	3.17	-6.99	-0.76	3.84	0.66	-1.77

月份	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
1	0.32	1.24	-4.12	0.53	-1.29	-1.24	0.11	-0.86	-1.45	-2.6
2	1.61	-0.6	0.1	-5.71	-0.51	2.37	-0.88	-4.19	-3.93	0.73
3	1.08	-0.99	1.76	-0.34	-0.94	-0.91	1.46	1.17	-2.27	-4.2
4	0.88	-1.42	0.17	-0.08	-1.87	-1.47	0.53	1.56	-1.75	0.3
5	-0.42	0.2	0.3	2.19	0.24	0.5	2.16	-0.01	-1.14	0.72
6	0.86	-0.43	1.09	-0.54	1.14	-1.05	-1.16	-1.29	-1.34	0.22
7	0.87	0.32	-0.13	-1.21	0.31	-0.8	0.22	0.52	-0.13	-1.03
8	0.51	-0.19	1.25	0.12	-0.76	0.7	1.41	-0.58	-0.31	0.18
9	0.42	-0.02	0.16	-0.43	0.47	-1.16	-0.59	0.09	-0.23	-0.29
10	-0.29	-0.88	-1.43	-0.76	1.3	0.89	0.99	-0.75	-0.14	0.68
11	2.43	-2.4	-0.87	0.77	0.68	0.91	-1.66	1.38	-1.91	0.92
12	0.15	2.13	0.74	0.13	-1.5	-4.54	-5.01	3.07	-0.75	-0.92

月份	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
1	0.74	1.71	3.9	1.43	1.45	1.63	-3.44	0.31	3.7	0
2	-0.92	-1.38	1.09	-0.66	-0.26	3.57	-1.41	-2.62	1.76	-0.37
3	-2.96	1.08	0.34	-0.06	1.58	0.02	0.93	0.1	1.4	0.49
4	-0.07	-0.02	0.84	-0.79	2.55	-1.38	0.56	0.4	-2.24	-2.34
5	0.03	-1.4	-1.85	-0.12	0.37	-0.72	0.29	-0.46	0.18	0.14
6	0.54	-0.62	-0.4	-1.55	0.59	-1.61	0	1.02	0	1.11
7	-0.27	1.07	0.85	0.4	-0.23	-1.4	0.94	0.74	-0.76	-0.61
8	-0.72	-2.02	0.13	0.09	0.39	-1.4	-0.93	-0.3	-0.76	-0.56
9	-0.47	-1.32	-0.03	0.45	1.58	-0.38	0.27	0.71	-0.91	-0.8
10	-0.56	-1.04	-1.48	-2.01	0.35	-0.62	1.52	-1.63	1.52	-1.42
11	2.24	-0.92	0.37	-0.2	2.94	-2.68	-0.03	1.6	-1.19	3.7
12	-0.74	1.71	0.35	-0.38	-0.62	-0.7	-0.06	0.6	3.54	-3.52

月份	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
1	-2.55	0.5	3.8	-0.17	-1.43	-0.14	0.98	4.27	3.49	-2.15
2	0.5	3.14	-0.8	-0.75	0.25	0.18	0.17	-0.63	2.94	2.58
3	1.66	1.42	1.92	-1.98	-1.07	2	-1.58	-0.64	1.72	3.04
4	2.66	1.84	1.44	-0.27	0.76	-0.05	-0.18	-0.37	2.02	-0.51
5	-0.29	-0.32	0.76	1.41	0.9	-0.07	-0.1	0.31	1.27	-1.21

6	1.27	0.63	1.04	0.27	-0.06	0.51	-0.3	1.33	-0.87	-0.23
7	1.31	-0.05	-0.7	0.21	-0.38	-1.38	-0.28	1.07	-1.07	-0.37
8	-0.62	1.32	0.72	0.82	0.52	-1.28	0.14	1.12	-0.03	0.35
9	0.09	0.08	2.44	-0.02	-0.05	-0.31	0.3	1.43	-1.51	-0.56
10	-0.33	1.59	-0.91	-0.34	2.07	-1.15	1.52	1.72	0.14	1.6
11	-3.1	0.81	1.25	1.67	-1.12	0.06	-0.67	0.9	-0.73	2.24
12	1.49	1.48	0.9	0.67	-1.28	2.26	3.59	1.93	0.05	0.19
月份	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	-1.13	3.3	0.64	0.26	3.34	0.4	-1.39	-0.03	3.97	-3.95
2	0.23	2.23	2.98	1.07	2.37	0.3	1.16	4.94	2.91	-1.46
3	-0.82	1.48	2.19	-0.79	1.24	0.34	2.2	3.3	-1.18	1.65
4	0.15	0.58	-1.02	3.24	-0.36	-0.18	1.29	4.3	0.87	0.31
5	-0.4	-0.01	0.96	-0.26	-1.56	0.39	-0.29	1.17	0.41	1.27
6	-0.71	-1.77	-0.53	1.92	-0.22	0.75	1.29	-0.04	0.54	3.16
7	-1.1	-0.55	-1.08	2.09	-1.48	-1	2.15	-0.03	1.87	1.87
8	1.35	-0.68	-0.97	1.35	-0.04	-1.2	1.86	-0.41	0.21	1.77
9	0.19	-1.03	-0.03	0.04	-0.39	-0.02	-0.7	2.11	1.33	2.05
10	-0.47	-1.07	-1.38	0.07	1.21	-0.94	-1.4	3.12	-0.04	-0.6
11	1.05	-1.94	-1.27	2.72	1.77	-1.44	1	-0.75	0.21	-1.9
12	0.01	-0.22	-1.31	-0.37	0.43	2.87	2.62	2.24	-0.03	-1.09

实验九 主成分分析

一．实验目的

掌握应用主成分分析方法对气象要素场进行正交分解的方法。

二．实验内容

对辽宁省 54 站的月平均温度场进行主成分分析。

三．主成分分析的基本思路

主成分分析能够把随时间变化的气象要素场分解为空间函数部分和时间函数(主分量)方法。空间函数部分概括场的地域分布特点，这部分是不随时间变化的；而时间函数部分则由空间点(变量)的线性组合所构成，称为主分量，这些主分量的头几个占有原空间点(变量)的总方差的很大部分。研究主分量随时间变化的规律就可以代替对场的随时间变化的研究。

设我们研究 p 个变量，它可以是 p 个气象要素，也可以是气象要素场中 p 个空间点的值。每个变量有 n 个观测量，每个数据表示为 x_{it}，i＝1，2，…，p；t＝1，2，…，n。写成矩阵形式为

$$X=\left(x_{it}\right)=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} \tag{9.1}$$

我们希望将 X 进行线性组合，组成 p 个新的变量 y_1，y_2，…，y_p，或者说，新变量是由原变量 x_1，x_2，…，x_p 进行线性组合来构成

$$Y=\left(y_{it}\right)=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p1} & y_{p2} & \cdots & y_{pn} \end{pmatrix} \tag{9.2}$$

满足 $y_{it}=v_{1i}x_{1t}+v_{2i}x_{2t}+\cdots+v_{pi}x_{pt}=\sum_{k=1}^{p}v_{ki}x_{kt}$ (i＝1，2，…，p；t＝1，2，…，n)，将此式用矩阵表示即为

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p1} & y_{p2} & \cdots & y_{pn} \end{pmatrix} = \begin{pmatrix} v_{11} & v_{21} & \cdots & v_{p1} \\ v_{12} & v_{22} & \cdots & v_{p2} \\ \cdots & \cdots & \cdots & \cdots \\ v_{1p} & v_{2p} & \cdots & v_{pp} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} \tag{9.3}$$

或者

$$Y = V'X \tag{9.4}$$

其中

$$V = \begin{pmatrix} V_1 & V_2 & \cdots & V_p \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix} \tag{9.5}$$

是待定的，显然，不同的 V 阵，对于不同的线性变化。要给定一定的原则，使组成的新变量会有助于对变量要素场的分析。这些原则是：

（1）要求新变量最大限度地，集中地反应原 p 个变量的方差。

（2）要求这些新变量互相独立，这样就可以用较少的新变量描述气象要素场的主要特征。这些新变量就是主成分或主分量。

根据上面的原则，求解系数矩阵 V 的过程，就转化为求解 X 阵的协方差矩阵 S 的特征值和特征向量问题。可应用 Jacobi 方法进行求解，方法如下：

设 $\boldsymbol{A}$ 为实对称阵，Jocabi 迭代是作一相似矩阵序列。令初始矩阵 $\boldsymbol{A}=\boldsymbol{A}_0$。则第 k 步的相似矩阵为

$$\boldsymbol{A}_k = \boldsymbol{R}_k^{-1}\boldsymbol{A}_{k-1}\boldsymbol{R}_k \tag{9.6}$$

其中

$$\boldsymbol{R}_k=\left(\begin{array}{ccccccccc}
1 & & & & & & & & 0\\
 & \ddots & & & 0 & & & \cdot^{\cdot^{\cdot}} & \\
 & & \cos\theta & & & & \sin\theta & & \\
 & & & 1 & & 0 & & & \\
 & 0 & & & \ddots & & & 0 & \\
 & & & 0 & & 1 & & & \\
 & & -\sin\theta & & & & \cos\theta & & \\
 & \cdot^{\cdot^{\cdot}} & & & 0 & & & \ddots & \\
0 & & & & & & & & 1
\end{array}\right)\begin{array}{l} \\ \\ \to p\text{行} \\ \\ \\ \\ \to q\text{行} \\ \\ \\ \end{array}$$

$$\begin{array}{ccc}\downarrow & \qquad\qquad & \downarrow\\ p & & q\\ \text{列} & & \text{列}\end{array}$$

它的元素为 $r_{pp}=r_{qq}=\cos\theta$，$r_{pq}=\sin\theta$，$r_{qp}=-\sin\theta$，$r_{ii}=1(i\neq p,\ q)$，其余元素为 0。显然这是正交变换，即 $\boldsymbol{R}^{-1}=\boldsymbol{R}'$。所有 $A_k(k=1，2，\cdots)$ 都是对称矩阵，A_k 与 A_{k-1} 只是在第 p，q 行及列上不一样，它们之间的关系如下：

$$\left.\begin{aligned}a_{ip}^{(k)}&=a_{pi}^{(k)}=a_{ip}^{(k-1)}\cos\theta-a_{iq}^{(k-1)}\sin\theta\\ a_{iq}^{(k)}&=a_{qi}^{(k)}=a_{ip}^{(k-1)}\sin\theta+a_{iq}^{(k-1)}\cos\theta\end{aligned}\right\}\quad i\neq p,q$$

$$a_{pp}^{(k)}=a_{pp}^{(k-1)}\cos^2\theta+a_{qq}^{(k-1)}\sin^2\theta-2a_{pq}^{(k-1)}\sin\theta\cos\theta$$

$$a_{qq}^{(k)}=a_{pp}^{(k-1)}\sin^2\theta+a_{qq}^{(k-1)}\cos^2\theta+2a_{pq}^{(k-1)}\sin\theta\cos\theta$$

$$a_{pq}^{(k)}=a_{qp}^{(k)}=\left(a_{pp}^{(k-1)}-a_{qq}^{(k-1)}\right)\sin\theta\cos\theta+a_{pq}^{(k-1)}\left(\cos^2\theta-\sin^2\theta\right)$$

要使 A 逐步称为对角方式，就是要使 $a_{pq}^{(k)}=a_{qp}^{(k)}=0$。

令 $\delta=a_{pq}^{(k-1)}$， $\mu=\frac{1}{2}\left(a_{pp}^{(k-1)}-a_{qq}^{(k-1)}\right)$，则旋转角 θ 应满足

$$\tan 2\theta=-\frac{\delta}{\mu} \tag{9.7}$$

旋转角满足(9.7)式时，则有 $a_{pq}^{(k)}=a_{qp}^{(k)}=0$。

若限制 θ 值在 $-\frac{\pi}{4}<\theta<\frac{\pi}{4}$，由三角公式可推出

$$G=\begin{cases}-\operatorname{sign}(\mu)\dfrac{\delta}{\sqrt{\mu^2+\delta^2}} & (\mu\neq 0)\\ 1 & (\mu=0)\end{cases}$$

则 $\sin\theta=\dfrac{G}{\sqrt{2\left(1+\sqrt{1-G^2}\right)}}$ $\qquad\cos\theta=\sqrt{1-\sin^2\theta}$

每经一次正交变换，主对角线元素平方和增加 $2\left(a_{pq}^{(k-1)}\right)^2$，对角线外元素的平方和相应减小此量。不断进行正交变换，主对角线外元素都趋于 0，矩阵 A_k 趋于一对角矩阵。对角线上的元素就是矩阵的各个特征值，$\boldsymbol{V}=\boldsymbol{R}_1\boldsymbol{R}_2\cdots\boldsymbol{R}_k$ 的每一列对于相应的特征向量。

迭代过程采用限制的循环法，若矩阵 $\boldsymbol{A}$ 的非对角元的平方和记为 S_1，然后计算 $\mu_1=\sqrt{S_1}/n$，这时将 μ_1 看作为一个限制；在遍及矩阵的一次扫描中，所有大于或等于 μ_1 的非对角元都要消去。然后由

$$\mu_2=\mu_1/n$$

作为一个新的限制值，再次扫描矩阵，将大于或等于 μ_2 的非对角元都消去。重复多次，直到满足条件，$\mu_i\le\varepsilon\mu_1$ 为止。这里 ε 为给定的误差限，它将保证非对角元的平方和的最终值小于 $\varepsilon^2 S_1$。

根据 Jacobi 法求出求出 S 阵的特征值和特征向量，并将特征值从大到小进行排序 $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_p$，对应的特征向量为 V_1，V_2，…，V_p。根据(9.4)式，就可求出主成分。

p 个主分量的总方差为 S 阵 p 个特征值之和，即 $\sum_{i=1}^{p}\lambda_i$，也等于原 X 阵 p 个变量的总方差 $\sum_{i=1}^{p}s_{ii}$。则第 k 个主分量的方差贡献大小为

$$R_k=\lambda_k/\sum_{i=1}^{p}\lambda_i \tag{9.8}$$

称前 m 个($m<p$)主分量占总方差的百分率为累计方差贡献率，或称

累计解释方差，表示为

$$G(m)=\sum_{i=1}^{m}\lambda_i/\sum_{i=1}^{p}\lambda_i \tag{9.9}$$

在实际应用时，我们没有必要取全部的主成分进行分析，主成分的个数只要取到能够达到问题要求的精度即可。在气象上，一般要求累计方差贡献率至少应达到85%。

主成分 y_k 与原指标 x_i 的相关系数，称为因子荷载，记为 a_{ik}，并且有

$$a_{ik}=\frac{\sqrt{\lambda_k}v_{ki}}{\sqrt{s_{ii}}} \tag{9.10}$$

它是主成分与原指标之间相关程度的一种度量指标，其值的大小可以看作主成分概括原指标信息的能力。

四．讨论

为什么说主成分分析可以起到因素降维的目的？因素降维在气象分析中有什么现实意义？

五．附资料（单位：0.1℃）

月份	1	2	3	4	5	6	7	8	9	10	11	12
沈阳	-115	-78	7	98	172	217	245	236	173	95	3	-79
辽中	-111	-77	5	93	169	216	243	233	174	98	4	-76
新民	-117	-83	1	91	167	214	242	231	170	93	-3	-83
康平	-129	-94	-8	87	166	213	237	225	166	87	-14	-95
法库	-131	-96	-8	86	164	211	236	223	161	84	-14	-97
大连	-46	-31	26	95	157	197	230	239	200	138	59	-11
复县	-77	-53	15	97	166	209	239	238	186	116	33	-43
皮口	-72	-49	14	86	147	191	228	236	189	121	36	-41
长海	-48	-32	20	80	134	179	220	239	205	143	61	-13
庄河	-79	-51	14	88	150	194	230	236	183	114	30	-47
旅顺	-44	-33	23	95	158	200	234	239	196	134	59	-11
丹东	-80	-50	14	86	148	193	229	233	179	111	27	-50
宽甸	-128	-88	-4	79	147	190	223	222	159	88	1	-87
凤城	-106	-68	8	89	155	198	232	232	171	99	9	-75

东沟	-84	-54	11	85	146	192	229	235	180	110	28	-52
辽阳	-111	-75	11	103	177	220	246	235	174	101	8	-74
锦州	-85	-56	16	101	172	214	241	237	185	111	19	-56
黑山	-106	-73	4	92	168	213	239	231	174	98	3	-75
义县	-108	-74	5	95	172	214	238	228	170	94	-1	-81
北镇	-102	-69	9	96	170	213	238	231	175	100	7	-71
锦县	-93	-63	9	92	166	211	240	234	181	106	13	-63
锦西	-87	-58	15	98	169	211	240	236	185	111	20	-57
建昌	-97	-68	11	104	175	212	233	219	165	96	5	-69
绥中	-80	-53	17	99	168	210	240	236	186	113	26	-49
兴城	-84	-57	13	93	161	205	237	235	183	110	22	-53
抚顺	-141	-100	-7	87	160	205	235	223	156	82	-12	-101
清原	-155	-114	-18	77	149	195	228	217	145	69	-24	-116
新宾	-166	-122	-22	72	145	189	222	211	139	64	-28	-124
阜新	-112	-79	1	94	171	215	242	230	170	90	-6	-83
彰武	-123	-88	-4	87	165	213	240	227	166	87	-11	-92
铁岭	-126	-90	0	95	170	216	243	231	169	94	-2	-87
开原	-141	-103	-9	87	164	210	236	223	159	83	-14	-101
昌图	-135	-98	-11	85	162	209	235	223	163	86	-16	-100
西丰	-172	-131	-22	76	153	200	229	214	144	66	-31	-124
营口	-91	-62	13	99	169	215	248	242	187	112	21	-56
熊岳	-90	-60	16	104	172	214	244	236	180	110	25	-54
盖县	-81	-54	21	110	175	220	248	241	187	114	26	-48
营县	-96	-65	12	100	173	217	244	235	177	105	18	-60
朝阳	-105	-72	12	111	185	224	245	230	172	98	3	-76
建平	-140	-106	-23	78	154	198	221	203	143	68	-30	-111
北票	-105	-72	9	106	182	222	245	232	173	96	0	-78
叶寿	-104	-73	7	104	177	218	239	224	166	92	-2	-77
凌源	-105	-71	10	106	179	219	239	222	165	93	-1	-78
喀左	-104	-69	13	110	182	220	241	225	168	96	3	-75
鞍山	-96	-64	16	105	178	220	247	238	181	108	17	-61

台安	-105	-72	7	93	169	216	243	234	177	101	8	-71
岫岩	-106	-70	5	86	153	196	230	229	166	94	7	-74
海城	-109	-74	11	101	175	217	245	235	175	102	12	-71
本溪	-119	-81	4	97	168	211	240	231	167	94	2	-83
溪县	-139	-100	-7	85	155	199	231	223	156	83	-9	-100
恒仁	-135	-96	-4	86	154	197	228	221	153	83	-5	-96
草河	-123	-88	-9	73	140	185	220	217	150	78	-10	-93
盘山	-100	-70	6	92	164	212	243	237	181	104	13	-65
大洼	-97	-68	8	93	165	214	245	240	184	108	17	-61

实验十 气象资料的插补和订正

一．实验目的

掌握气象序列的差值订正法。

二．实验内容

（1）根据给定的沈阳市 1951～2001 年和本溪市 1956～2001 年的年平均温度序列，采用差值订正法对本溪市 1951～1955 的年平均气温进行插补订正。

（2）用差值订正法将本溪市 1956～2001 年的多年年平均气温延长到 1951～2001 年的多年平均气温。

三．差值订正法的理论基础

观测资料表明，相邻测站气象要素的差值的年际变化较之它本身的年际变化要小得多，这就是所谓的差值稳定性。

大气环流影响的范围是很大的，在距离不太远的两个测站，可以认为是处于同一气候背景下，其同一要素的逐年变化是比较一致的。某些连续变化的气象要素，如温度、湿度、气压、风等。两站间的差值是稳定的，为此，可以利用差值法，将短年代的气象资料订正到长序列，并将某些短缺的资料补上。

四．论述

论述差值订正法的适用范围。

五．附资料(单位：℃)

见实验一资料。

附录 A 逐步回归分析程序

```
PROGRAM MAIN    !    Coded by REN C Y, 2006.5.27
PARAMETER(M=54,N=4,PI=3.1415926)    ! M:样本容量  N:变量的个数
DIMENSION T(N+4,M),XY(N,M),Y(M),A(N,N),XBA(N),SGM(N),      &
                 YY(M)
CHARACTER FILENAME*20
WRITE(*,*)'**************************************************'
WRITE(*,*)'**                                              **'
WRITE(*,*)'*    宁省多年年月平均气温与地理因子的回归分析          *'
WRITE(*,*)'**                                              **'
WRITE(*,*)'**************************************************'
WRITE(*,*)'  本程序用来建立辽宁省 54 站 1960～1990 年多年月平均'
WRITE(*,*)'  气温及年平均气温与海拔高度、经度、纬度的多元线性'
WRITE(*,*)'  和逐步回归方程  '
WRITE(*,*)'**************************************************'
WRITE(*,*)'请输入文件名:'
WRITE(*,*)'注意: '
WRITE(*,*)'1)文件的格式为 ASCII 文本格式  '
WRITE(*,*)'2)文件名不能超过 20 位           '
WRITE(*,*)' 3)与本程序置于同一目录之下    '
WRITE(*,*)' 4)前三列为自变量,后五列依次为 1,4,7,10 月和年均气温'
READ(*,*)FILENAME
WRITE(*,*)'文件名为:',FILENAME
OPEN(2,FILE=FILENAME)
DO I=1,M
      READ(2,*)(T(J,I),J=1,N+4)    !  读入原始数据
ENDDO
WRITE(*,*)' '
WRITE(*,*)'选择要预报的平均气温的月份,按回车键继续'
WRITE(*,*)' 1:  代表一月平均气温'
WRITE(*,*)' 2:  代表四月平均气温'
WRITE(*,*)' 3:  代表七月平均气温'
WRITE(*,*)' 4:  代表十月平均气温'
WRITE(*,*)' 5:  代表年平均气温    '
READ(*,*)ID_MONTH
WRITE(*,*)'请输入引入因子和剔除因子的 F 临界值:'
```

```
READ(*,*)FISN
DO I=1,M
	DO J=1,N-1
		XY(J,I)=T(J,I)
	ENDDO
	XY(N,I)=T(N-1+ID_MONTH,I)
ENDDO
B=0.0
DO I=1,M
	B=B+XY(N,I)
ENDDO
B=B/FLOAT(M)    ！求出预报量的平均值
CALL SRA(M,N,FISN,XY,Y,A,XBA,SGM,YY)
PAUSE '按任意键退出程序'
END

SUBROUTINE SRA(M,N,FISN,XY,Y,A,XBA,SGM,YY)
DIMENSION XY(N,M),Y(M),A(N,N),XBA(N),SGM(N),YY(M)
INTEGER SSS
COMMON K,L,R,YN,H,AMX
! ARGUMENT:
! M: 样本容量
! N: 变量个数
! FISN: 引入和剔除因子的置信限
! XY: 原始资料矩阵
! A:   相关系数矩阵
! Y:   预报值的拟合值
! YY: 因变量与拟合值的差值
SSS=0.0
CALL PP(M,N,XY,XBA,SGM)
! 求出相关系数矩阵 A
DO I=1,N
	DO K=1,N
		C=0.0
		DO J=1,M
			C=C+(XY(I,J)-XBA(I))*(XY(K,J)-XBA(K))
		ENDDO
		A(I,K)=C/(SGM(I)*SGM(K))
```

```
        ENDDO
    ENDDO

    WRITE(*,*)'第零步相关系数矩阵 R(0)为:'
    DO  I=1,N
        WRITE(*,FMT='(10F10.3)')(A(I,J),J=1,N)
    ENDDO
    WRITE(*,*)' '
    WRITE(*,*)' '
    N1=N-1
    DO I=1,N
        XBA(I)=FLOAT(I)
        SGM(I)=0.0
        A(I,I)=1.0
    ENDDO

    L=0        ！方程中自变量的个数
    S=0.0
2   IF(L.EQ.N1) GOTO 3
    AMX=0.0

    ！找到方差贡献最大者，最大为 AMX，最大的因子序数为 K
    CALL IM(XBA,A,N)
    F1=AMX*FLOAT(M-L-2)/(A(N,N)-AMX)
    IF(F1.LE.FISN)GOTO 3
    XBA(K)=0.0          ！表示该因子已引入方程中
    SGM(K)=FLOAT(K)
    L=L+1

    ！矩阵变换,求复相关系数 R 和剩余标准差 YH
    CALL MA(A,N,M)
    WRITE(*,*)'引入第',K,'个因子'
    WRITE(*,*)'变换后的矩阵为:'
    DO  I=1,N
        WRITE(*,FMT='(10F10.3)')(A(I,J),J=1,N)
    ENDDO
    WRITE(*,*)' '
    S=S+1.
```

```
      IF(S.EQ.1.0.OR.S.EQ.2.0)GOTO 2
      AMX=-1.E10

      ! 剔除因子
      CALL IM(SGM,A,N)
      F1=-AMX*FLOAT(M-L-1)/A(N,N)
      IF(F1.GT.FISN)GOTO 40
      SGM(K)=0.0
      XBA(K)=FLOAT(K)
      L=L-1
      CALL MA(A,N,M)
      WRITE(*,*)'剔除第',K,'个因子',F1
      WRITE(*,*)'变换后的矩阵为:'
      DO  I=1,N
          WRITE(*,FMT='(10F10.3)')(A(I,J),J=1,N)
      ENDDO
40    GOTO 2
3     DO 52 I=1,N1
      IF(SGM(I).NE.0)SSS=SSS+1           ! SSS 表示方程中自变量因子的个数
52    A(I,1)=SGM(I)                      ! A 的第一列表示引入方程中自变量
                                         ! 因子的序号
      CALL PP(M,N,XY,XBA,SGM)            ! 重新计算平均值 XBA,离差平方和
                                         ! 的开方值 SGM
      C=0.0
      WRITE(*,*)'引入的变量序号和待定系数'
      DO 60 I=1,N1
      IF(A(I,1))50,60,50
50    A(I,2)=A(I,N)*SGM(N)/SGM(I)        ! A(I,2)为待定系数
      S=A(I,2)
      WRITE(*,51)I,S
51    FORMAT(10X,2HI=,I6,5X,2HB=,F17.6)
      C=C+S*XBA(I)
60    CONTINUE
      S=XBA(N)-C                         ! S 为 b0，即回归方程的常数项
      WRITE(*,*)'回归方程的常数项为      b0＝',S
      WRITE(*,*)'回归方程的复相关系数  R＝',R
      WRITE(*,*)'回归方程的剩余标准差 Sr＝',YN
      WRITE(*,*)'引入方程的因子个数      m＝',SSS
```

```
      IF(SSS.NE.0)THEN
         F005=(R*R/SSS)/((1.-R*R)/(M-SSS-1))
      ELSE
         F005=-999
      ENDIF
      WRITE(*,*)'回归方程总体回归效果 F=',F005

      DO J=1,M
         AMX=S
         DO I=1,N1
            IF(A(I,1).NE.0.0)AMX=A(I,2)*XY(I,J)+AMX
         ENDDO
         Y(J)=AMX              ! 预报值的拟合值
      ENDDO
      DO J=1,M
         YY(J)=XY(N,J)-Y(J)    ! 预报量的实测值XY与拟合值Y的差值
      ENDDO
      RETURN
      END

      SUBROUTINE MA(A,N,M)
      DIMENSION A(N,N)
      COMMON K,L,R,YN,H,AMX
      IF(A(K,K).LT.1.E-5) WRITE(5,10)K
10    FORMAT(1X,2HK=,I6)
      C1=1.0/A(K,K)
      DO 20 I=1,N
         IF(I.EQ.K)GOTO 20
         C2=A(I,K)*C1
         DO 21 J=1,N
            IF(J.NE.K)A(I,J)=A(I,J)-A(K,J)*C2
21       CONTINUE
20    CONTINUE
      DO 35 J=1,N
         IF(J.EQ.K)GOTO 35
         A(K,J)=A(K,J)*C1
         A(J,K)=-A(J,K)*C1
```

```
35    CONTINUE
      A(K,K)=C1
      R=SQRT(1.-A(N,N))
      YN=H*SQRT(A(N,N)/FLOAT(M-L-1))
      RETURN
      END

      SUBROUTINE IM(DS,A,N)
      DIMENSION DS(N),A(N,N)
      COMMON K,L,R,YN,H,AMX
      N1=N-1
      DO I=1,N1
            IF(DS(I).EQ.0.0) CYCLE
            V1=A(I,N)*A(N,I)/A(I,I)
            IF(V1.LE.AMX) CYCLE
            AMX=V1
            K=I
      ENDDO
      RETURN
      END

      SUBROUTINE PP(M,N,XY,XBA,SGM)
      !求出各个因子的平均值 XBA,离差平方和的开方值 SGM
      DIMENSION XY(N,M),XBA(N),SGM(N)
      COMMON K,L,R,YN,H,AMX
      DO I=1,N
            C=0.0
            DO J=1,M
                  C=C+XY(I,J)
            ENDDO
            XBA(I)=C/FLOAT(M)
            C=0.0
            DO J=1,M
                  C=C+(XY(I,J)-XBA(I))**2
            ENDDO
            SGM(I)=SQRT(C)
      ENDDO
      H=SGM(N)
```

RETURN
END

附录 B 谐波分析程序

```
      PROGRAM MAIN    ! Coded by REN C Y, 2002.8.9
      WRITE(*,*)'************************************************'
      WRITE(*,*)'*                                              *'
      WRITE(*,*)'                 谐波分析计算程序                '
      WRITE(*,*)'*                                              *'
      WRITE(*,*)'************************************************'
      WRITE(*,*)'请输入时间序列的长度 N'
      READ(*,*)N
      WRITE(*,*)' '
      PI=4.*ATAN(1.)
      CALL XIEBO(N,PI)
      END

      SUBROUTINE XIEBO(N,PI)
      REAL A0,T(N),AK(N/2),BK(N/2),C(N/2),FAI(N/2),FK(N/2)
      INTEGER K_SORT(N/2)
      ! FK:  每个谐波的 F 值
      ! K_SORT:  排序后的谐波序号
      CHARACTER FILENAME*20
      CHARACTER FILENAME2*20
      WRITE(*,*)' '
      WRITE(*,*)'请输入文件名:'
      WRITE(*,*)'注意: 1)文件的格式为 ASCII 文本格式，按列排列 '
      WRITE(*,*)'      2)文件名不能超过 20 位          '
      WRITE(*,*)'      3)与本程序置于同一目录之下    '
      READ(*,*)FILENAME
      OPEN(2,FILE=FILENAME)
      DO I=1,N
         READ(2,*)T(I)    ！ 读入原始数据
      ENDDO
      WRITE(*,*)'------------------------------------------------'
1001  FORMAT(A18,A20)
1002  FORMAT(A18,I4)
      WRITE(*,*)' '
      WRITE(*,*)'请输入结果保存的文件名'
```

```
READ(*,*)FILENAME2
OPEN(3,FILE=FILENAME2)
WRITE(3,1001)'文件名              :    ',FILENAME
WRITE(3,1002)'时间序列长度为:    ',N
WRITE(3,*)'------------------------------------------------'
WRITE(3,*)' '

TOTAL=0
DO I=1,N
    TOTAL=TOTAL+T(I)
ENDDO
A0=TOTAL/FLOAT(N)       ！计算时间序列的平均值
SUM=0.
DO I=1,N
    SUM=SUM+(T(I)-A0)**2/FLOAT(N)
ENDDO
SY=SUM
WRITE(3,*)'序列的平均值 A0 为:',A0
WRITE(3,*)' '

!谐波的个数取[N/2]
WRITE(3,*)'    K          Ak          Bk              Ck             φk          Fk'
DO K=1,N/2
    K_SORT(K)=K
    CO=0.0
    SI=0.0
    DO I=1,N
        FI=I-1
        AA=2*PI*K*FI/FLOAT(N)
        CO=CO+T(I)*COS(AA)
        SI=SI+T(I)*SIN(AA)
    ENDDO

    AK(K)=2*CO/FLOAT(N)
    BK(K)=2*SI/FLOAT(N)
    C(K)=SQRT((AK(K)**2+BK(K)**2))
    IF(BK(K).EQ.0) FAI(K)=PI/2
    FAI(K)=ATAN(ABS(AK(K)/BK(K)))
```

```
        IF(AK(K).LT.0..AND.BK(K).GT.0) FAI(K)=2*PI-FAI(K)
        IF(AK(K).LT.0..AND.BK(K).LT.0) FAI(K)=PI+FAI(K)
        IF(AK(K).GT.0..AND.BK(K).LT.0) FAI(K)=PI-FAI(K)
        FK(K)=C(K)**2*(N-2-1)/4/(SY-C(K)**2/2)
    ENDDO

    ! 将各个谐波的贡献从大到小排序
    DO I=1,N/2-1
        DO J=N/2-I,N/2-1
            IF(C(J).GE.C(J+1)) THEN
            ELSE
                TEMP=C(J)
                C(J)=C(J+1)
                C(J+1)=TEMP
                TEMP=AK(J)
                AK(J)=AK(J+1)
                AK(J+1)=TEMP
                TEMP=BK(J)
                BK(J)=BK(J+1)
                BK(J+1)=TEMP
                TEMP=FAI(J)
                FAI(J)=FAI(J+1)
                FAI(J+1)=TEMP
                TEMP=FK(J)
                FK(J)=FK(J+1)
                FK(J+1)=TEMP
                K_TEMP=K_SORT(J)
                K_SORT(J)=K_SORT(J+1)
                K_SORT(J+1)=K_TEMP
            ENDIF
        ENDDO
    ENDDO
    DO K=1,N/2
        WRITE(*,1000)K_SORT(K),AK(K),BK(K),C(K),FAI(K),FK(K)
        WRITE(3,1000)K_SORT(K),AK(K),BK(K),C(K),FAI(K),FK(K)
    ENDDO
1000 FORMAT(1X,I3,5F10.4)
    WRITE(3,*)'-----------------------------------------------'
```

```
      WRITE(*,*)
2001  WRITE(*,*)'确认有多少个谐波是显著存在的,输入显著谐波的个数:'
      READ(*,*)MM
      IF(MM.GT.N/2) THEN
         WRITE(*,*)'输入错误,请重新输入!!!!!!!!!!!!!!!'
         GOTO 2001
      ELSE
      ENDIF
      WRITE(3,*)' '
      WRITE(3,1005)'应用',MM,'个谐波进行模拟'
1005  FORMAT(1X,A4,I2,A14)
      WRITE(*,*)' '
      WRITE(3,*)' '
      SUM=0.
      DO I=1,MM
         SUM=SUM+C(I)**2/2
      ENDDO
      SUM=SUM/SY*100.          ! 显著的波占总方差的比重
      IF(SUM.GT.100) SUM=100.
      WRITE(3,1007)'选定的波动占总方差的比重为:',SUM,'%'
      WRITE(3,*)' '
1007  FORMAT(A28,F7.3,A1)
      WRITE(3,*)'          原值          模拟值'
      DO I=1,N
         SUM=A0
         DO J=1,MM       ! N/2
            SUM=SUM+C(J)*SIN(2*PI*K_SORT(J)*I/FLOAT(N)+FAI(J))
         ENDDO
         WRITE(*,1006)T(I),SUM
         WRITE(3,1006)T(I),SUM
      ENDDO
1006  FORMAT(1X,2F15.5)
      WRITE(*,*)'-----------------------------------------------'
      PAUSE '按任意键退出程序'
      END
```

附录 C 谱分析程序

```
PROGRAM MAIN      ! Coded by REN C Y, 2006.3.5
WRITE(*,*)'***********************************************'
WRITE(*,*)'                   时间序列的谱分析                  '
WRITE(*,*)'***********************************************'
WRITE(*,*)'请输入时间序列的长度 N'
READ(*,*)N
WRITE(*,*)'请输入最大时间后延 m'
READ(*,*)M
CALL Spec_Analysis(N,M)
END

SUBROUTINE Spec_Analysis(N,M)
REAL X(N),R(0:M),RR(0:M),G(0:M),G_SMOOTH(0:M)
! X : 原始时间序列
! R : 相关函数
! RR: 自相关函数
! N : 时间序列长度
! M : 最大时间后延
! G : 粗谱
! G_SMOOTH: 平滑谱
CHARACTER FILENAME*20
CHARACTER FILENAME2*20
LOGICAL L_SPEC ! 序列是否存在显著周期波动的标志,有则输出

WRITE(*,*)'请输入文件名:'
WRITE(*,*)'注意: 1)文件的格式为 ASCII 文本格式, 按列排列 '
WRITE(*,*)'      2)文件名不能超过 20 位           '
WRITE(*,*)'      3)与本程序置于同一目录之下   '
READ(*,*)FILENAME
OPEN(2,FILE=FILENAME)
DO I=1,N
    READ(2,*)X(I)   ! 读入原始数据
ENDDO
PI=4.*ATAN(1.)
AVG=0.0
```

```
      DO I=1,N
         AVG=AVG+X(I)
      ENDDO
      AVG=AVG/FLOAT(N)
      SUM=0.0
      DO I=1,N
         SUM=SUM+(X(I)-AVG)**2
      ENDDO
      DX=SUM/FLOAT(N)
      WRITE(*,*)'请输入结果保存的文件名'
      READ(*,*)FILENAME2
      OPEN(3,FILE=FILENAME2)
      WRITE(3,*)'--------------------------------------------------'
      WRITE(3,1001)'文件名            :    ',FILENAME
      WRITE(3,1002)'时间序列长度为:   ',N
      WRITE(3,1002)'最大时间后延为:   ',M
1001  FORMAT(A18,A20)
1002  FORMAT(A18,I4)
      WRITE(3,*)'--------------------------------------------------'
      DO I=0,M
         SUM=0.0
         DO J=1,N-I
            SUM=SUM+(X(J)-AVG)*(X(J+I)-AVG)
         ENDDO
         R(I)=SUM/FLOAT(N-I)                 ! 相关函数
         RR(I)=R(I)/DX                       ! 标准化相关函数
      ENDDO
1003  FORMAT(I10,F10.5,F10.5)

      ! 粗谱估计
      G(0)=(RR(0)+RR(M))/FLOAT(2*M)
      G(M)=(RR(0)+(-1)**M*RR(M))/FLOAT(2*M)
      DO I=1,M-1
         G(0)=G(0)+RR(I)/FLOAT(M)
         G(M)=G(M)+(-1)**I*RR(I)/FLOAT(M)
         G(I)=(RR(0)+RR(M)*COS(I*PI))/FLOAT(M)
         DO J=1,M
            G(I)=G(I)+2*(RR(J)*COS(I*PI*J/FLOAT(M)))/FLOAT(M)
```

```
      ENDDO
   ENDDO

   ! 选择求平滑谱的谱估计方法
2001 WRITE(*,*)'请输入谱平滑的方法: Hamming 平滑请输入 1'
   WRITE(*,*)'                    Hanning 平滑请输入 2'
   READ(*,*)ID_SMOOTH_METHOD
   IF(ID_SMOOTH_METHOD.EQ.1) THEN
      A1=0.54
      A2=0.46
   ELSEIF(ID_SMOOTH_METHOD.EQ.2) THEN
      A1=0.50
      A2=0.50
   ELSE
      WRITE(*,*)'平滑谱方法选择错误,请重新输入:'
      GOTO 2001
   ENDIF

   ! 求平滑谱
   SUM=0.
   WRITE(3,*)'后延(波数)  相关函数  标准化相关函数    粗谱    平滑谱'
   DO I=0,M
      IF(I.EQ.0) THEN
         G_SMOOTH(I)=A1*G(I)+A2*G(I+1)
         IF(G_SMOOTH(I).LT.0) G_SMOOTH(I)=0.
         SUM=SUM+G_SMOOTH(I)
      ELSEIF(I.EQ.M) THEN
         G_SMOOTH(I)=A1*G(I)+A2*G(I-1)
         IF(G_SMOOTH(I).LT.0) G_SMOOTH(I)=0.
         SUM=SUM+G_SMOOTH(I)
      ELSE
         G_SMOOTH(I)=A2*G(I-1)/2.+A1*G(I)+A2*G(I+1)/2.
         IF(G_SMOOTH(I).LT.0) G_SMOOTH(I)=0.
         SUM=SUM+G_SMOOTH(I)
      ENDIF
      WRITE(3,1010)I,R(I),RR(I),G(I),G_SMOOTH(I)
   ENDDO
   WRITE(3,*)'--------------------------------------------------'
```

```
1010  FORMAT(I6,F12.3,F12.3,F10.3,F10.3)
      G_AVG=SUM/FLOAT(M+1)
      V=(2.*N-M/2.)/FLOAT(M)
      WRITE(*,1011)'请输入给定信度下自由度为',V,'的卡方分布      &
                  的临界值:'
1011  FORMAT(A25,F6.2,A18)
      READ(*,*)X2
      L_SPEC=.False.
      WRITE(3,*)' 存   在   的   显   著   的   周   期   波   动'
      WRITE(3,*)'             __________________          '
      WRITE(3,*)'   波数   角频率      频率     周期     卡方值'
      DO I=1,M
         IF(G_SMOOTH(I).GE.G_AVG*X2/V) THEN
            L_SPEC=.True.
            WRITE(3,1005)I,I*PI/FLOAT(M),I/(2.*M),2.*M/FLOAT(I), &
                         V*G_SMOOTH(I)/G_AVG
         ELSE
         ENDIF
      ENDDO
1005  FORMAT(4X,I3,4F10.3)
      IF(.NOT.L_SPEC) THEN
         WRITE(3,*)'      !!!时间序列中没有显著的周期波动!!!'
      ENDIF
      PAUSE   '按任意键退出程序'
      END
```

附录 D 主成分分析程序

```
PROGRAM MAIN    ! Coded by REN C Y, 1997.3.20
WRITE(*,*)'************************************************'
WRITE(*,*)'*                                              *'
WRITE(*,*)'            Principal Component Analysis        '
WRITE(*,*)'*                                              *'
WRITE(*,*)'************************************************'
WRITE(*,*)'请输入变量或空间点的个数 P'
READ(*,*)M
WRITE(*,*)'请输入样本数 N'
READ(*,*)N
CALL PCA(N,M)
END

SUBROUTINE PCA(N,M)
REAL A1(M,M),A2(M,M),A4(M,M),X(N,M),Y(N,M)
CHARACTER FILENAME*20
CHARACTER FILENAME2*20
WRITE(*,*)' '
WRITE(*,*)'请输入文件名:'
WRITE(*,*)'注意: 1)文件的格式为 ASCII 文本格式，按列排列 '
WRITE(*,*)'          2)文件名不能超过 20 位          '
WRITE(*,*)'          3)与本程序置于同一目录之下    '
READ(*,*)FILENAME
OPEN(2,FILE=FILENAME)
DO I=1,N
     READ(2,*)(X(I,J),J=1,M)   ！ 读入原始数据
ENDDO
WRITE(*,*)' '
WRITE(*,*)'请输入结果保存的文件名'
READ(*,*)FILENAME2
OPEN(3,FILE=FILENAME2)
WRITE(*,*)'请输入累积贡献率的标准(%):'
READ(*,*)THERESHOLD
CALL RCOEF(X,N,M,A1)
！ 此时计算出来的 A1(M,M)为相关系数矩阵
```

```
      WRITE(*,*)'原始资料的相关系数矩阵:'
      WRITE(3,*)'原始资料的相关系数矩阵:'
      DO I=1,M
          WRITE(*,1001)(A1(I,J),J=1,M)
          WRITE(3,1001)(A1(I,J),J=1,M)
      ENDDO
1001  FORMAT(100F6.3)

      CALL JACOBI(A1,A2,M,MN)
      !  此时的 A1(M,M)对角线方向存放的是已排好序的特征值
      ! A2(M,M)存放的是排好序的特征向量.
      ! A2(M,M)中的第一个 M 为特征向量的维数,第二个为向量的序数
      !  如 A2(I,J)为第 J 个特征向量的第 I 维分量.
1017  FORMAT(1X,8(F8.4))
      SUME=0.
      DO 100 I=1,M
          A1(I,1)=A1(I,I)
          ! A1(I,1),I=1,M 为排好序的特征值
          SUME=SUME+A1(I,1)
100   CONTINUE

      SUMEE=0.
      DO 101 I=1,M
          A1(I,2)=A1(I,1)*100/SUME
          !  A1(I,2),I=1,M 为每个特征值的贡献率
          SUMEE=SUMEE+A1(I,1)
          A1(I,3)=SUMEE*100/SUME
          !  A1(I,3),I=1,M 为累积贡献率
101   CONTINUE
      WRITE(*,2003)
      WRITE(3,2003)
2003  FORMAT(//7X, 'COLUMN1          COLUMN2          COLUMN3')
      DO 10 I=1,M
          WRITE(*,1002)(A1(I,J),J=1,3)
          WRITE(3,1002)(A1(I,J),J=1,3)
          IF(A1(I,3).GE.THERESHOLD)THEN
              ICR2=I
```

```
                GOTO 401
            ELSE
            ENDIF
10      CONTINUE
401     CONTINUE
1002    FORMAT(3F15.6)
        WRITE(*,410)
        WRITE(3,410)
410     FORMAT(//7X,'MATRIX OF FACTORS LOADING      CO-DEGREE ')
        DO I=1,M
            SUM=0.0
            DO J=1,ICR2
                A4(I,J)=SQRT(A1(J,1))*A2(I,J)
                SUM=SUM+A4(I,J)**2
            ENDDO
            WRITE(*,1003)(A4(I,J),J=1,ICR2),SUM
            WRITE(3,1003)(A4(I,J),J=1,ICR2),SUM
        ENDDO
1003    FORMAT(1X,100F14.5)

        ! Calculate the principal component
        WRITE(*,3001)
        WRITE(3,3001)
3001    FORMAT(//7X,'The principal component')
        DO 20 J=1,N
            DO 21 I=1,ICR2
                Y(J,I)=0.0
                DO 22 KL=1,M
                    Y(J,I)=Y(J,I)+X(J,KL)*A2(KL,I)
22              CONTINUE
21          CONTINUE
            WRITE(*,1005)(Y(J,I),I=1,ICR2)
            WRITE(3,1005)(Y(J,I),I=1,ICR2)
20      CONTINUE
1005    FORMAT(5X,100F11.3)
        pause '按任意键退出程序'
        END
```

```
      SUBROUTINE RCOEF(X,N,M,A)
!  计算相关系数矩阵
      DIMENSION X(N,M),A(M,M)
      AN=N
      DO 100 I=1,M
      DO 100 J=1,M
         SX1=0.0
         SX2=0.0
         SX1X1=0.0
         SX2X2=0.0
         SX1X2=0.0
         DO 101 K=1,N
            SX1=SX1+X(K,I)
            SX2=SX2+X(K,J)
            SX1X1=SX1X1+X(K,I)**2
            SX2X2=SX2X2+X(K,J)**2
            SX1X2=SX1X2+X(K,I)*X(K,J)
101      CONTINUE
         R=(SX1X2-SX1*SX2/AN)/SQRT((SX1X1-SX1*SX1/AN)      &
           *(SX2X2-SX2*SX2/AN))
         A(I,J)=R
         A(J,I)=R
100   CONTINUE
      RETURN
      END

      SUBROUTINE JACOBI(A,B,N,N1)
!  计算特征值和特征向量
      DIMENSION A(12,12),B(12,12)
      ANORM=0.0
      DO 100 I=1,N
         DO 101 J=1,N
            IF(I-J)2,1,2
1           B(I,J)=1.0
            GO TO 101
2           B(I,J)=0.0
            ANORM=ANORM+A(I,J)*A(I,J)
101      CONTINUE
```

```
100     CONTINUE
        ANORM=SQRT(ANORM)
        FNORM=ANORM*1.0E-09/FLOAT(N)
        THR=ANORM
23      THR=THR/FLOAT(N)
3       IND=0
        DO 102 I=2,N
            I1=I-1
            DO 103 J=1,I1
                IF(ABS(A(J,I))-THR)103,4,4
4               IND=1
                AL=-A(J,I)
                AM=(A(J,J)-A(I,I))/2.0
                AO=AL/SQRT(AL*AL+AM*AM)
                IF(AM)5,6,6
5               AO=-AO
6               SINX=AO/SQRT(2.0*(1.0+SQRT(1.0-AO*AO)))
                SINX2=SINX*SINX
                COSX=SQRT(1.0-SINX2)
                COSX2=COSX*COSX
                DO 104 K=1,N
                    IF(K-J)7,10,7
7                   IF(K-I)8,10,8
8                   AT=A(K,J)
                    A(K,J)=AT*COSX-A(K,I)*SINX
                    A(K,I)=AT*SINX+A(K,I)*COSX
10                  BT=B(K,J)
                    B(K,J)=BT*COSX-B(K,I)*SINX
                    B(K,I)=BT*SINX+B(K,I)*COSX
104             CONTINUE
                XT=2.0*A(J,I)*SINX*COSX
                AT=A(J,J)
                BT=A(I,I)
                A(J,J)=AT*COSX2+BT*SINX2-XT
                A(I,I)=AT*SINX2+BT*COSX2+XT
                A(J,I)=(AT-BT)*SINX*COSX+A(J,I)*(COSX2-SINX2)
                A(I,J)=A(J,I)
                DO 105 K=1,N
```

```
                  A(J,K)=A(K,J)
                  A(I,K)=A(K,I)
105            CONTINUE
103        CONTINUE
102    CONTINUE
       IF(IND)20,20,3
20     IF(THR-FNORM)25,25,23
25     DO 110 I=2,N
           J=I
29         IF(A(J-1,J-1)-A(J,J))30,110,110
30         AT=A(J-1,J-1)
           A(J-1,J-1)=A(J,J)
           A(J,J)=AT
           DO 111 K=1,N
               AT=B(K,J-1)
               B(K,J-1)=B(K,J)
               B(K,J)=AT
111        CONTINUE
           J=J-1
           IF(J-1)110,110,29
110    CONTINUE
       ! WRITE(30,*)'THIS IS THE VECTOR'
       DO 19 I=1,N1
       ! WRITE(30,1000)I,(B(J,I),J=1,N1)
 19  CONTINUE
1000   FORMAT(1X,I3,12(F10.7))
       RETURN
       END
```