讲 座

Ⅳ.时间趋势分析(一)

上海医科大学* 詹绍康

按时间顺序排列的某变量的一系列观察值就组成了一组时间趋势资料。宏观的时间趋势资料中不同时间的观察值变化，可分析为四种成分：

总趋势 这是指指标随时间变化的总趋势。如某市四十年来传染病死亡率的下降趋势；肺癌死亡率的上升趋势等。

季节变异 对于所观察到的各年的资料，有时可见到明显的季节性变动。如冬春季呼吸道发病率高于其它季节；夏季的溺水死亡率高于其它季节。

随机变异 每个观察值都含有随机变异的成分；有时，随机变异很大，以至使人感到数值大小变幻莫测；有时，随机变异很小，甚至可以略而不计。从宏观或长期趋势来看，各种随机因素的影响又可能相互抵消。

螺旋 （或称循环）除了总趋势和季节变异外，有时还能见到在长期变化趋势中有一种颇有规律的波浪型的变化，并呈现一浪高一浪或一浪低一浪的情况。

作时间趋势分析，不仅要研究趋势的特征，而且要识别是什么因素在影响和促成这种变化趋势，并把某一种影响因素的作用从众多影响因素的综合作用中提炼出来再用某种方法给以度量。

总 趋 势

一、量测趋势的基本假定：测量时间趋势的方法有多种，但都有一个基本的假定：变量值都波动于一条基本的、或多或少是光滑的趋势线上下，这种波动常被认为是随机因素造成的。因此，趋势线就可以被想象成一种平均值的连线。尽管某一时间的观察值会有变化，但它们的平均值仍可看作是一个固定的值。由这些平均值所得的连线，不论是直线还是曲线，就代表了观察值的变化趋势。在一般情况下，以直观方法概括变化趋势已能满足分析需要。当然，此外还有一些更细致更客观的测量趋势的方法。

二、移动平均法：移动平均数有普通型和加权型两种。普通型移动平均只要顺次取n个数据分别计算平均数即可。下面介绍一种较常用的加权型移动平均数。

如果在时间t的观察值与该时间前后的观察值是有联系的，那么其平均值或基本趋势值可以由t−2，t−1，t，t+1，t+2五个时间点（当然不一定是五点，可以是三点、四点，也可以是六点、七点，“点”数就是移动平均的长度）的观察值的平均值所得。不难理解，离时间t越近的观察值对平均值的影响越大，远离时间t的观察值对平均值的影响越小。这就是说，要用在时间t前后的观察值以加权法求平均值，离时间t近者给以较大的权，远者给以较小的权。

现举例说明加权移动的平均分析方法。设某一段时间的五个连续的原始观察值为：

$${}_0X_1,\ {}_0X_2,\ {}_0X_3,\ {}_0X_4,\ {}_0X_5$$

第一轮先计算相邻两值之平均值，可得：

$${}_1X_{1.5},\ {}_1X_{2.5},\ 4X_{3.5},\ {}_1X_{4.5}$$

式中，$${}_1X_{1.5}=\tfrac{1}{2}({}_0X_1)+\tfrac{1}{2}({}_0X_2)$$

$${}_1X_{2.5}=\tfrac{1}{2}({}_0X_2)+\tfrac{1}{2}({}_0X_3)$$

……

第二轮再计算相邻两值的平均值，可得：

$${}_2X_2,\ {}_2X_3,\ {}_2X_4$$

式中，$${}_2X_2=\tfrac{1}{2}({}_1X_{1.5})+\tfrac{1}{2}({}_1X_{2.5})$$

$${}_2X_3=\tfrac{1}{2}({}_1X_{2.5})+\tfrac{1}{2}({}_1X_{3.5})$$

$${}_2X_4=\tfrac{1}{2}({}_1X_{3.5})+\tfrac{1}{2}({}_1X_{4.5})$$

显然

$${}_2X_3=\tfrac{1}{2}[\tfrac{1}{2}({}_0X_2)+\tfrac{1}{2}({}_0X_3)]+\tfrac{1}{2}[\tfrac{1}{2}({}_0X_3)+\tfrac{1}{2}({}_0X_4)]=\tfrac{1}{4}({}_0X_2)+\tfrac{1}{2}({}_0X_3)+\tfrac{1}{4}({}_0X_4)$$

此时权重系数的比例是1：2：1。仿照上述方法，再作两轮移动平均，可得

$${}_4X_3=\frac{1}{16}({}_0X_1)+\frac{4}{16}({}_0X_2)+\frac{6}{16}({}_0X_3)+\frac{4}{16}({}_0X_4)+\frac{1}{16}({}_0X_5)$$

---

* 邮政编码 200032

此时权重系数的比例是1：4：6：4：1。实际上，不论移动平均的长度是几，权重的比例就是二项式的系数。上述计算过程可以列表进行，简便而易于核对。

表1是16年来某地20岁女子初婚率数据（第2列），

表1 20岁女子初婚率的五点加权移动平均值计算表

年份	初婚率	移动的相邻两值之和				(6)÷16
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1960	0.106					
		0.236				
1961	0.130		0.506			
		0.270		1.047		
1962	0.140		0.541		2.080	0.130
		0.271		1.033		
1963	0.131		0.492		1.917	0.120
		0.221		0.884		
1964	0.090		0.392		1.633	0.102
		0.171		0.749		
1965	0.081		0.357		1.515	0.095
		0.186		0.766		
1966	0.105		0.409		1.664	0.104
		0.223		0.898		
1967	0.118		0.489		1.961	0.123
		0.266		1.063		
1968	0.148		0.574		2.227	0.139
		0.308		1.164		
1969	0.160		0.590		2.262	0.141
		0.282		1.098		
1970	0.122		0.508		2.034	0.127
		0.226		0.936		
1971	0.104		0.428		1.763	0.110
		0.202		0.827		
1972	0.098		0.399		1.611	0.101
		0.197		0.784		
1973	0.099		0.385		1.530	0.096
		0.188		0.746		
1974	0.089		0.361		⋮	⋮
		0.173		⋮		
1975	0.084		⋮			
⋮	⋮	⋮				

第3、4、5和6列都为各自前一列的相邻两值之和。第7列由第6列数据除以16而得，它就是5点加权平均值。用5点加权平均值作图，就能得较为光滑的趋势线，代表初婚率的变化趋势（图1）

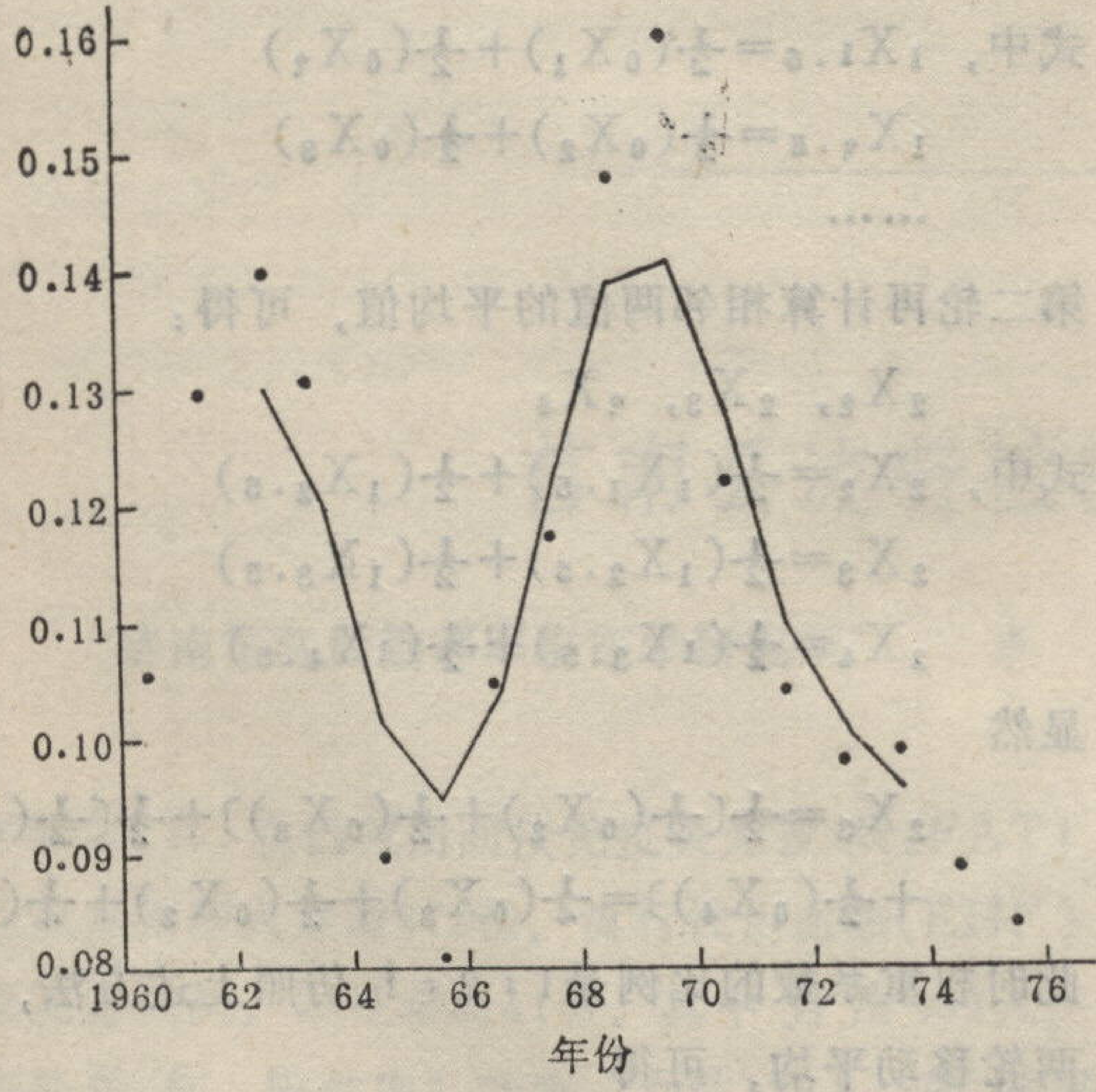

图1 20岁女子的初婚率及五点移动平均值

三、移动平均的性质：一般地说，移动平均的过程消除了一些过高或过低的数值，或者说移动平均过程抹去了短时间内的变异。但移动平均的作用并不只限于这一点，

如果原始数据呈周期性起伏，而移动平均中的长度又正好等于周期长度，那么移动平均的结果就完全消除了周期性的变异，因为每一个移动平均值所使用的原始数据包含了正好是一个周期内的不同数据，甚至可以说每一个移动平均值计算时所用的数据都是相仿的。这种性质，有时却是有用的。例如，以每12个月的意外死亡数计算移动平均，因12个月正好是季节的一个周期，这种移动平均显然就消除了季节变异。

用移动平均来代表趋势也有一些缺点：

1.对起始和最终的那些时间，无法计算移动平均值。在首尾各约有半个移动平均的长度是空白的（长度为奇数或偶数时略有差别）。

2.移动平均难以用数学函数来表达，因此难以用作客观预测。由于趋势分析的主要目的之一是作预测，而移动平均又难以满足要求，所以在作趋势分析中应用移动平均法的作用也是有限的。

四、消除季节变动看趋势：前面提到过，如果移动平均的长度与周期性变化中周期长度相等，移动平均的结果就消除了周期性变化的影响。要达此目的，对于按一天中24个小时收集的数据，就应取移动平均长度为24；对于按月份收集的数据，就应取移动平均长度为12；对于按季度收集的数据，就应取移动平均长度为4。

〔例1〕 表2数据是几年来每个季度的药品消耗量（千元），显然存在着季节变动，作长度为4的移动平均后，消除了季节变动而显示了总的变化趋势。为简便起见，这里采用不加权的4点移动平均。

表2 药品消耗量的不加权4点移动平均

年份	季度	药品消耗（千元）	四点合计	第4列两点合计	趋势
(1)	(2)	(3)	(4)	(5)	(6)
1981	1	114			
	2	142			
			547		
	3	155		1096	137.00
			549		
	4	136		1106	138.25
			557		
1982	1	116		1112	139.00
			555		
	2	150		1114	139.25
			559		
	3	153		1130	141.25
			571		
	4	140		1150	143.75
			579		
1983	1	128		1174	146.75
			595		
	2	158		1209	151.13
			614		
	3	169		1237	154.63
			623		
	4	159		1268	158.50
			645		
1984	1	137		1313	164.13
			668		
	2	180		1349	168.63
			681		
	3	192		1390	171.25
			709		
	4	172		1432	179.00
			723		
1985	1	165			
	2	194			

表2中第4列是第3列原始数据的四点移动合计，如

$$114+142+155+136=547$$

$$142+155+136+116=549$$

如果将第4列的数据除以4，就可得长度为4的移动平均值，如547÷4＝136.75，549÷4＝137.25。但这样做存在一个问题，算得的136.75不代表二季度，也不代表三季度，而是处于二、三季度之间，同样，137.25处于三、四季度之间，若取这两者的平均数$\frac{1}{2}$(136.75＋137.25)＝137.00，却正好是代表三季度的移动平均值。第5列就是为此而计算的第4列相邻两值的合计，再除以8，就是代表趋势的移动平均值了（图2）。

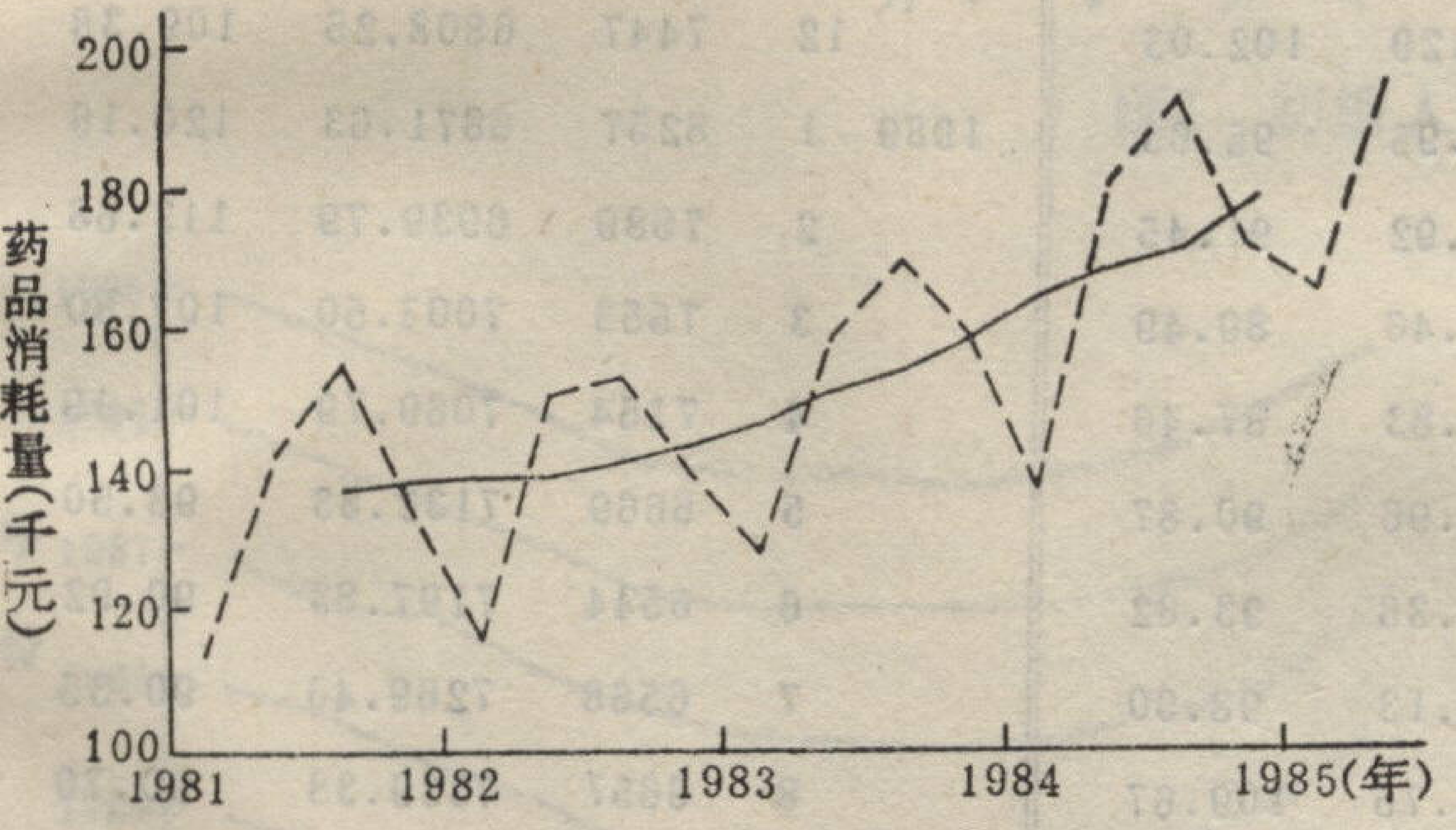

图2 药品消耗量的季节变动和变化趋势

把表2中第3列的原始数据和第6列的趋势数据画在图2中，可以见到原始数据有明显的季节消长变化，而移动平均值的连线，完全消除了季节消长，呈现的就是总的变化趋势。

季节变动

一、周期性变化：在某些资料中，会由于某种因素造成有规律的周期性数值变化。所谓周期性，就是以一定的时间间隔重现高峰和低谷。一个周期可以是一天、一周、一个月或一年等。这里以季节变动为例来说明周期性变化的分析方法。移动平均法可应用于任何一种周期性变化的分析

二、季节变动分析的基本思想 移动平均法用于分析季节变动的原理就在于对原始数据系列和消除了季节变动后的移动平均系列作比较。以一年中十二个月的变动为例，一方面，每一个月有一个原始数据；另一方面，每一个月又有一个移动平均值。这些移动平均值中的每一个都包含了十二个月的数值的信息。一方面它包含了每一月份的信息，另一方面每一个月份的信息又都只被包含一次。这些移动平均值与季节变动无关，而每个月的实际观察值却又在某种程度上受季节因素的影响，所以在分析中就可把注意力集中在移动长度等于季节变动周期的移动平均值与原始数值的比较方面。

三、中心化移动平均及其对原始数值的相对比：前面已提及，若取表2第4列的每一个数值除以4可得长度为4的移动平均系列：136.75，137.25，……。因为第3列的第一个数值——第一季度的药品消耗量是代表第一季度这一段时间的，以后各季度也是如此。所以四个季度的平均值实际代表的是第二与第三季度交界点的指标。同理，如果以前一年的11、12月及后一年的1至10月的数值作移动平均，它所代表的时间点是4月底和5月初的交点。如果要使每一个移动平均值能代表相对应的一个月的指标，应该使移动平均值所代表的时间点转移到该个月的中点（月中）来，这个过程就称为移动平均的中心化。

如果移动平均长度是奇数，那么移动平均值是自动中心化的，如一周内的7天（星期日到星期六），这7天的指标的平均值代表的是星期三。如果移动平均的长度是偶数，那么每次得到的是非中心移动平均。例如，以上一年的4～12月和下一年的1～3月的12个数值计算平均值，同时又以上一年的5～12月和下一年的1～4月的12个数值计算平均值，那么这两个都是非中心化的移动平均值，它们代表的时间点分别是9月底10月初和10月底11月初，把这两个均值再作平均，便成了中心化的移动平均，它代表10月份（或10月份的中心）的指标值。

表3是5年中每个月的初婚数和用上述方法算得的中心化移动平均值。当然，在计算时还需有1984年下半年及1990年上半年各月的初婚数。第3列的计算方法是：

$$\frac{\text{初婚数}}{\text{同月中心化移动平均值}}\times 100$$

如1985年1月，$\dfrac{4777}{3991.92}\times 100=119.67$

第3列是相对比，显示了季节变动的轮廓，并可作为进一步分析的依据。

四、季节指数：把表3第3列抄出来列于表4，可称为移动平均百分数。

移动平均百分数之间的差异是由多种因素造成的，但与原始数值相比，差别缩小了。这是因为在计算过程中已消除了趋势效应、部分随机效应及短长度的周期效应。

表4中数值差异，是由三种原因造成的：季节效应、螺旋效应和随机效应。然后，对于任一横行数值间差

表3　　5年中各月初婚人数及中心化移动平均值

年.月	初婚数 (1)	中心化移动平均 (2)	$\frac{(1)}{(2)}\times 100$ (3)	年.月	初婚数 (1)	中心化移动平均 (2)	$\frac{(1)}{(2)}\times 100$ (3)	年.月	初婚数 (1)	中心化移动平均 (2)	$\frac{(1)}{(2)}\times 100$ (3)
1985. 1	4777	3991.92	119.67	1986 9	4619	5010.58	92.18	1988 5	5950	6306.88	94.34
2	4633	4033.92	114.85	10	4749	5074.71	93.58	6	5812	6381.71	91.07
3	4391	4075.50	107.74	11	5032	5138.54	97.93	7	5779	6461.50	89.44
4	4159	4120.38	100.94	12	5604	5199.75	107.77	8	5810	6540.67	88.83
5	3913	4196.96	93.84	1987 1	6276	5254.54	119.44	9	6065	6615.33	91.68
6	3815	4222.00	90.36	2	6017	5303.96	113.44	10	6186	6685.75	92.53
7	3823	4275.25	89.42	3	5782	5354.21	107.99	11	6712	6747.79	99.47
8	3824	4328.75	88.34	4	5521	5411.29	102.03	12	7447	6808.25	109.38
9	4018	4380.79	91.72	5	5235	5473.96	95.63	1989 1	8257	6871.63	120.16
10	4172	4431.38	94.15	6	5072	5545.92	91.45	2	7889	6939.79	113.68
11	4495	4479.38	100.35	7	5034	5625.46	89.49	3	7553	7003.50	107.80
12	4959	4525.63	109.58	8	4991	5706.83	87.46	4	7154	7069.79	101.19
1986 1	5425	4573.67	118.61	9	5256	5783.96	90.87	5	6669	7132.83	93.50
2	5269	4623.33	113.96	10	5482	5855.38	93.62	6	6544	7197.83	90.92
3	5004	4673.29	107.08	11	5803	5921.13	98.00	7	6568	7269.46	90.35
4	4760	4722.38	100.80	12	6560	5981.75	109.67	8	6657	7339.38	90.70
5	4464	4768.79	93.61	1988 1	7229	6043.63	119.28	9	6819	7405.54	92.08
6	4374	4818.04	90.78	2	7017	6108.79	114.87	10	6951	7475.25	92.99
7	4417	4880.38	90.51	3	6633	6176.63	107.39	11	7460	7548.50	98.83
8	4422	4947.00	89.39	4	6384	6239.67	102.31	12	8259	7621.54	108.36

表4　　各月的移动平均百分数

月份	1985年	1986年	1987年	1988年	1989年
1	119.67	118.61	119.44	119.28	120.16
2	114.85	113.96	113.44	114.87	113.68
3	107.74	107.08	107.99	107.39	107.80
4	100.94	100.80	102.03	102.31	101.19
5	93.84	93.61	95.63	94.34	93.50
6	90.36	90.78	91.45	91.07	90.92
7	89.42	90.51	89.49	89.44	90.35
8	88.34	89.39	87.46	88.83	90.70
9	91.72	92.18	90.87	91.68	92.08
10	94.15	93.58	93.62	92.53	92.99
11	100.35	97.93	98.00	99.47	98.83
12	109.58	107.77	109.67	109.38	108.36

异是不能用季节效应来解释的。因为它们所代表的是同一月份。从这些移动平均百分数可以看出，1月份的观察值约比移动平均值高19%，8月份观察值约比移动平均值低10%。同一横行之间的差异是由螺旋及随机效应造成的。

如果对各横行求平均值，理论上又消除了螺旋效应和随机效应的作用而获得了表达季节效应的一种度量。

对于表4数据，可以有两种图示法表示季节变动趋势：

1.季节分析图：这类似于相关分析中的点图。但在季节分析图中，横轴只取12个可能的值（12个月），每一个值代表一个区间（月）而不是一个点。因此，对应于每一个横轴取值会有几个纵轴取值。为避免点子重合，可把横轴取值均匀分散地画在该月这个区间内（图3）。

2.阶层图：这是一种显示各年度是否有一致或相仿的季节变动趋势的分析图，不论其变化是突然的还是逐渐的都可用（图4）。

所谓季节指数，就是不同年份相同月份的移动平均百分数的平均值。由于算术平均值易受个别偶然出

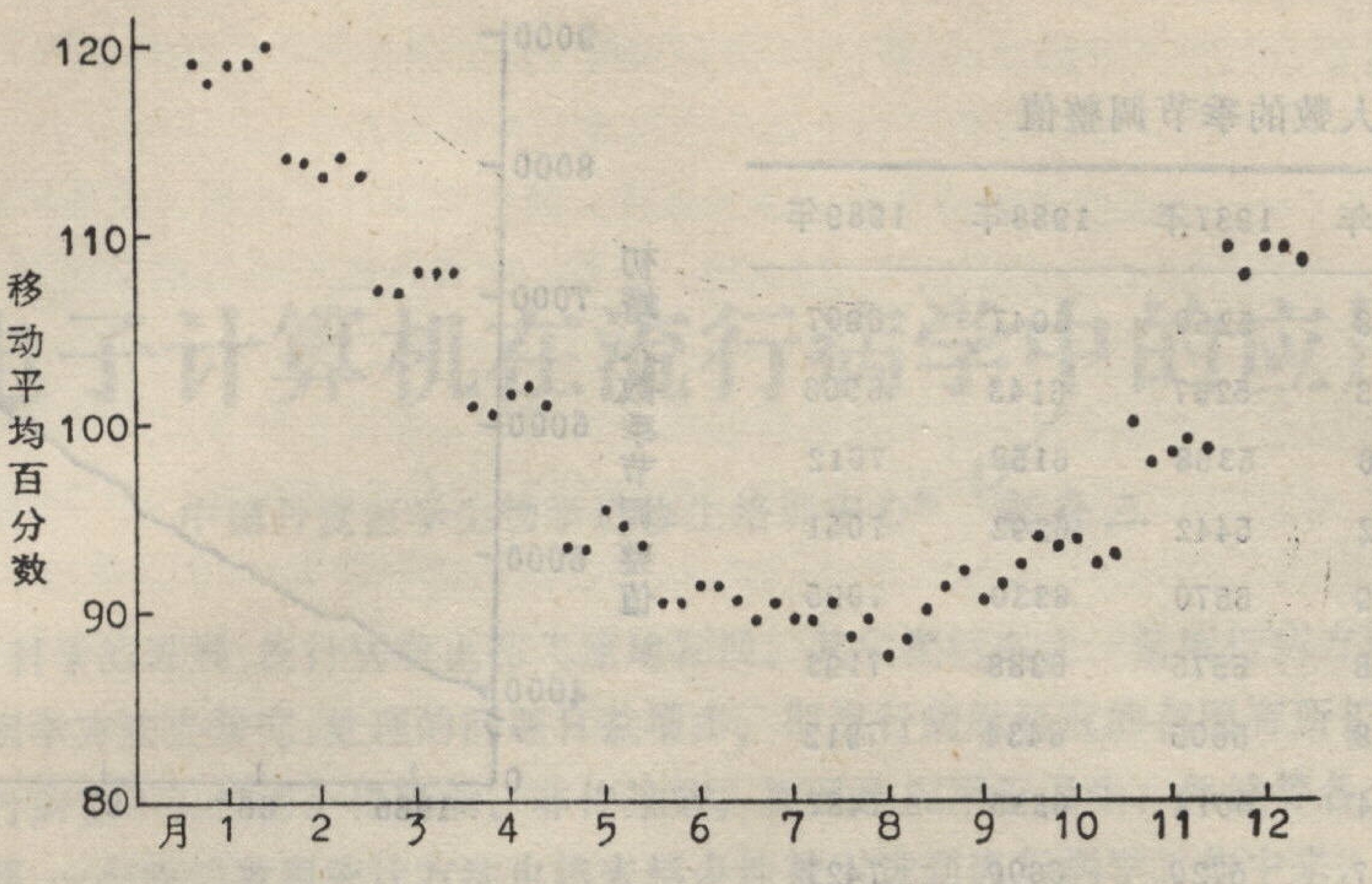

图3　初婚人数季节变化（季节分布图）

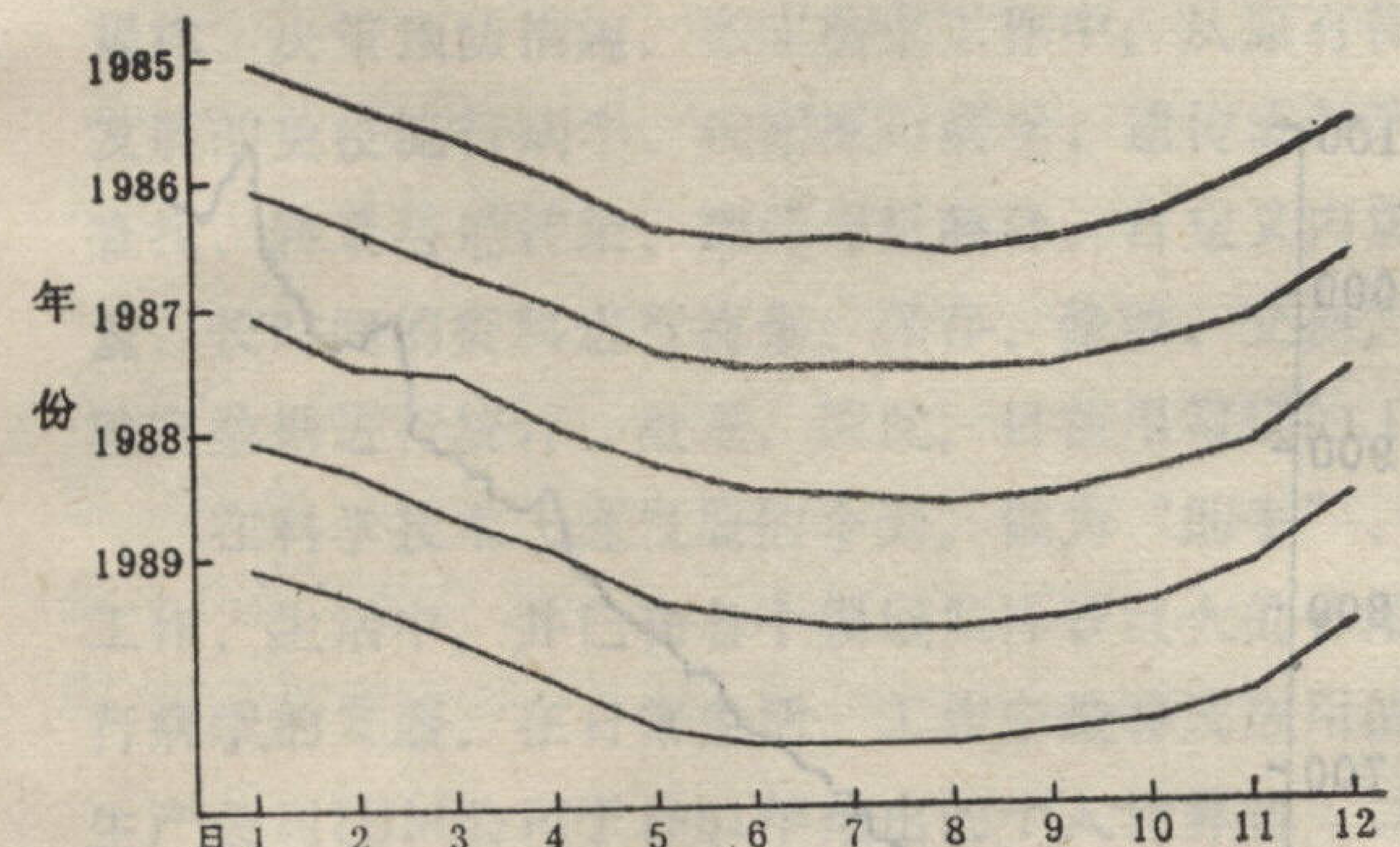

图4　初婚人数季节变化（阶层图）

现的极端值影响，因此一般认为中位数更好一些。另一个更好的办法是去掉最大值和最小值后再求平均值，这就是中心项均值。为了使该指标更直观、更易解释，常把中心项均值的总和调整到1200（因为是12个月，所以有12×100＝1200），把每个月份的中心项均值乘以系数K即可：

$$K=\frac{1200}{\text{中心项均值之和}}$$

对本例，1月份五个数值中118.61最小，120.16最大，去掉这两个再求和，得

$$119.67+119.44+119.28=358.39$$

把358.39除以3，即得中心项均值119.46。12个中心项均值之和是1199.22，故得

$$K=\frac{1200}{1199.22}=1.000650$$

用K乘以各中心项均值就可得调整的季节指数：

$$1.000650\times 119.46=119.54$$

$$1.000650\times 114.16=114.23$$

五、季节指数的应用：季节指数有两个用途；一

表5　季节指数计算用表

月份	中心项合计	中心项均数	调整的季节指数	(4)−100
(1)	(2)	(3)	(4)	(5)
1	358.39	119.46	119.54	19.54
2	342.49	114.16	114.23	14.23
3	322.93	107.64	107.71	7.71
4	304.16	101.39	101.46	1.46
5	281.79	93.93	93.99	−6.01
6	272.77	90.92	90.98	−9.02
7	269.28	89.76	89.82	−10.18
8	266.56	88.85	88.91	−11.09
9	275.48	91.83	91.89	−8.11
10	280.19	93.40	93.46	−6.54
11	296.30	98.77	98.83	−1.17
12	327.32	109.11	109.18	9.18
合计		1199.22	1200.00	

是分析，二是模拟。

1.利用季节指数计算季节调整值，以分析有无异乎寻常的季节变动。

$$\text{季节调整值}=\frac{\text{原始观察值}}{\text{调整的季节指数}}\times 100$$

例如，从表3查得1985年1月和2月的原始观察值为4 777和4 633，从表5查得1月和2月的调整的季节指数为119.54和114.23，因此分别得季节调整值为：

$$\frac{4777}{119.54}\times 100=3996$$

$$\frac{4633}{114{,}23}\times 100=4056$$

由此可得表6。

表6 初婚人数的季节调整值

月份	1985年	1986年	1987年	1988年	1989年
1	3996	4538	5250	6047	6907
2	4056	4613	5267	6143	6906
3	4077	4646	5368	6158	7012
4	4099	4692	5442	6292	7051
5	4163	4749	5570	6330	7095
6	4193	4808	5575	6388	7193
7	4256	4918	5605	6434	7312
8	4301	4974	5614	6535	7487
9	4373	5027	5720	6600	7421
10	4464	5081	5866	6619	7437
11	4548	5092	5872	6791	7548
12	4542	5133	6008	6821	7565

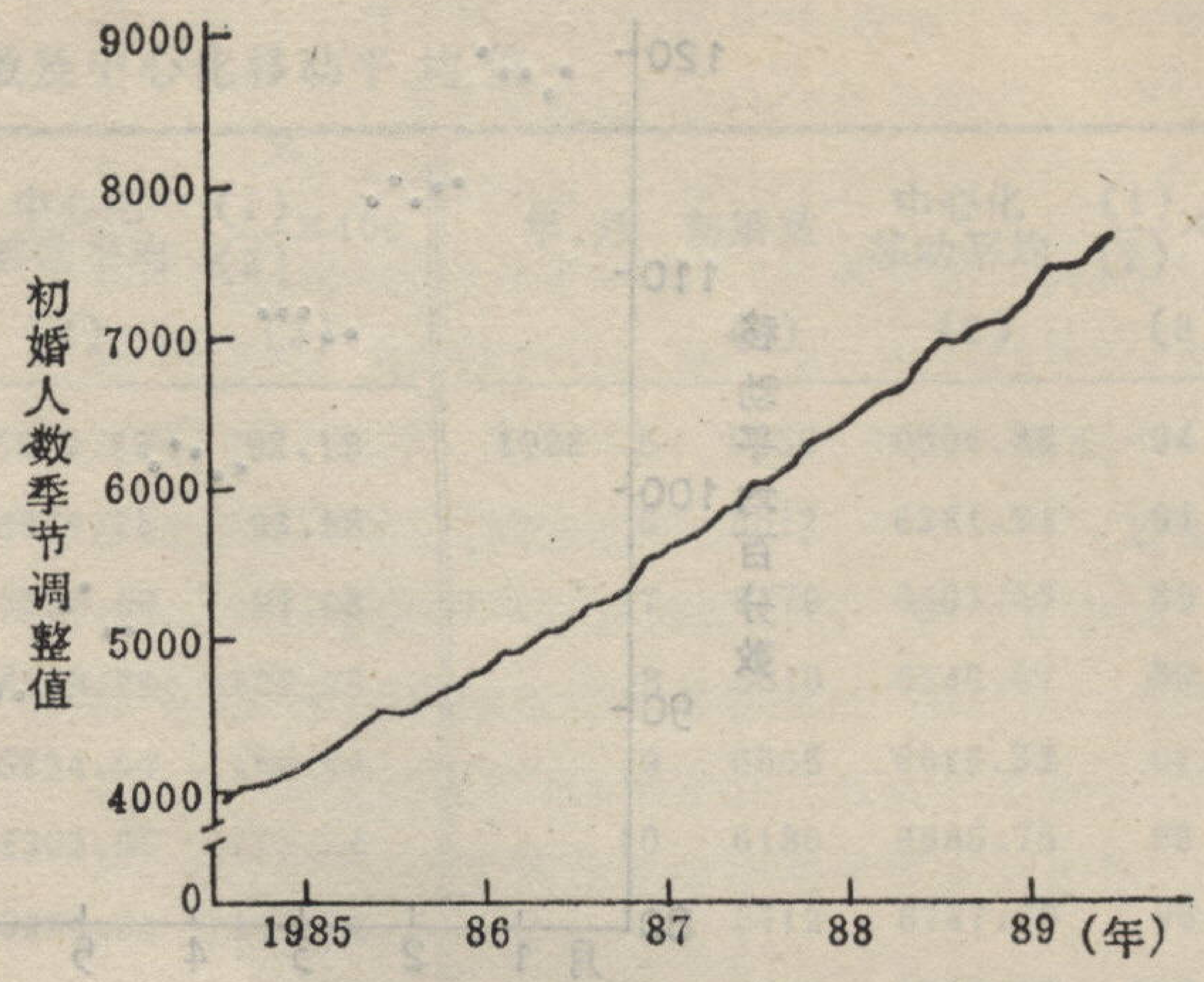

图5 初婚人数的季节调整值

用表6数据可画成图5,以观察有无不同于寻常的季节变动。在计算过程中，季节调整值已消除了一般季节因素的影响,如果各年度具有相仿的季节影响因素，则在图中就不再呈现季节波动。图5表明，这几年的季节变动是近乎一致的。

在另一份资料中，计算了某地7年来各月食物中毒人数的季节调整值，并用此画成了图6。从图6可见,从1986年开始，似乎有一种新的因素在影响季节变化。因为在消除了通常的季节变动影响因素的作用后，在后三年又出现了新的季节变化规律。

2.利用计算所得的趋势线和季节指数，可以模拟和预报某指标的变化情况。预报或预测方法将另作讨论。

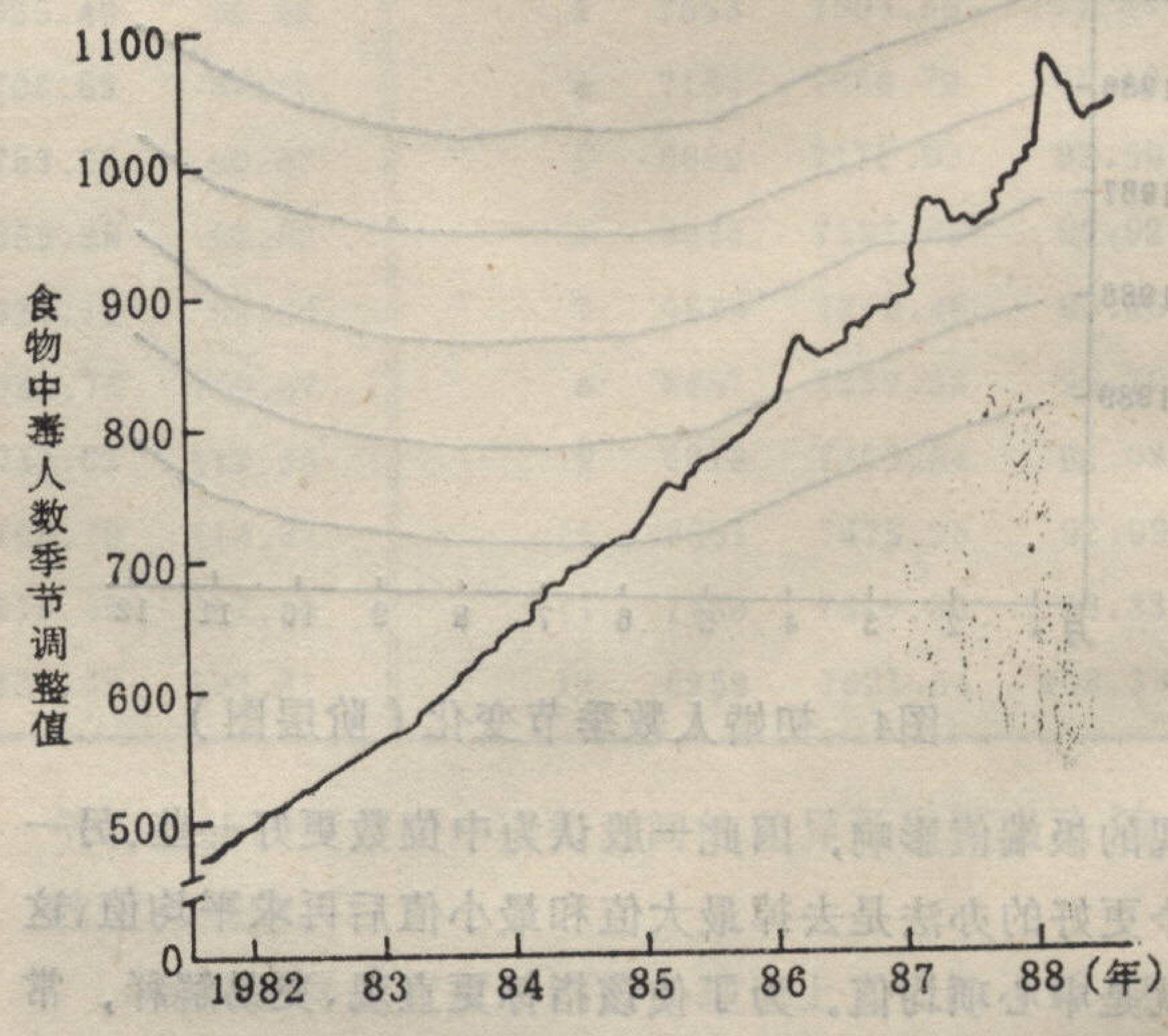

图6 食物中毒人数的季节调整值

·书 讯·

《细胞素抗癌疗法》

本书由第二军医大学杜平教授和戚中田副教授主编。该书是以Prances R. Balkwill 1989编著的《癌症的细胞素治疗》(Cytokines in Cancer Therapy)为蓝本,并收集了近2年来有关最新资料详编而成。是国内目前介绍癌症生物免疫疗法的一本唯一专著。由上海科技出版社出版，约15万字。来函订阅请寄**上海第二军医大学微生物学教研室(邮政编码200433)张晓萍收**。

《肿瘤坏死因子及相关细胞毒因子》

本书由第二军医大学微生物学教研室积10余年研究INF和各类淋巴因子的经验和知识编写而成，由余伟民讲师和焦炳华副教授主编。全书约30万字，共16章，定价7.0元(含邮资)，上海科技出版社出版。订阅来信请寄**上海第二军医大学微生物学教研室周炳荣同志收(邮政编码200433)**。

(谢正旸 供稿)