

第二章

经济预测数据的收集与处理技术



数据来源：统计资料、调查资料

一、数据的分类

- 纵向数据：时间序列数据

同一统计指标，由同一统计单位按时间顺序记录

- 横向数据：横截面数据

同一统计指标，在同一时间，按不同统计单位记录

- 混合数据

- 虚拟变量数据

二、数据的整理与加工

- **Step1:** 审查资料是否完整、合理，详细程度是否合乎要求
- **Step2:** 判断是否有异常数据

(1) 根据经验进行判断

例：某厂产品销售数据

1	2	3	4	5	6	7	8	9	10	11	12
320	390	150	350	340	490	380	375	370	400	420	410

(2) 用滤波的方法鉴别：利用正态分布来确定数据允许的变动范围，超出该范围的即为异常数据

$$Y_{\text{上(下)}} = \bar{Y} \pm k \cdot s$$

其中： \bar{Y} 为样本均值

s 为样本标准差（均方差 $\sigma = \sqrt{D\xi} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$)

k 由样本数量 n 及概率 P_1 ， P_2 ， 查 k 值表得到

K 值表

<div> <div> <div>P_1</div> <div>P_2</div> </div> <div> <div>k</div> <div>n</div> </div> </div>		0.95		0.99	
		0.99	0.999	0.99	0.999
10					
.					
.					
.					
30		3.35	4.26		
.					

P_2 : 落在 $\bar{Y} \pm k \cdot s$ 范围内的数据个数与整个样本数据个数的最小百分比;
 P_1 : P_2 的置信度 (或置信水平)

例: 某地区连续**30**年降水量如下 (单位: 毫升)

1243	1236	1230	1240	1251	1266	1273	1252	1301	1274
1245	1328	1362	1333	1371	1275	1264	1282	1310	1304
1323	1352	1845	1628	1350	1347	1344	1320	1383	1360

试找出异常数据。

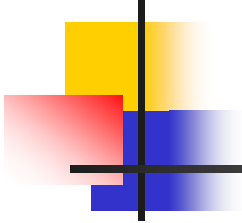
解: $\bar{Y} = 1330$ $s = 123$ $P_1 = 0.95$ $P_2 = 0.99$

$n = 30$ 时, $k = 3.35$

$$Y_{(上)} = 1330 + 3.35 \times 123 = 1742.05$$

$$Y_{(下)} = 1330 - 3.35 \times 123 = 917.95$$

所以, **1845**为异常数据



■ Step3: 对异常数据进行加工

(1) 剔除: 数据多

(2) 插补法

散点图呈水平趋势: 平均值

散点图呈线性趋势:
$$x_k = \frac{x_{k+1} + x_{k-1}}{2}$$

散点图呈非线性趋势:
$$x_k = \sqrt{x_{k+1} \cdot x_{k-1}}$$

(3) 残缺数据修补法: 线性插值法

例：某市工业总产值数据缺**1977**及**1978**年资料

年份	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
产值 (万元)	31.4	33.5	38.6	40.3			53.4	60.1	64.8	70.5

解： $\hat{x}_t = a + bt$

前三期平均值 $\bar{x}_1 = \frac{31.4 + 33.5 + 38.6}{3} = 34.5$

后三期平均值 $\bar{x}_2 = \frac{60.1 + 64.8 + 70.5}{3} = 65.1$

$$b = \frac{\bar{x}_2 - \bar{x}_1}{t_2 - t_1} = \frac{65.1 - 34.5}{9 - 2} = 4.37 \quad a = \bar{x}_1 - bt_1 = 34.5 - 4.37 \times 2 = 25.76$$

$$\hat{x} = 25.76 + 4.37t$$

1977年数据为 $Y_5 = 25.76 + 4.37 \times 5 = 47.61$

1978年数据为 $Y_6 = 25.76 + 4.37 \times 6 = 51.98$

年份	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
产值 (万元)	31.4	33.5	38.6	40.3	47.61	51.98	53.4	60.1	64.8	70.5