



常用数据分析方法

培训主要内容：

- 1、了解数据、总体、样本的含义及随机抽样的一般方法；
- 2、掌握排列图、因果分析图的作图方法和应用；
- 3、掌握分层法、统计图表法的应用；
- 4、了解直方图的原理、作用、作图方法及应用；

	培 训 要 求	
--	---------	--

瞎子国里有一只眼睛的人就是国王！

有一些统计分析工具可供书业公司用来解决经营问题，它们有助于收集和分析数据以便为决策提供依据。

❖ 对帐单（检查表）；

❖ 流程图；

❖ 散布图；

❖ 直方图；

❖ 排列图；

❖ 控制图；

❖ 因果分析图；

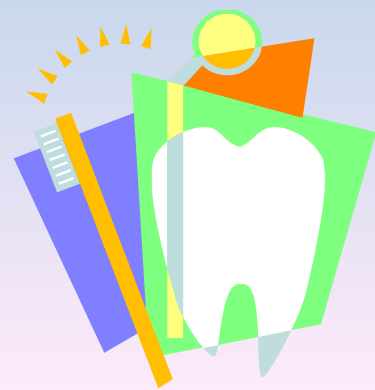
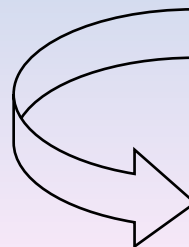
	统计分析工具	
--	--------	--

第一部 数据分析概述

1、什么是数据？

数据是对图书销售业务全过程记录下来的、可以鉴别的符号。数据是销售业务全过程的属性数量、位置及相通关系等等的抽象表示。

数据表现形式



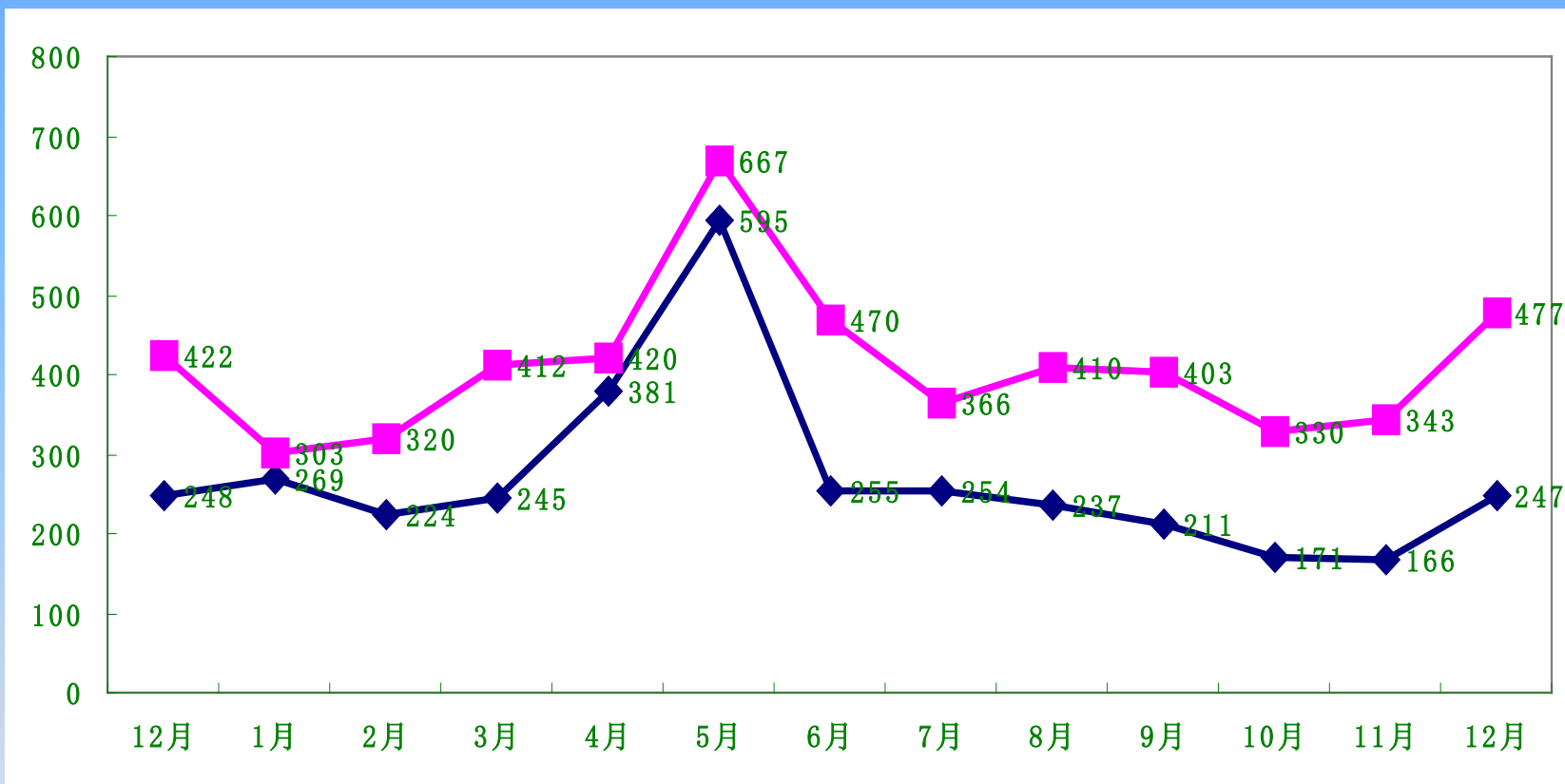
表现形式一

2006年客户销售分析报表									
A客户	12月	1月	2月	3月	4月	5月	6月	7月	8月
B客户	99	98	98	100	101	108	108	114	110
C客户	67	70	73	74	75	77	82	80	80
D客户	87	84	88	85	79	85	91	92	93
E客户	148	143	143	145	137	141	136	134	139
F客户	75	76	76	77	79	78	82	78	78
G客户	87	86	92	100	103	100	97	89	91
H客户	132	134	141	142	160	154	150	136	137
合计	696	692	712	724	735	743	746	723	728

数据
列表

数据表现形式二

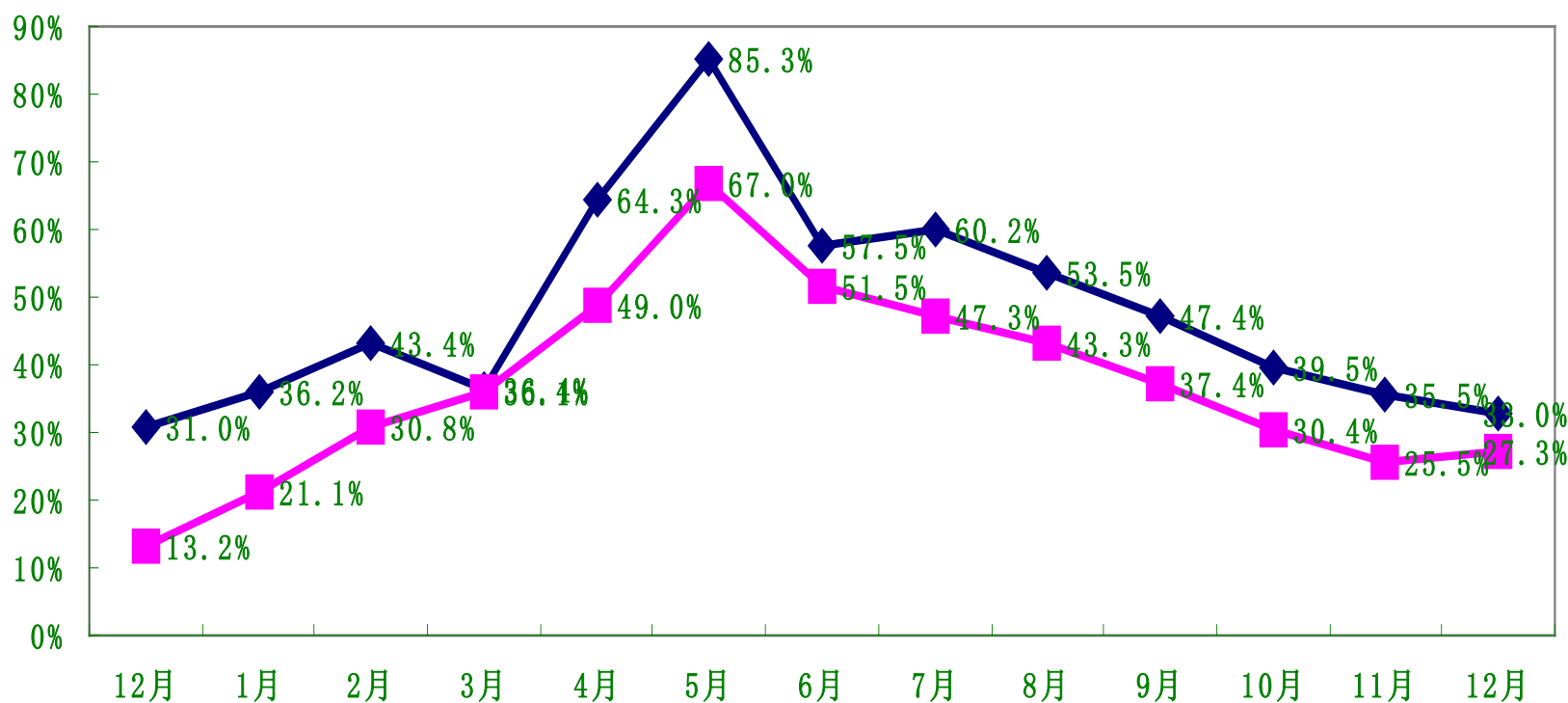
销售码洋趋势图



描述性分析：数据分时段走势、差距变化情况以及相应的时间背景

数据表现形式三

销售码洋结构比



图形可以帮助我们 将数据转换成信息

2、数据分析目的

- ✚ 让数据说话；
- ✚ 行动的向导；
- ✚ 杜绝浪费；
- ✚ 提供决策的依据。

3、数据的分类

✧ 按数据的性质一般可归纳为：

- ✦ 定性数据。往往用于频数或频率分析。
- ✦ 定量型数据。可以用数字量化的数值。

✧ 按数据的来源可分为：

- ✦ 人力资源数据；
- ✦ 财务数据；
- ✦ 营销数据；
- ✦ 采购数据；
- ✦ 仓储数据；
- ✦ 生产数据；
- ✦ 编辑数据。

4、运用统计方法应遵循的原则

 坚持用数据说话的基本观点。

 有目的地收集数据。

 掌握数据的来源。

 认真整理数据。

5、统计分析流程

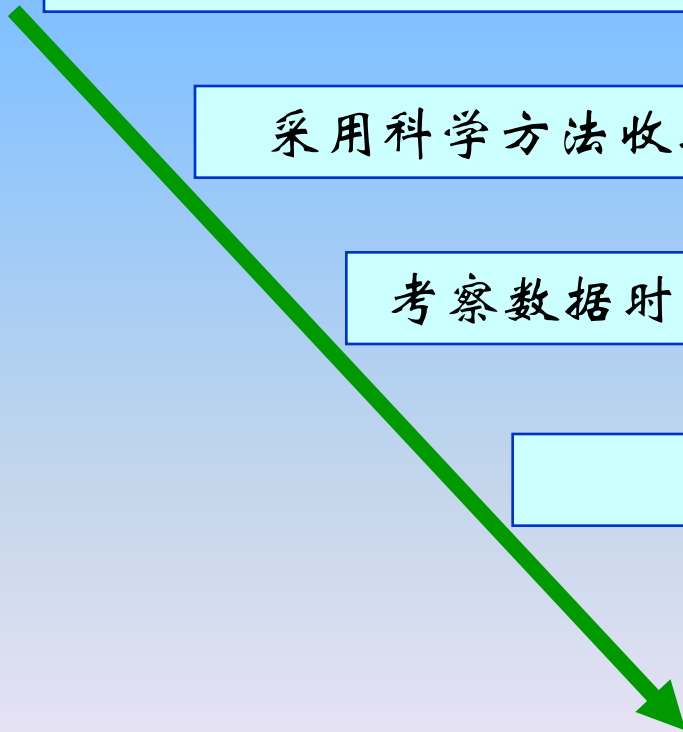
确定问题确定分析目标

采用科学方法收集数据

考察数据时效性整理数据

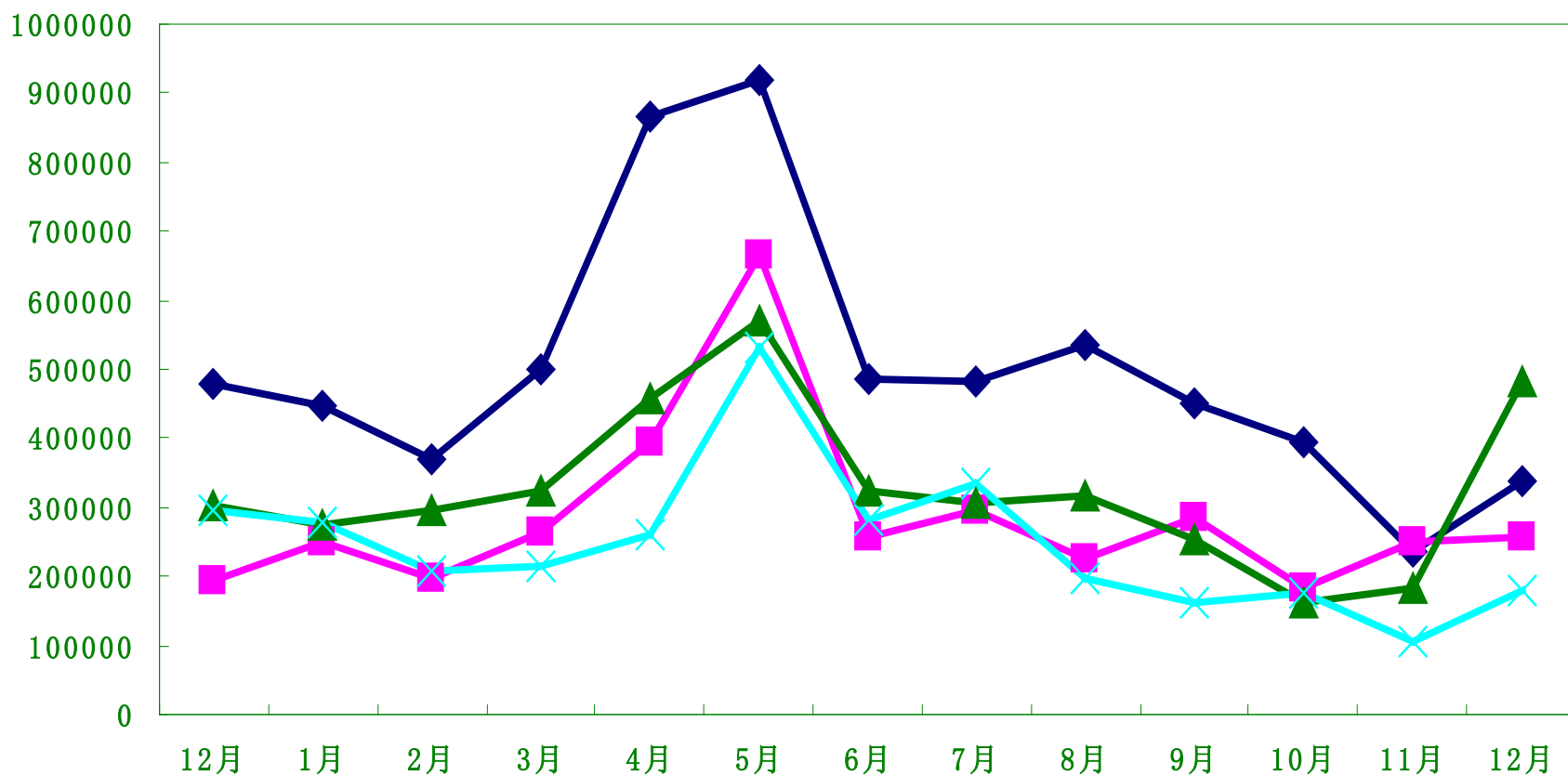
统计分析

出具分析报告，提出解决
意或建议



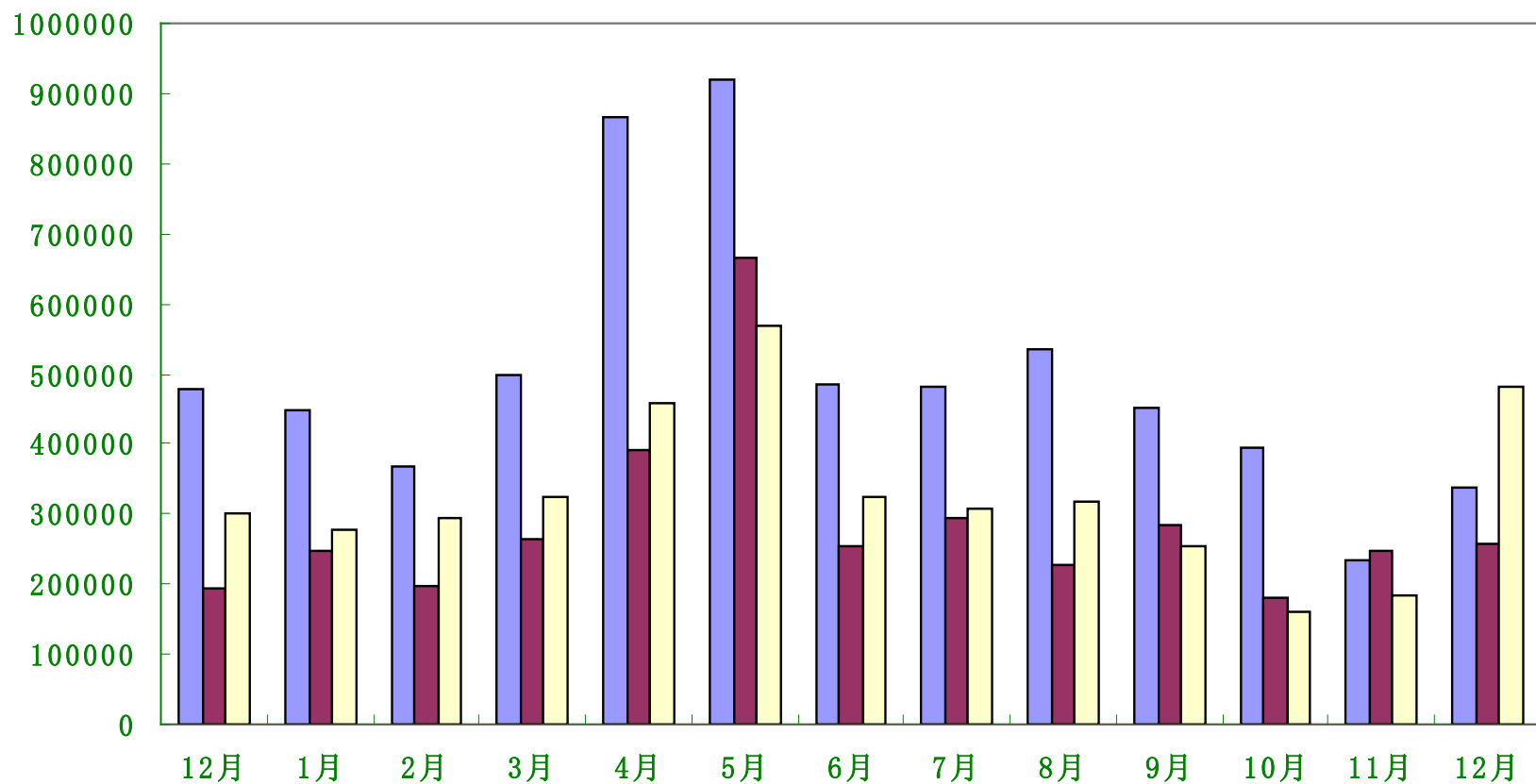
5、数据分析误区

误区一：展示元素不宜大于3个



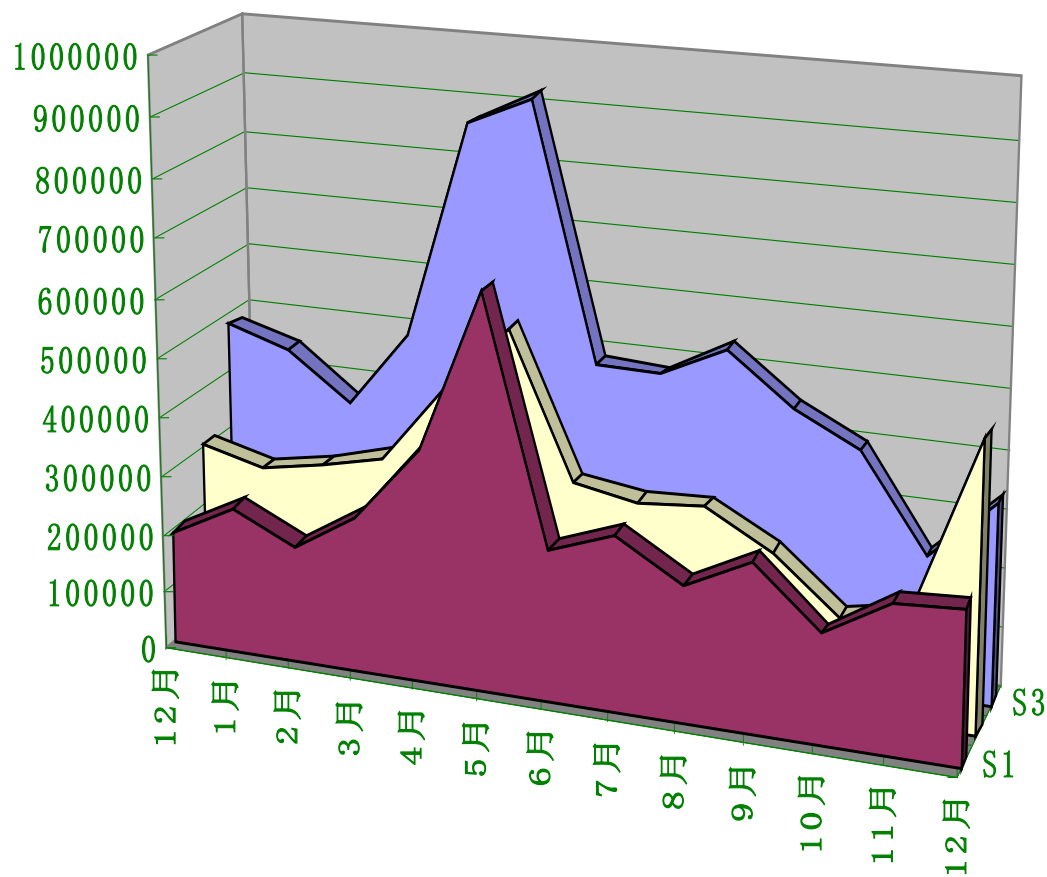
5、数据分析误区

误区二:时间序列数据最好使用折线图, 而不宜使用柱状图等



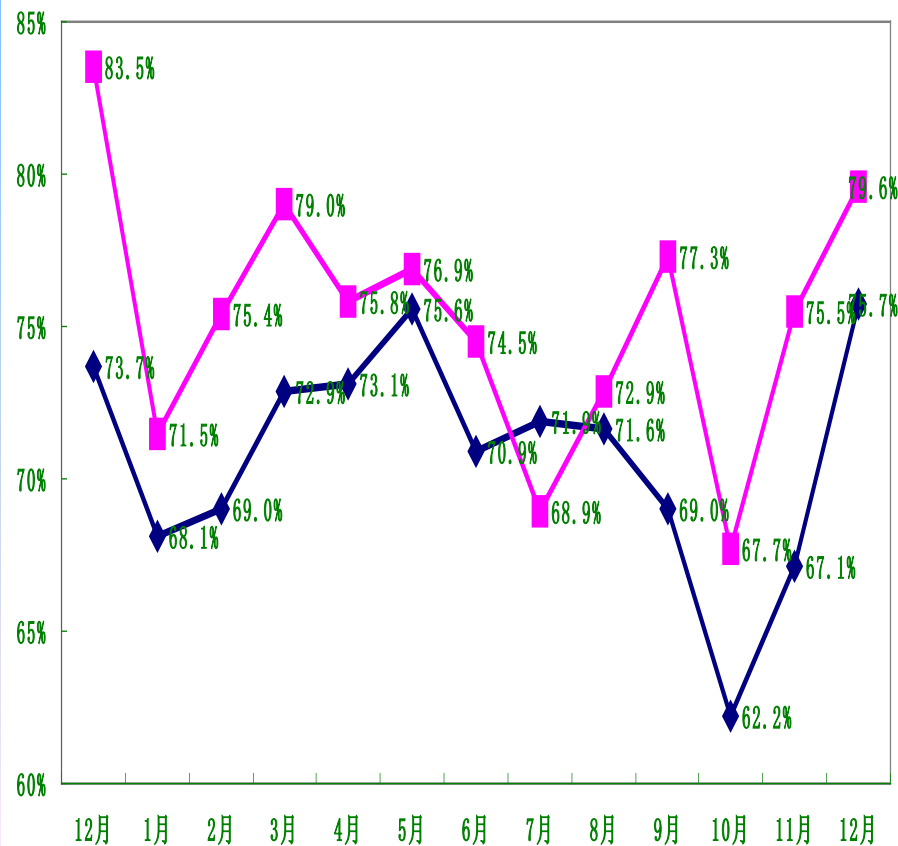
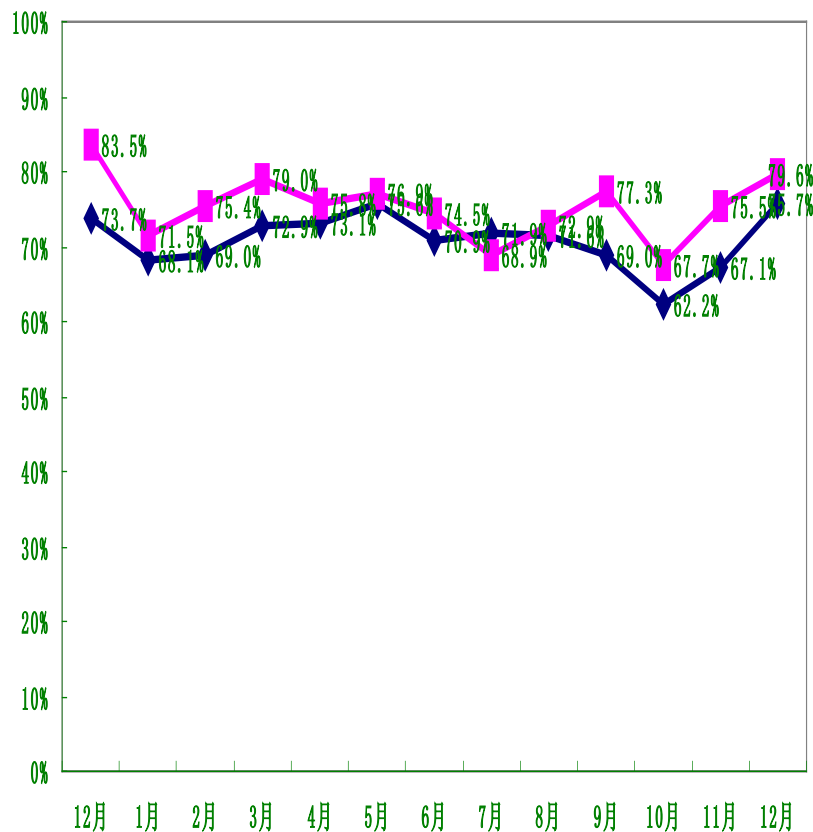
5、数据分析误区

误区三:研究用数据最好不使用三维立体图



5、数据分析误区

误区四：为避免图表的欺骗性，图线最好占据2/3至3/4的高度（调整Y轴刻度）



第二部 常用的数据分析方法

总体：

总体又叫母体，是研究对象的全体。

出版商一个批次到货的**全部**都可以称为总体。

个体：

构成总体的基本单位，称为个体。

每册书都是一个个体。

来货检验常用抽样方法进行，即从来货总件数中抽出一部分件数，并测试每件的有关册数是否够数的特性数据，进行统计分析后，对总体作出估计和判断。

	总体、个体	
--	-------	--

样本：

样本又叫子样，是从总体中抽出来一部分个体的集合。

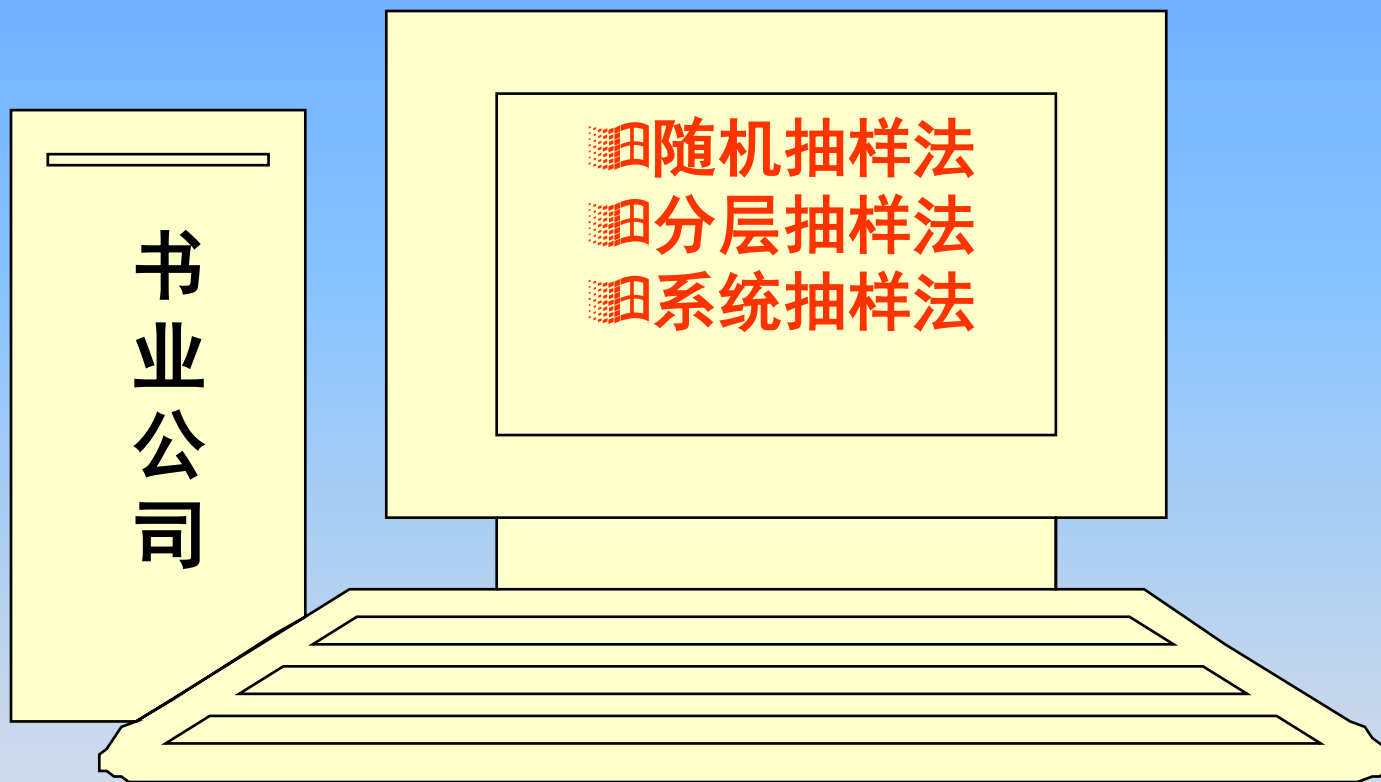
样本中每个个体叫样品，样本中所包含样品数目称为样本大小，又叫样本量，常用 n 表示。

对样本的特性进行测定，所得的数据称为样本值。

当样本个数越多时，分析结果越接近总体的值，样本对总体的代表性就越好。

	样 本	
--	-----	--

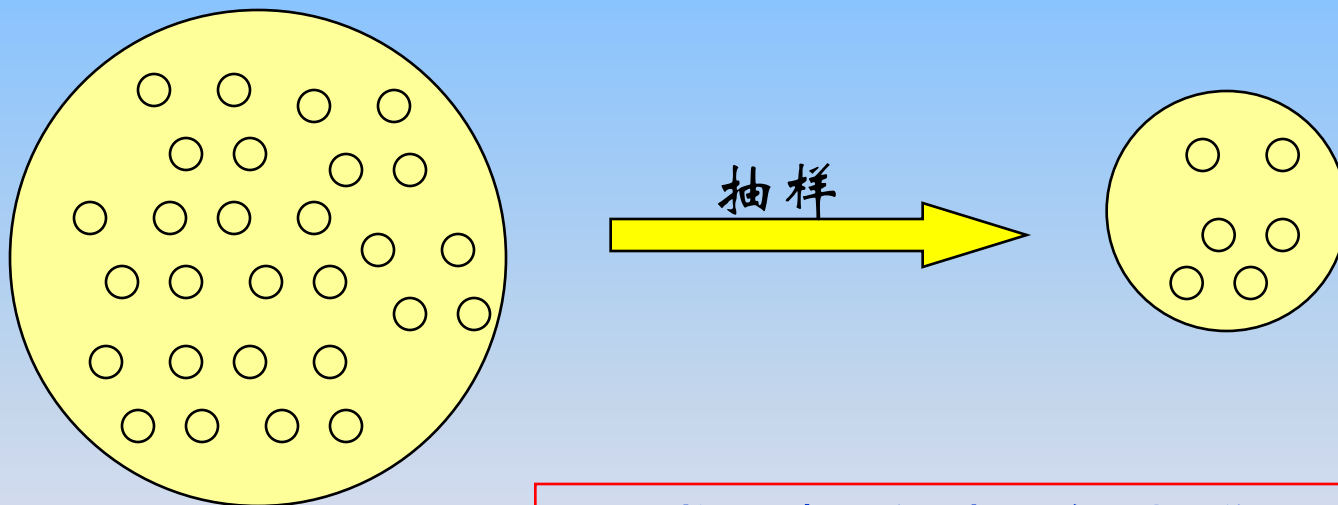
学用的统计抽样方法主要有以下三种



抽 样 方 法

随机抽样

指总体中每一个个体都有同等可能的机会被抽到。这种抽样方法事先不能考虑抽取哪一个样品，完全用偶然方法抽样，常用抽签或利用随机数表来抽取样品以保证样品代表性。

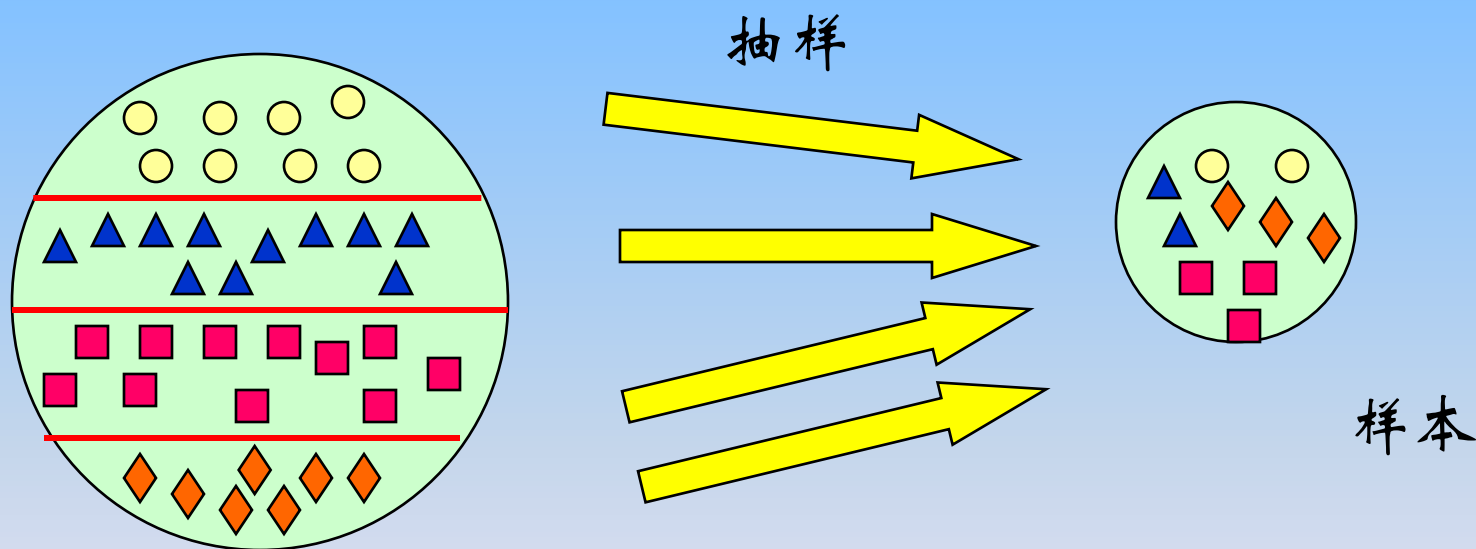


当图书品种不多时，随机抽样是一种有效的抽样方法；

抽 样 方 法

分层抽样

分层抽样是先将总体按照研究内容密切有关的主要因素分类或分层，然后在各层中按照随机原则抽取样本。分层抽样可以减少层内差异，增加样本的代表性。



当到货产品较多时，分层抽样是一种有效的抽样方法；

抽样方法

系统抽样

从总体中每隔K个个体抽取一个个体的抽样方法，比值K是总体容量N与样本容量n之比；

例，从具有1000个个体的总体中抽取50个个体。

1, 2, K

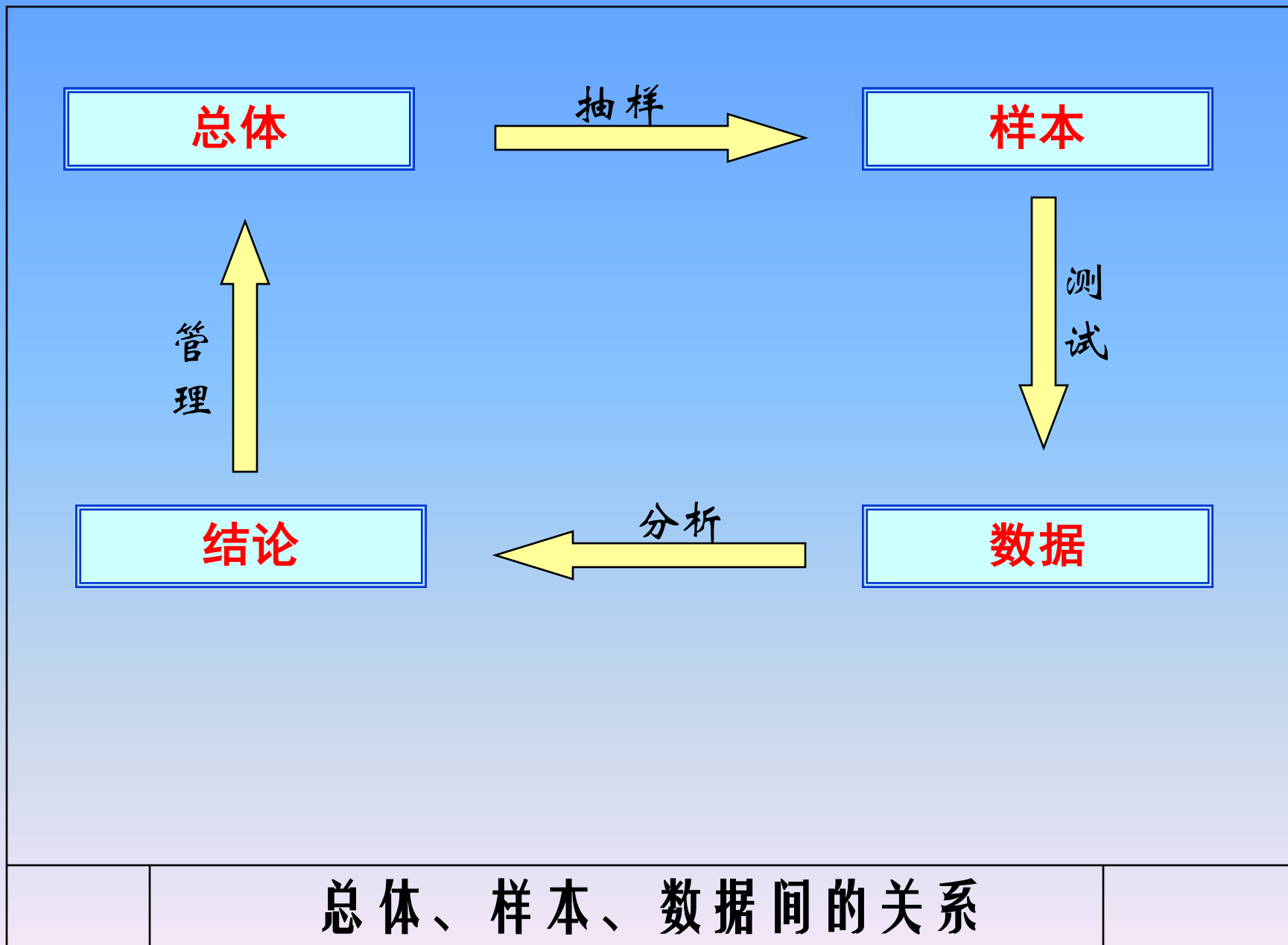
K+1, K+2,, 2K

2K+1, 2K+2,, 3K

直到 N为止

当出版商批量发货及产品特别多时，并且易作某种次序的整理时，系统抽样比分层抽样好；

	抽 样 方 法	
--	---------	--



抽样的目的是通过样本来反映总体。

在书业公司经营管理中，常常将测试的样本数据，通过整理加工，找出它们的特性，从而推断总体的变化规律、趋势和性质。

一批数据的分布情况，可以用中心倾向及数据的分散程度来表示，表示中心倾向的有平均值、中位值等，表示数据分散程度的有方差、标准偏差、极差等。

描述总体数据离散程度的参数为方差 σ^2 ，描述总体数据中心倾向的数为均值 μ 。若利用样本参数近似描述总体状况时，可以利用样本方差 S^2 近似代替总体方差 σ^2 ，利用样本均值 \bar{X} 近似代替总体均值 μ 。

样本平均值

$$\bar{X} = \frac{X_1 + X_2 + X_3 \dots\dots\dots + X_n}{n}$$

样本中位值

中位值是按照数据大小顺序排列位于中间的数值，中位值记为 \tilde{X}

若n为偶数，则取位于中间两个数值的平均值为中位值；

样本极差

样本极差表示一组数据分布的范围，是指数据中最大值与最小值的差：

$$R = X_{\max} - X_{\min}$$

样本方差和样本标准偏差

样本方差和样本标准差就是用来度量数据波动幅度大小的一个重要特性值。样本方差是一组数据中每一个数值与平均值之差的平方和的平均值，通常记为 S^2 ；样本方差的平方根 S 称作样本标准偏差，它与样本方差一样，是反映一组数据分散程度的特性值：

在书业公司日常管理中**最强调的是事实管理**，就是要掌握事实，要掌握事实就必须设计检查表收集数据。

业务记录用检查表

检查项目	8月1日	8月2日	8月3日	8月19日	合计
报订	2	4				
开单	10	13				
拣货	2	8				
返单	4	8				
收货	1	2				
合计	19	35				
检查数	100	100				
不良率	19	35				

检 查 表

层别法是所有手法中最基本的概念，即将多种多样的数据，因应用目的的需要分类成不同的“类别”，使之方便以后的分析；

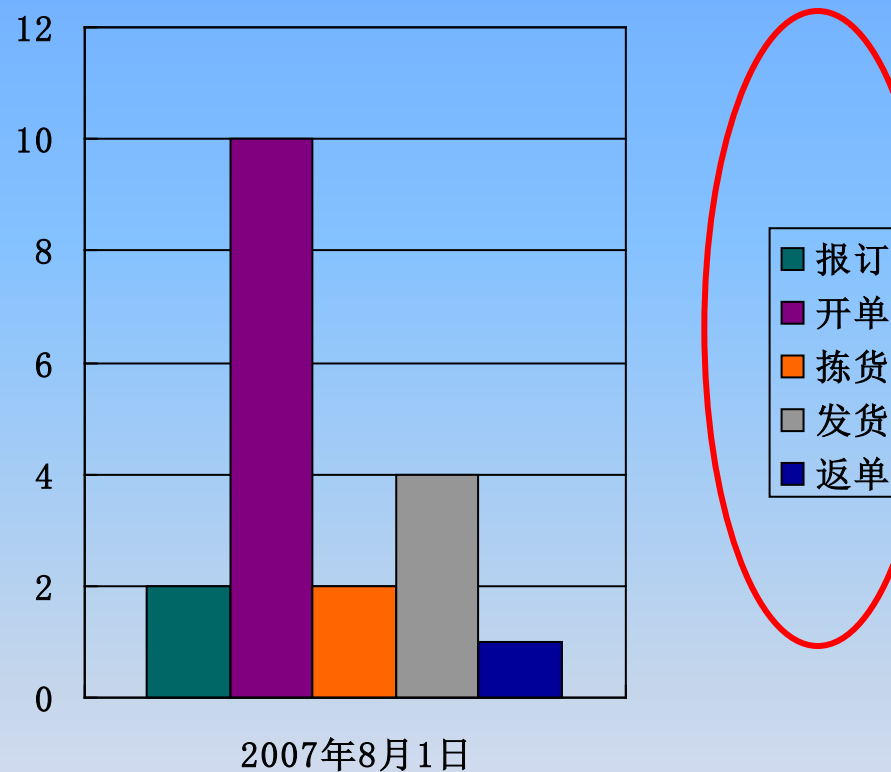


层 别 法

用在检查表上

不良项目	8月1日
报订	2
开单	10
拣货	2
发货	4
返单	1
合计	19
检查数	100
不良率	19

用在排列图上



层别法

排列图（帕累拉图）

- ✚关键的少数，次要的多数。
- ✚通常所说的80/20法则。
- ✚针对问题所收集的数据按其重要性的顺序，记录在帕累托分析表中。
- ✚公司80%的销售来自于20%教辅图书产品。

公司要解决的问题很多，但从何入手呢？

事实上，大部分的问题，只要能找出几个影响较大的因素，并加以处置及控制，就可解决问题的80%以上。柏拉图是根据收集的数据，以不良原因、不良状况发生的现象，有系统地加以项目别分类，计算出各项目所产生的数据（如作业准确率、损失金额）及所占的比例，再依照大小顺序排列，再加上累积值的图形

排列图的作用

产品销售排列图

销售名次	产品名称	销售码洋	所占%	备注
合计				

- ① 将用于排列图所记录的数据进行分类。
- ② 确定数据记录的时间。
- ③ 按分类项目进行统计。
- ④ 计算累计频率。
- ⑤ 准备坐标纸，画出纵横坐标。
- ⑥ 按频数大小顺序作直方图。
- ⑦ 按累计比率作排列曲线。
- ⑧ 记载排列图标题及数据简历。

	排列图的作图方法步骤	
--	------------	--

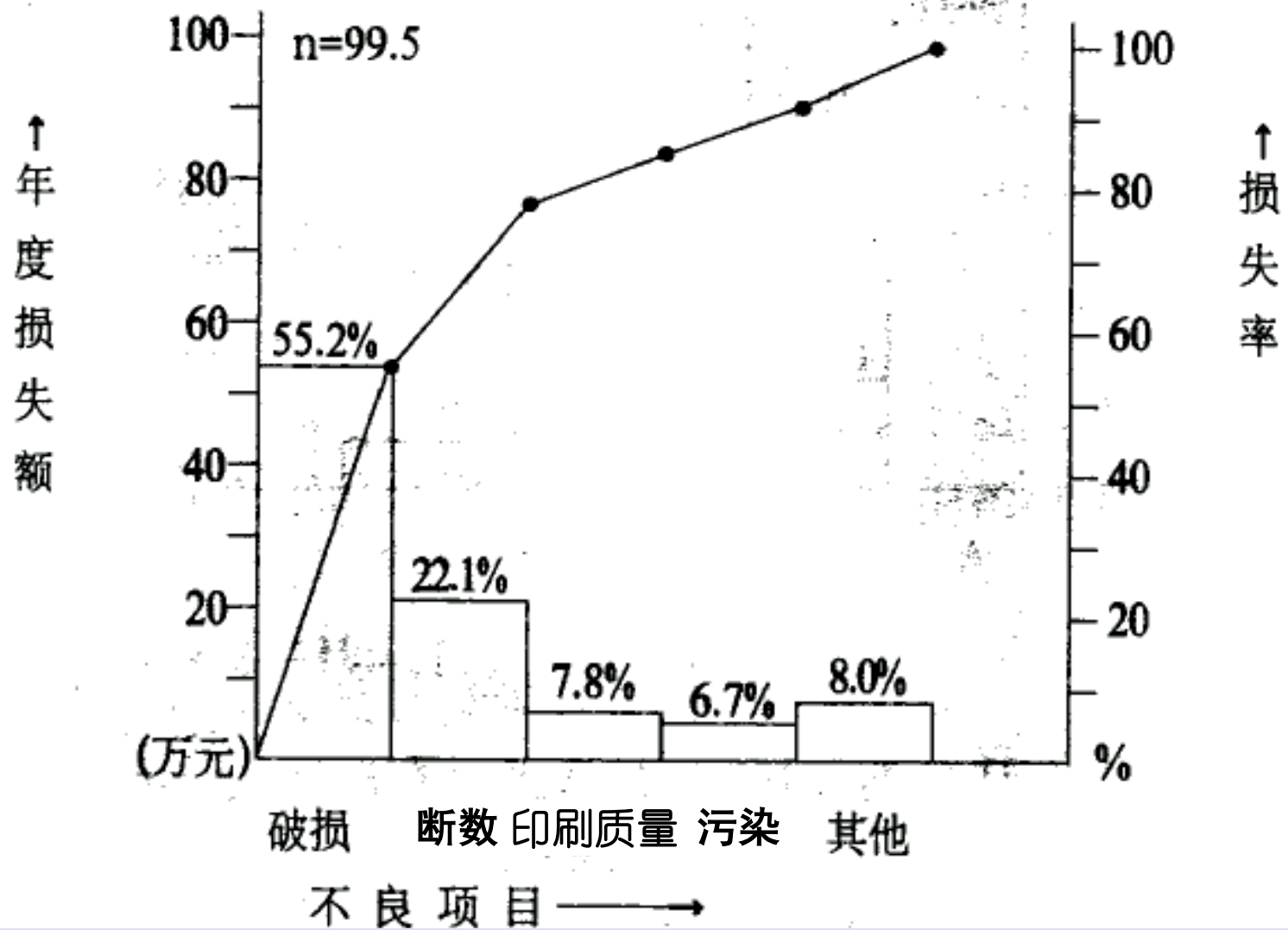
如一公司由物流中心人员操作问题，
导致图书损耗。记录两季教辅销售期
所产生库存损耗不良情况数据，并分
别将不良项目归结为下表

	排列图：例1	
--	--------	--

物流中心由于操作问题导致图书损耗

顺 位	不良项目	年度 损失额 (万元)	损 失 百 分 比 (%)	
			本 项 目	累 计
1	破 损	55	55.2	
2	断 数	22	22.1	77.3
3	印刷质量	7.8	7.8	85.1
4	污染	6.7	6.7	91.8
5	其 他	8	8.0	100
6	合 计	99.5	100	

排列图：例2



排列图：例2

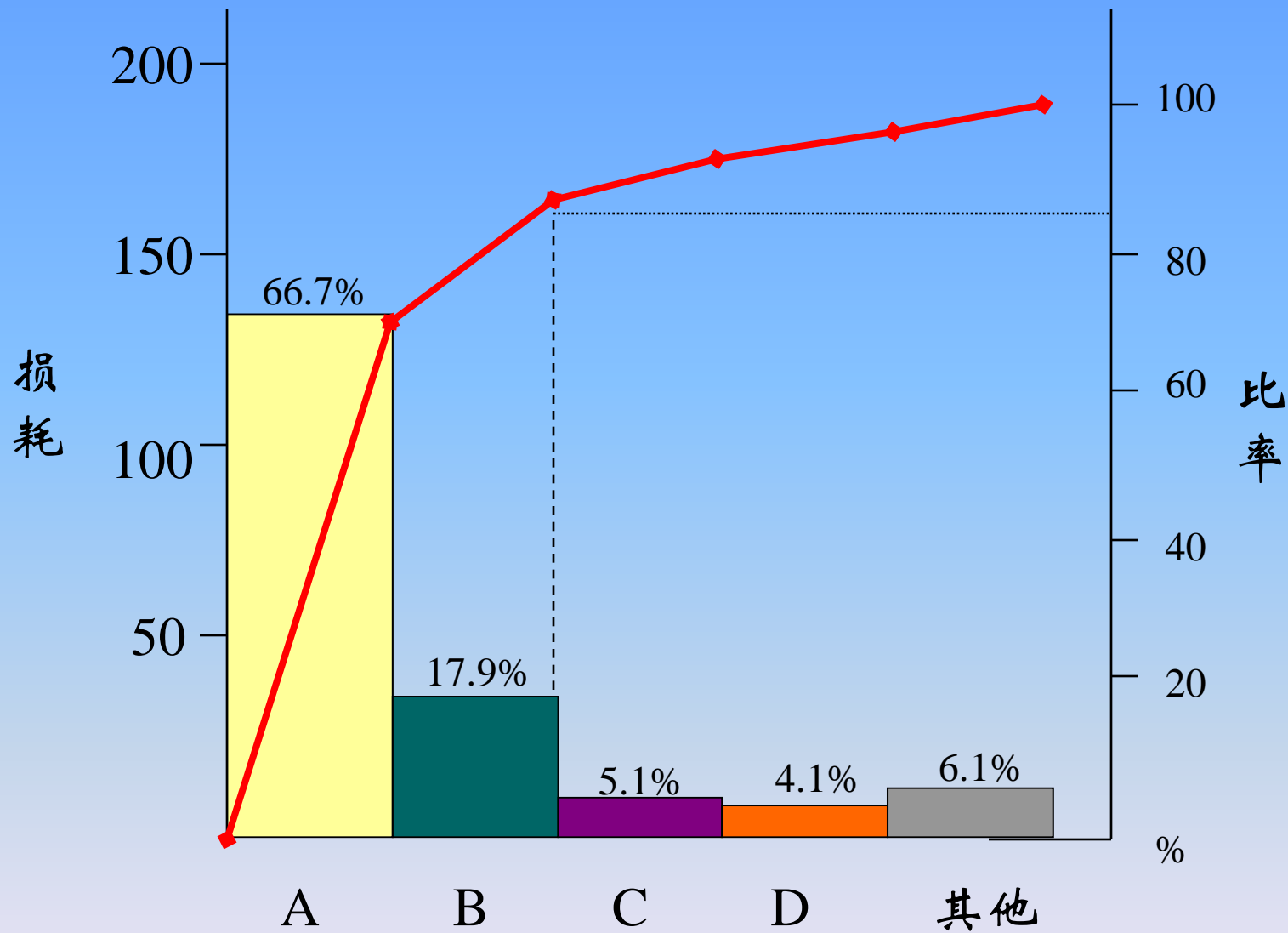
上例中物流库存损耗项目，此损耗为全年损耗码洋的总和，再将产品类别用柏拉图法分析如下：

序号	产品	损耗	占损耗比率(%)	累积比率(%)
1	A	130	66.7	66.8
2	B	35	17.9	84.7
3	C	10	5.1	89.8
4	D	8	4.1	93.9
5	其他	12	6.1	100
合计		195	100	

排列图：练习

序号	产品	损耗	占损耗总数比率(%)	累积比率(%)
1	A	130	66.7	
2	B	35	17.9	84.6
3	C	10	5.1	89.7
4	D	8	4.1	93.8
5	其他	12	6.2	100
合计		195	100	

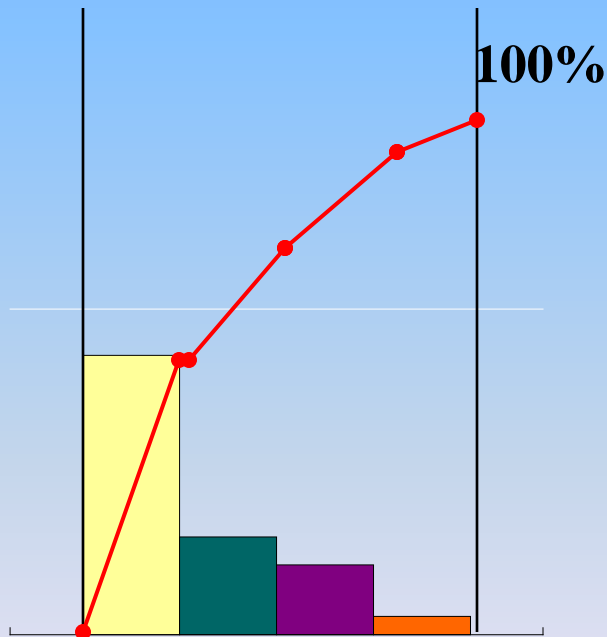
排列图：练习



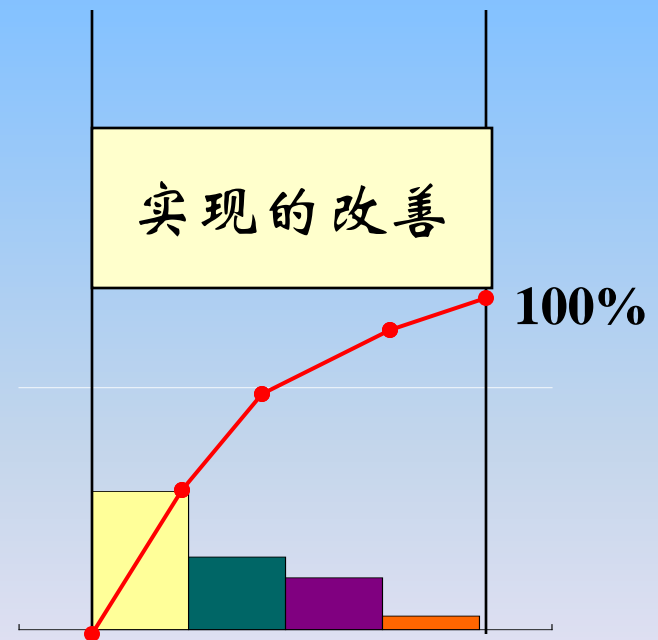
1、利用排列图寻找图书损耗的改善重点；

2、利用排列图验证改善图书损耗的效果；

之前

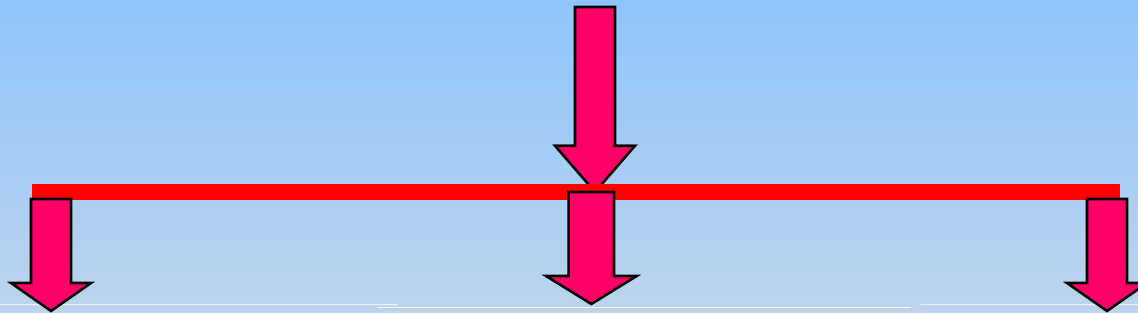
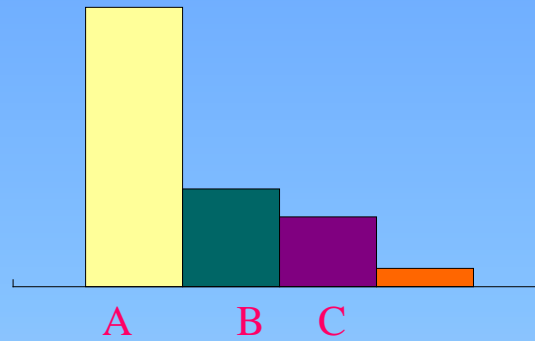


之后



排列图的应用

3、利用排列图对图书损耗进行分层研究；

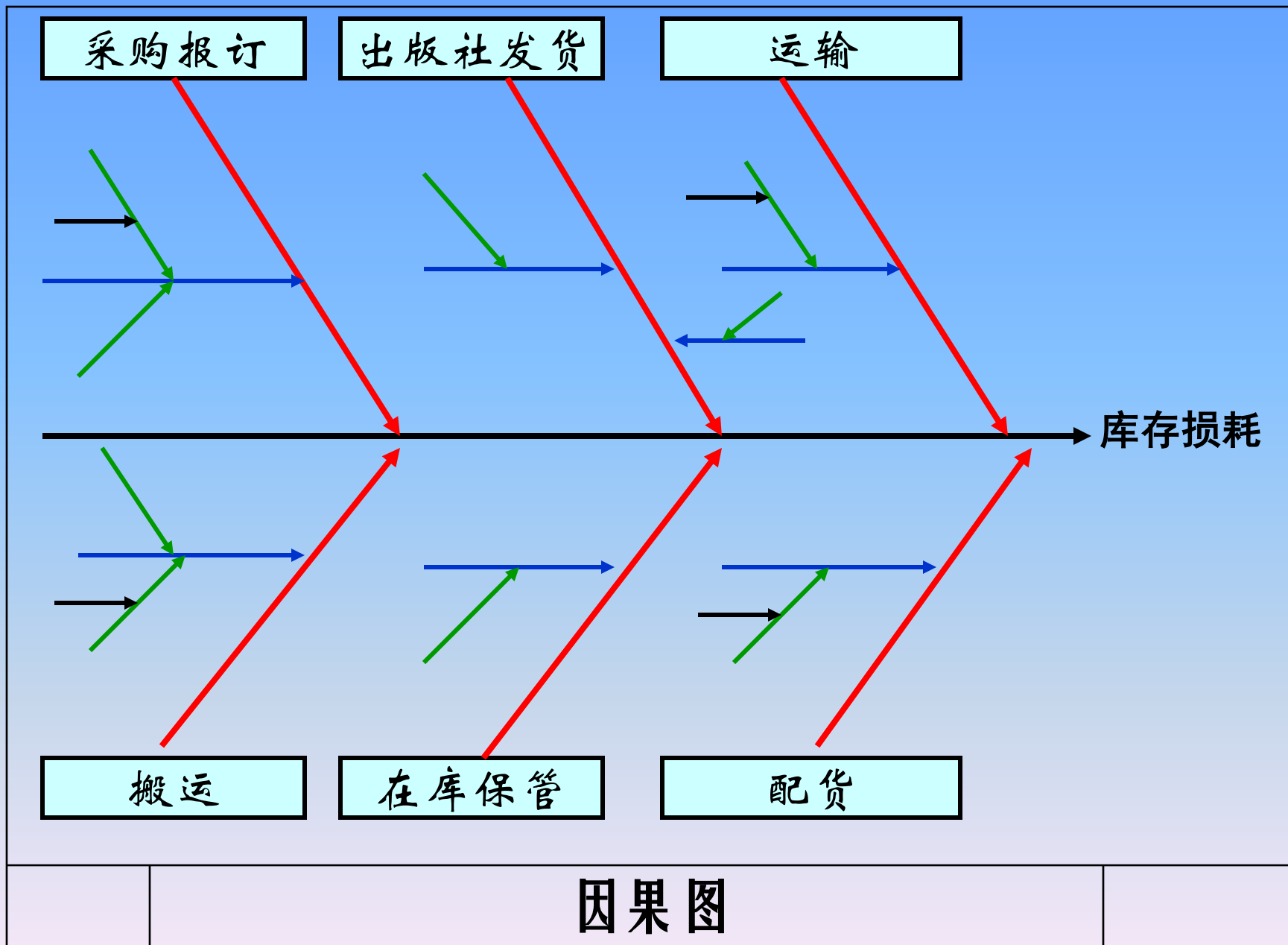


排列图的应用

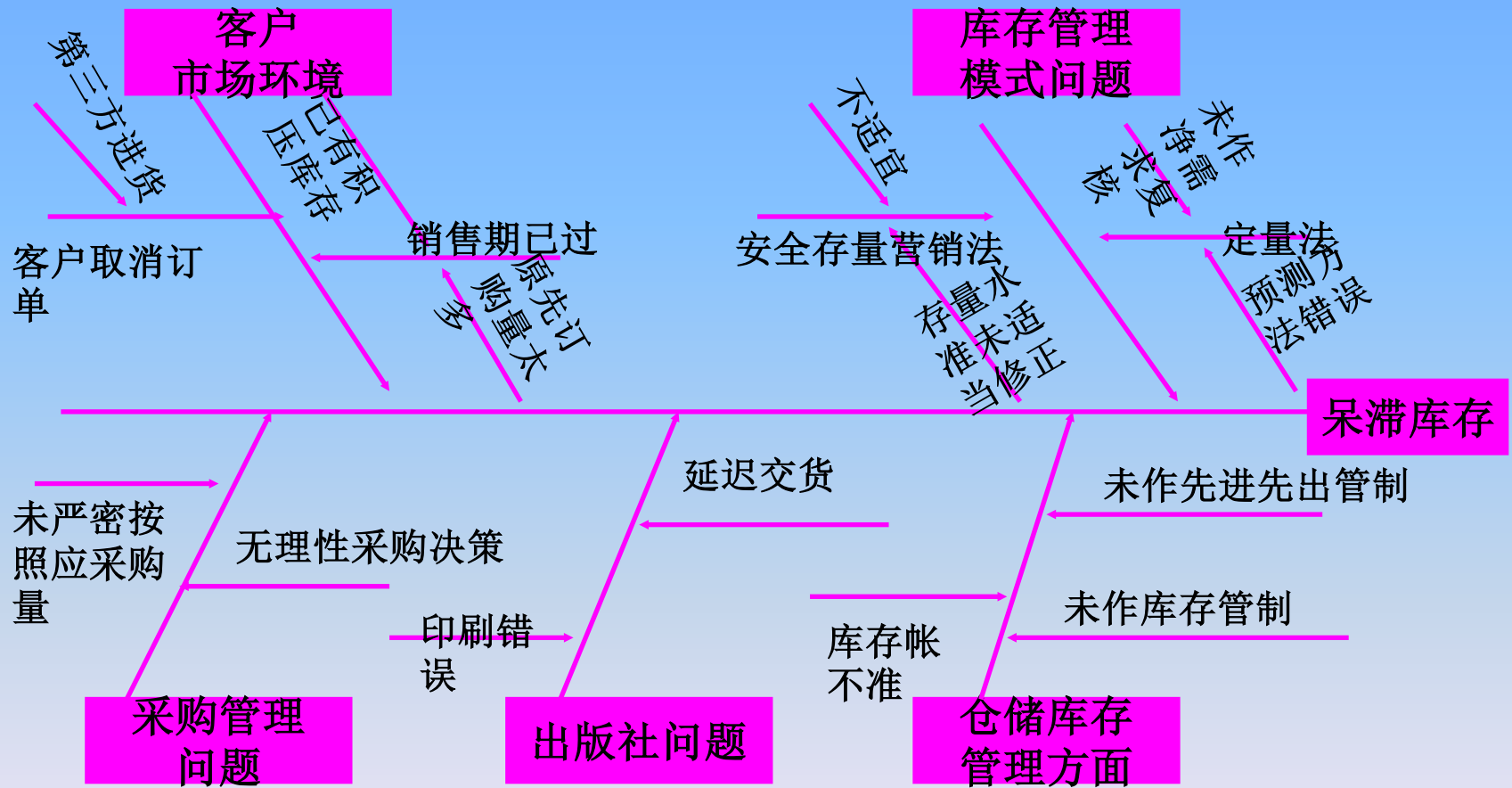
因果图

某项结果之形成，必定有其原因，应设法利用图解法找出其原因来。

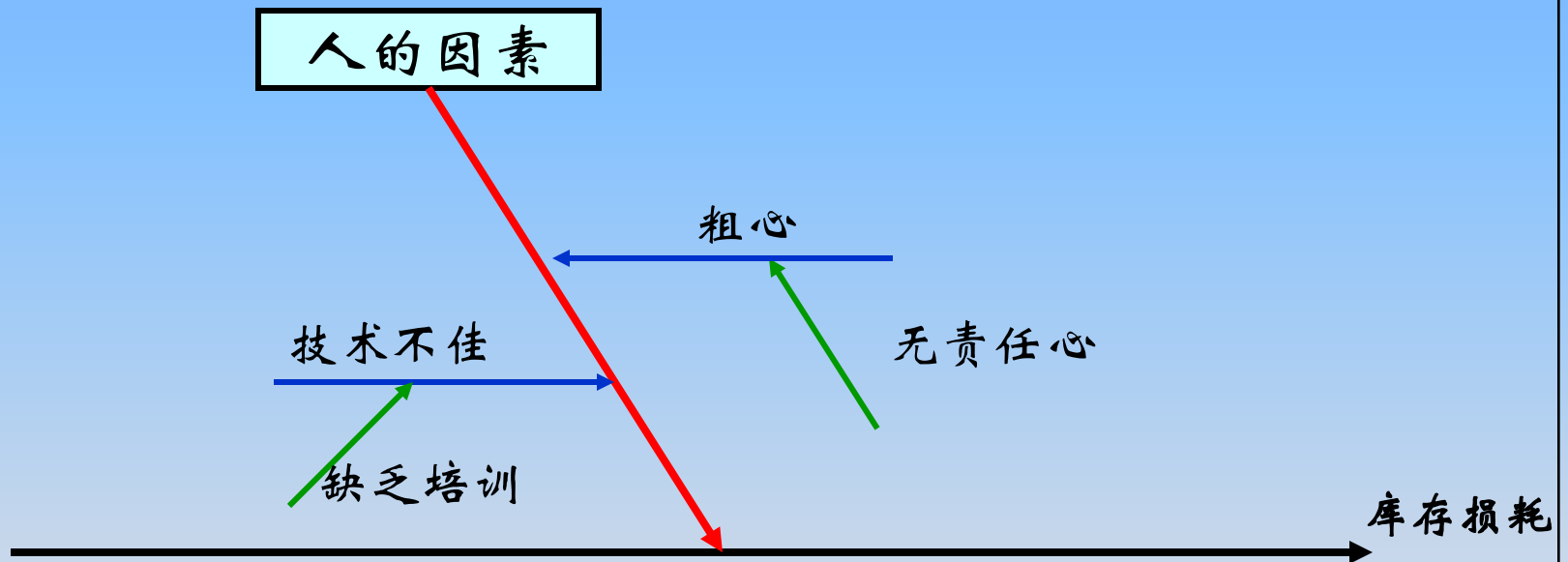
因果图



滞库图书因果分析



运用因果图有利于找到问题的症结所在，然后对症下药，解决问题。因果图再经营管理活动中，尤其是在物流作业中、问题分析中有着广泛的用途。



因果图

对策表

当利用鱼刺图确定了经营问题产生的主要原因后，有必要采取措施去消除这些原因，以达到改进的目的。这时，可以采用对策表的方法。

对策表的目的

用以针对经营问题产生的原因制定对策或措施，作为实施时的依据。

	对策表	
--	-----	--

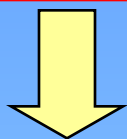
1	操作者不能掌握 物流作业流程	<input type="checkbox"/> 组织学习作业流程 <input type="checkbox"/> 强货责任心 <input type="checkbox"/>				
2	软件不稳定	<input type="checkbox"/> 做好日常维护 <input type="checkbox"/> 更换软件				
3	来货不准确	<input type="checkbox"/> 及时与出版商沟通确认 <input type="checkbox"/>				
4	发货不准确	<input type="checkbox"/> 更换物流人员 <input type="checkbox"/> 改拣货作业流程				

对策表的格式

频数分布表是一种把分散和不规则的数据，整理成一个能顺着其度量的尺度，清楚地显示出该数据的集中趋势和离散程度的一种统计方法。

	频数分布表	
--	-------	--

测定客户报订《教材完全解读》100次订单



教材完全解读订货频数分布表

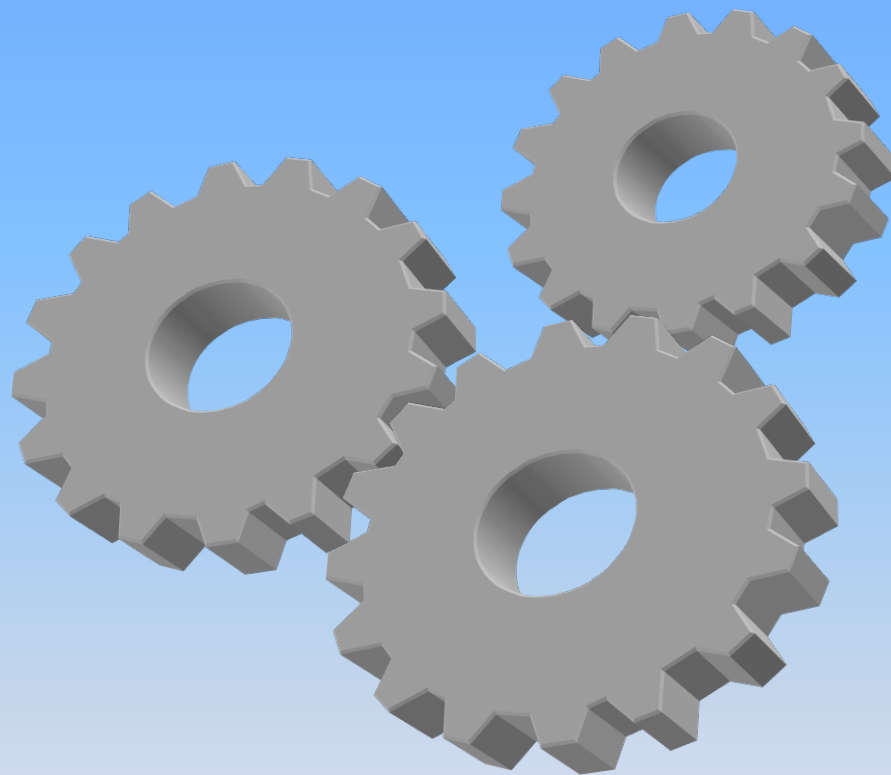
组号	下界限~上界限	组中值	频数符号	频数
1	11.405~11.505	11.455		1
2	11.505~11.605	11.555		2
3	11.605~11.705	11.655		7
4	11.705~11.805	11.755		13
5	11.805~11.905	11.855		24
6	11.905~12.005	11.955		25
7	12.005~12.105	12.055		16
8	12.105~12.205	12.155		10
9	12.205~12.305	12.255		1
10	12.305~12.405	12.355		1
Σ				100

频数分布表

1、从数据中找出最小值S和最大值L。

$$S = 11.45$$

$$L = 12.35$$



	频数分布表编制步骤1	
--	------------	--

2、决定组数。

$$m = 1 + 3.3 \lg n$$

当 $n = 100$ 时

$$m = 1 + 3.3 \lg 100 = 1 + 6.6 = 7.6 \approx 8$$

3、计算组距。

$$\text{组距 } h = \frac{\text{全距}}{\text{组数}} = \frac{L - S}{m}$$

组距 尽可能取为10、5、1、0.5、0.1、0.05

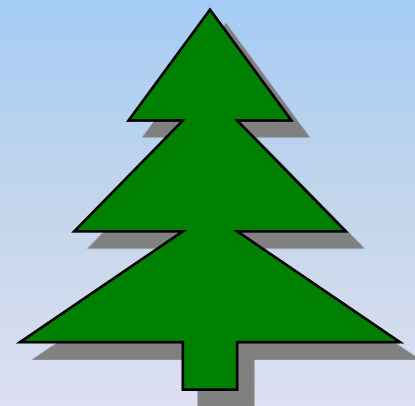
$$\text{组距 } h = \frac{12.35 - 11.45}{8} = \frac{0.9}{8} = 0.1125 \approx 0.1$$

4、求界限值。

在划分界限肘，必须明确端点的归属，所以在决定组的界限值肘，可以从每一个界限值上加上或减去1/2测量单位。

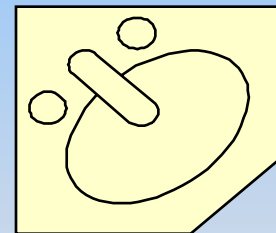
5、计算组中值。

各组的下界限值与上界限值的
平均值称为该组的组中值。



6、统计频数。

落在各组中的数据个数称为频数。

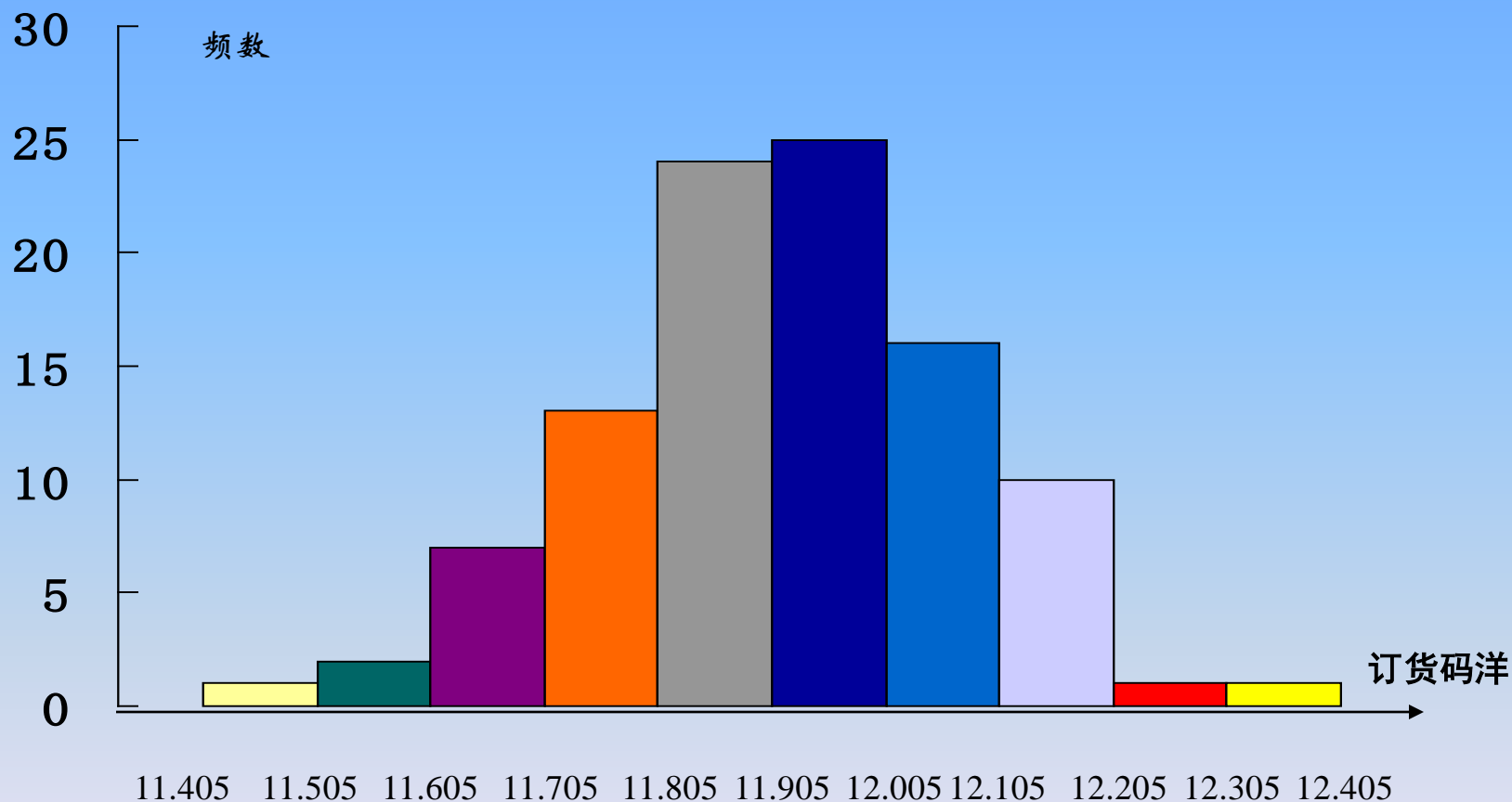


7、列频数分布表。

组号	下界限~上界限	组中值	频数符号	频数
1	11.405~11.505	11. 455		1
2	11.505~11.605	11. 555		2
3	11.605~11.705	11. 655		7
4	11.705~11.805	11. 755		13
5	11.805~11.905	11. 855		24
6	11.905~12.005	11. 955		25
7	12.005~12.105	12. 055		16
8	12.105~12.205	12. 155		10
9	12.205~12.305	12. 255		1
10	12.305~12.405	12. 355		1
Σ				100

频数分布表编制步骤7

以坐标横轴表示组距，坐标纵轴表示频数，所画出的矩形图称为频数直方图，简称直方图。

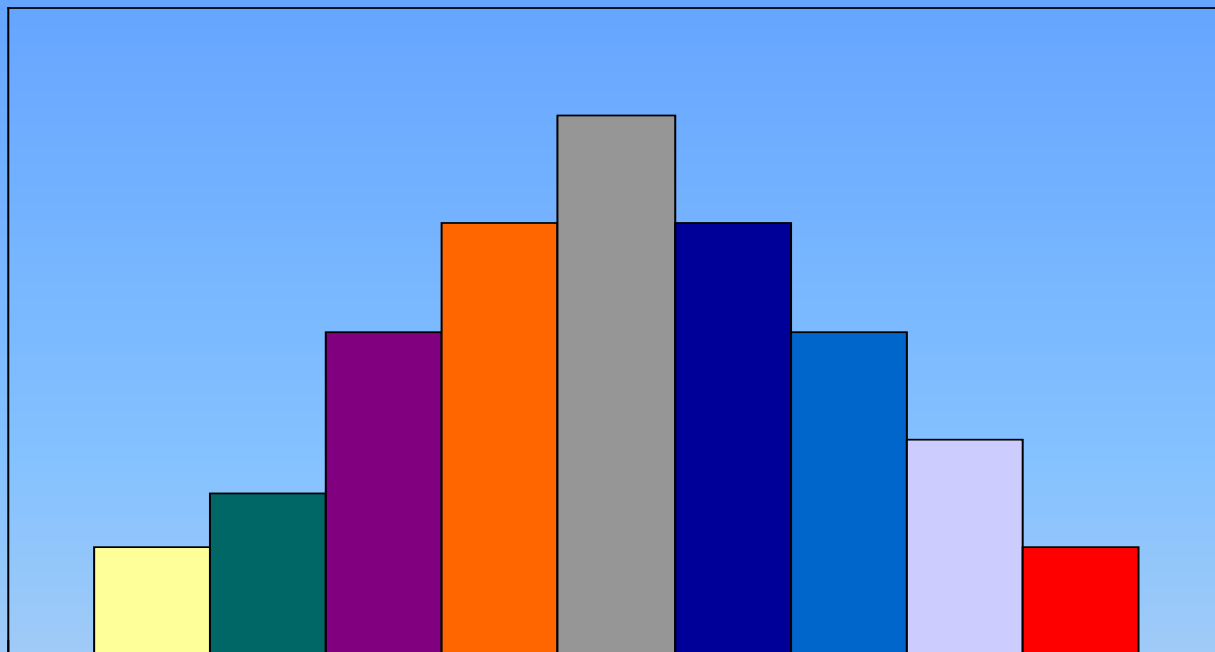


频数直方图

1、判断分布类型

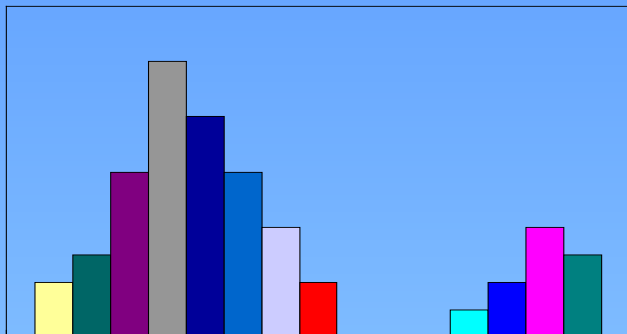
客户订货特性值的分布，一般都是服从正态分布或近似正态分布。当客户订货特性值的分布不具有正态性时，往往是公司营销政策不理想，或教辅供应能力不足。因而，由客户订货特性值所作的直方图的形状，可以推测营销政策是否稳定，或供货是否充足，由此可对客户订货作出初步判断。根据客户订货特性值的频数分布，可将直方图分为正常型直方图和异常型直方图两种类型。

	直方图在销售中应用	
--	-----------	--

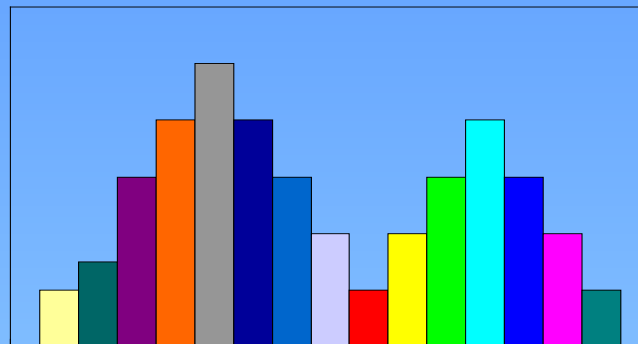


看直方图时应着眼于图形的整体形状，根据形状判断它是正常型还是异常型。正常型直方图具有“中间高，两边低，左右对称”的特征，它的形状像“山”，字。因此，根据客户订货特性值的频数分布所画出来的直方图是正常型时，就可初步判断为经营过程是稳定的，或供货能力是充足的。

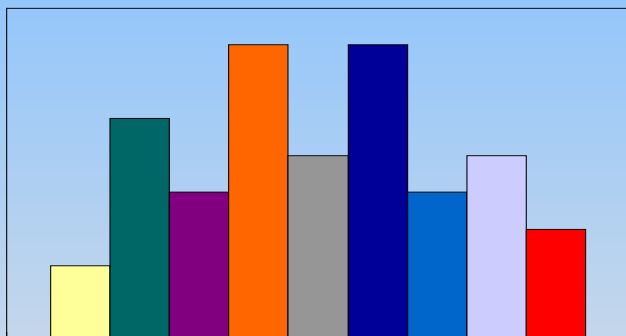
	正 常 型 直 方 图	
--	-------------	--



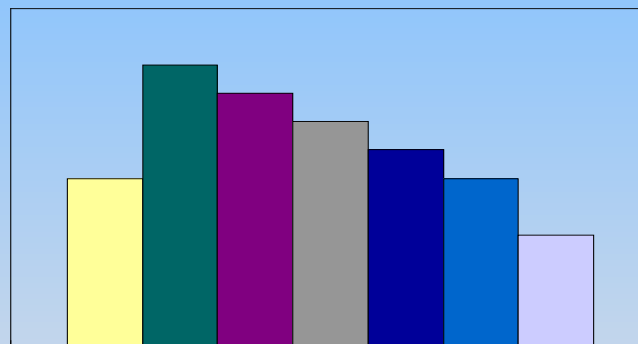
孤岛型直方图



双峰型直方图

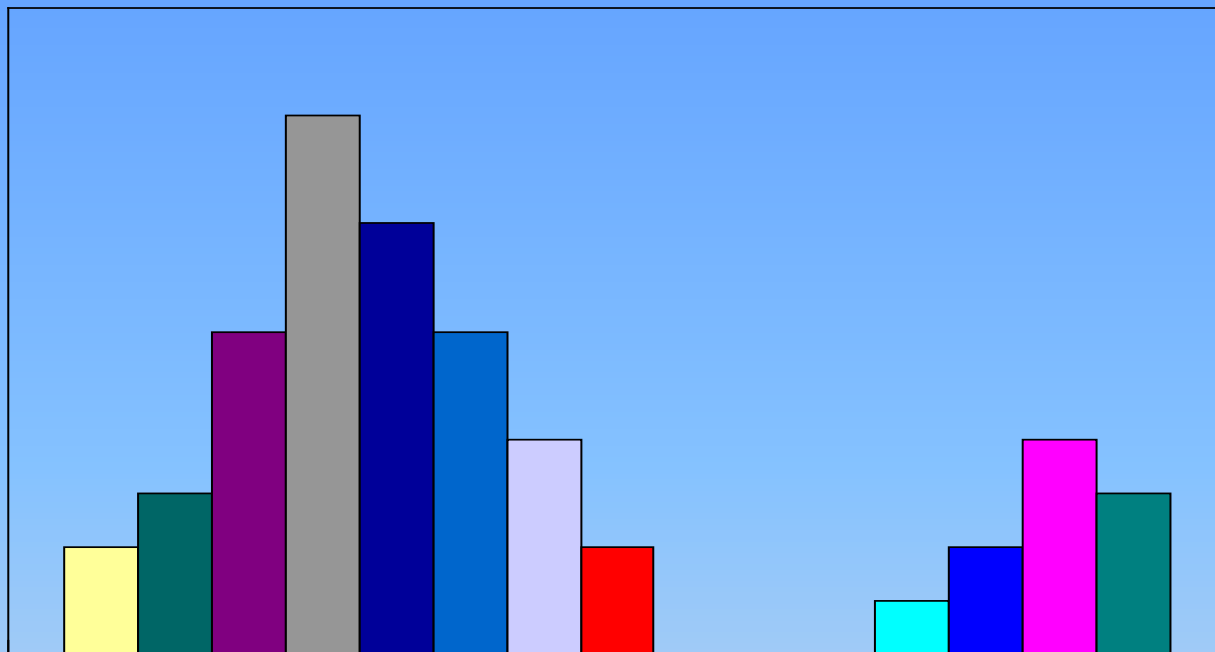


折齿型直方图



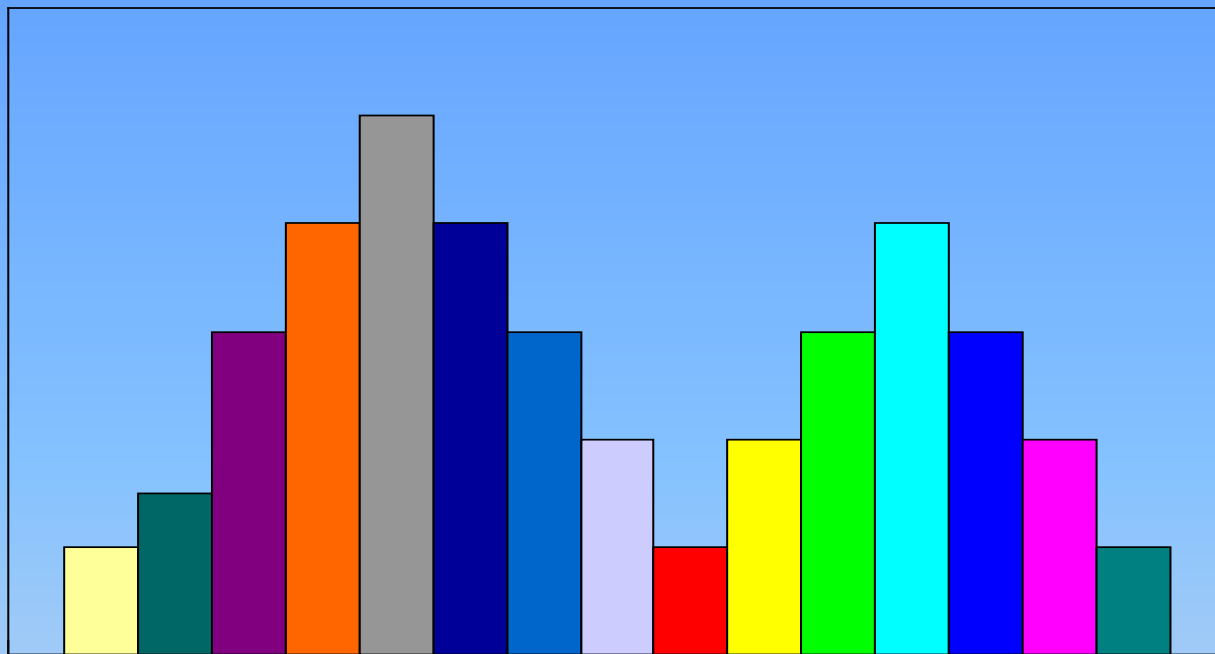
绝壁型直方图

不正常直方图



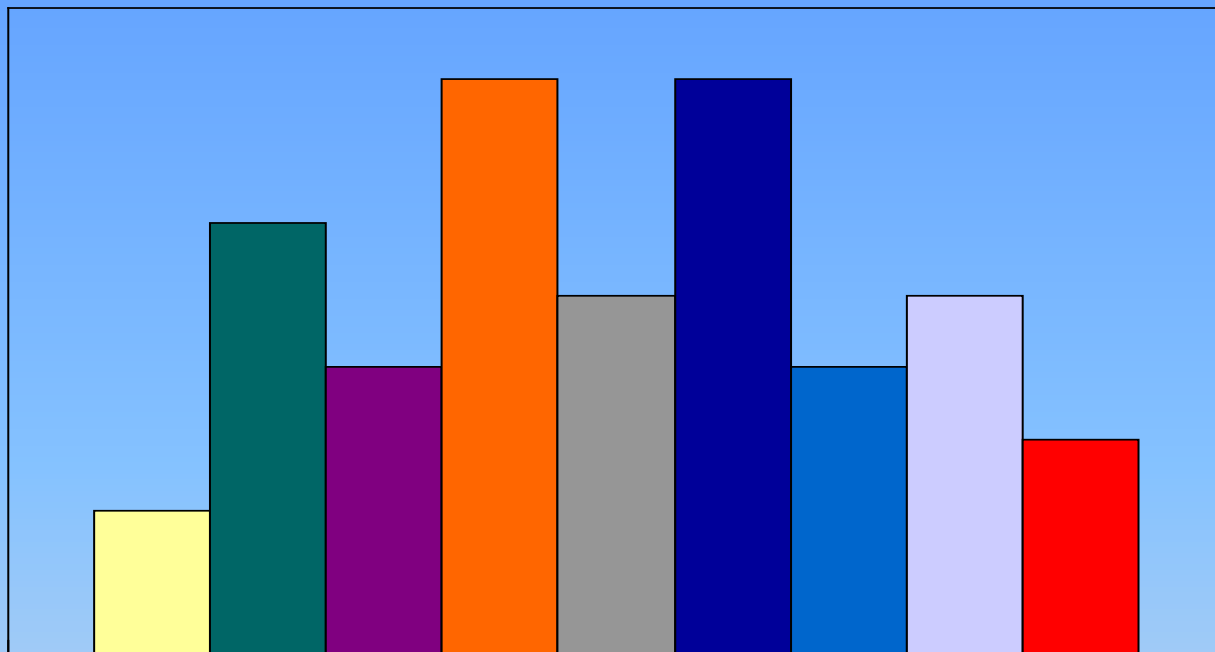
在主体直方图的左侧或右侧出现孤立的小块，像一个孤立的小岛。出现孤岛型直方图，说明有特殊事件发生。造成原因可能是一时供货发生变化，或者一段时间内出版商断故障，或者短时间内从第三方进货等。所以，只要找出原因，就能使直方图恢复到正常型。

孤 岛 型 直 方 图



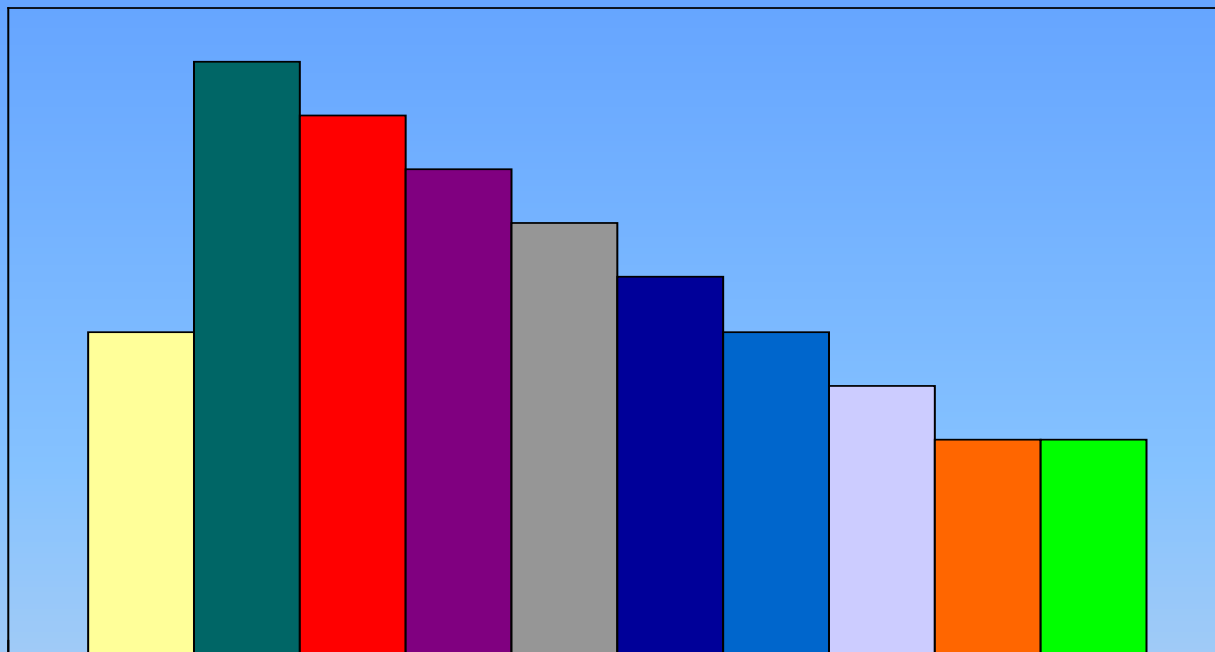
双峰型直方图是指在直方图中有左右两个峰，出现双峰型直方图，这是由于观测值来自两个总体、两种分布，数据混在一起。往往是由于将两季教辅销售列示在一起所造成的。

	双峰型直方图	
--	--------	--



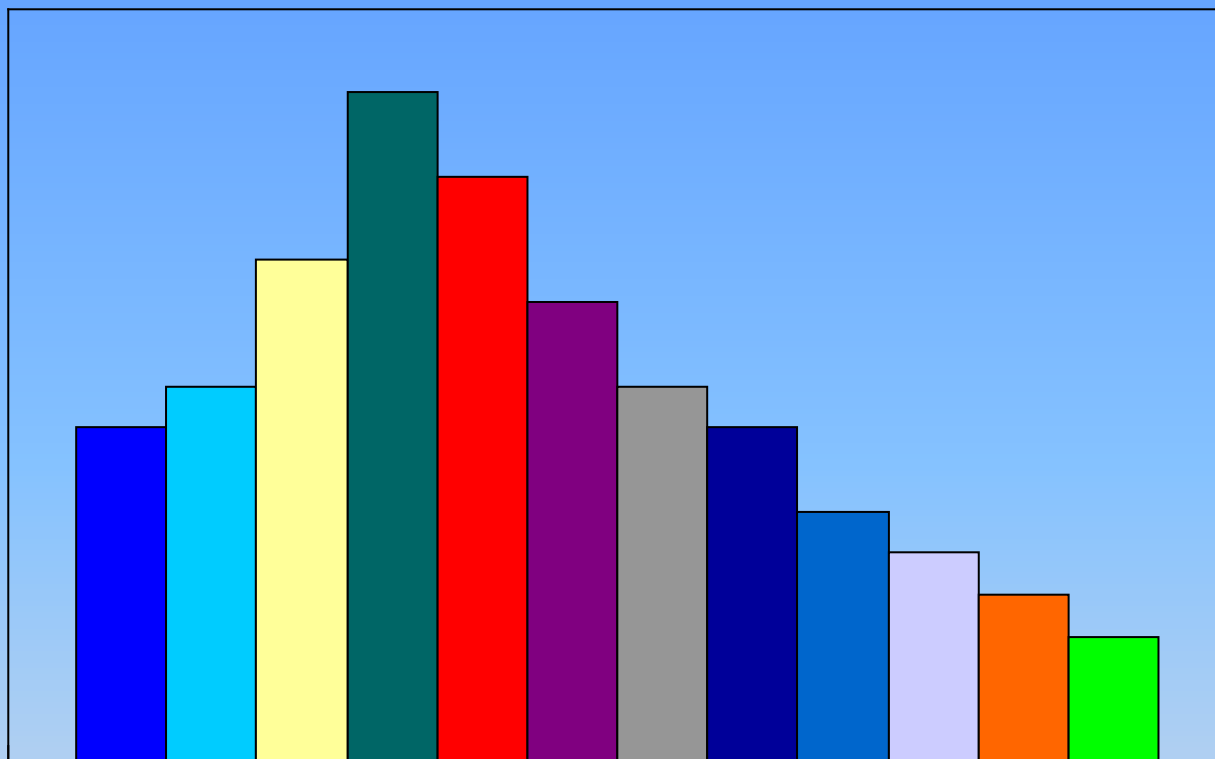
折齿型直方图形状凹凸相隔，象梳子折断齿一样。出现折齿型直方图，多数是由于备货不合理，或处理数据时分组不适当等原因造成。

折齿型直方图



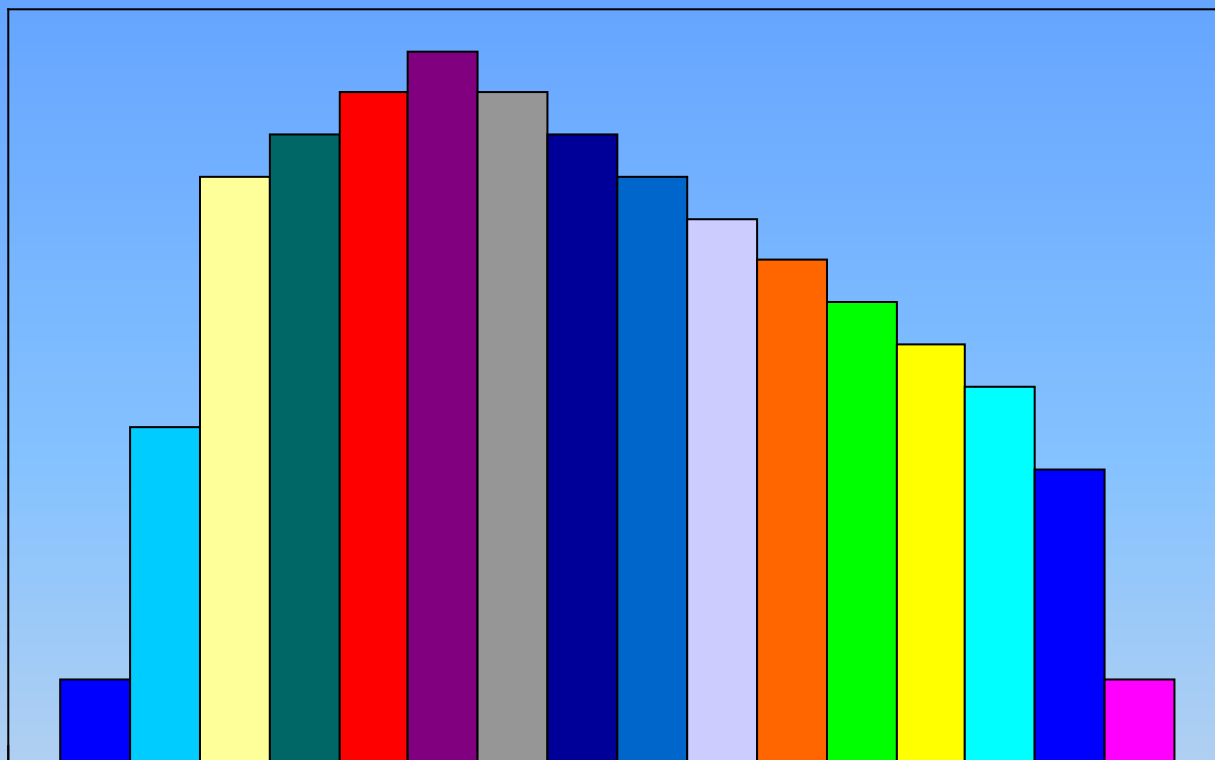
绝壁型直方图左右不对称，并且其中一侧像高山绝壁的形状，当用剔除了客户退货特性值数据作直方图时，往往会出现绝壁型直方图。此外，亦可能是操作者的工作习惯，习惯于偏标准下限，于是出现左边绝壁的直方图。

绝壁型直方图



某种原因使下（上）限受到限制时，容易发生“偏左型”（偏右型）。

偏态型直方图



与双峰型类似，由于多个总体、多种分布混在一起。

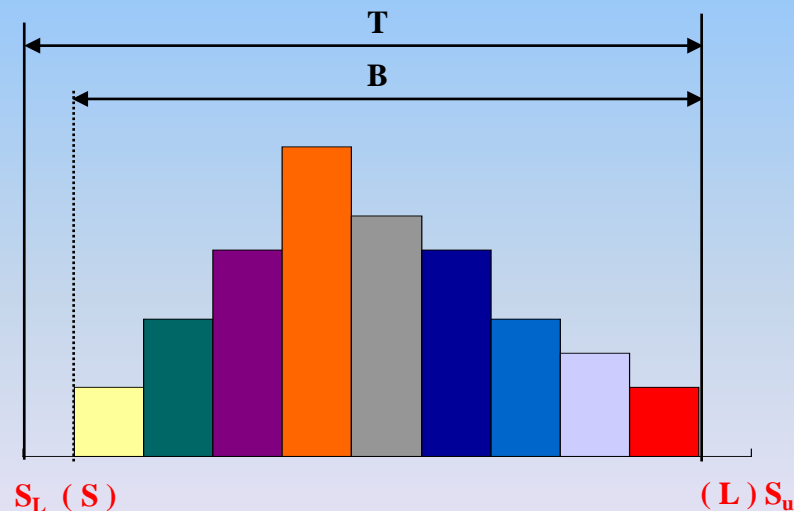
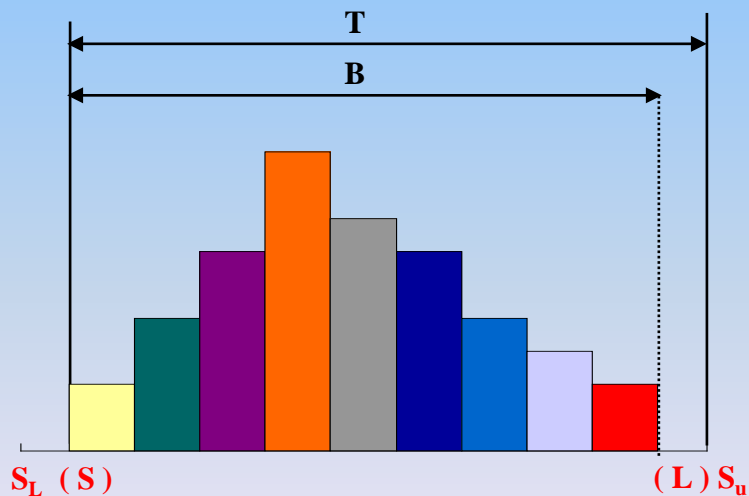
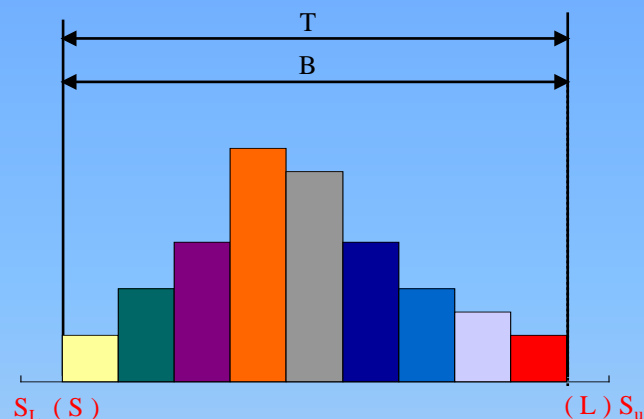
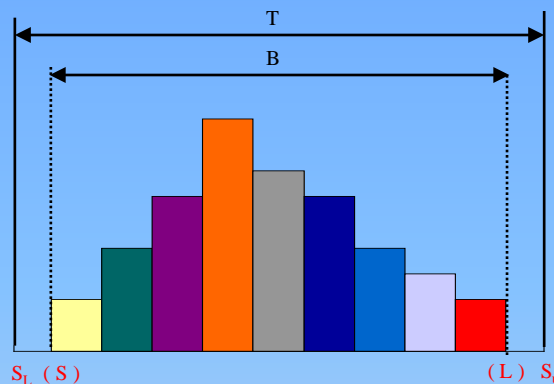
平顶型直方图

对于正常型直方图，将其分布范围 $B=[S, L]$ (S 为一批数据中的最小值， L 为一批数据中的最大值) 与标准范围 $T=[S_L, S_u]$ ， S_L 为标准下界限， S_u 为标准上界限) 进行比较，就可以看出订货特性值的分布是否在标准范围内，从而可以了解销售过程是否处于所希望的状态。为了方便，可在直方图上标出标准下界限值和标准上界限值。

直方图与标准比较

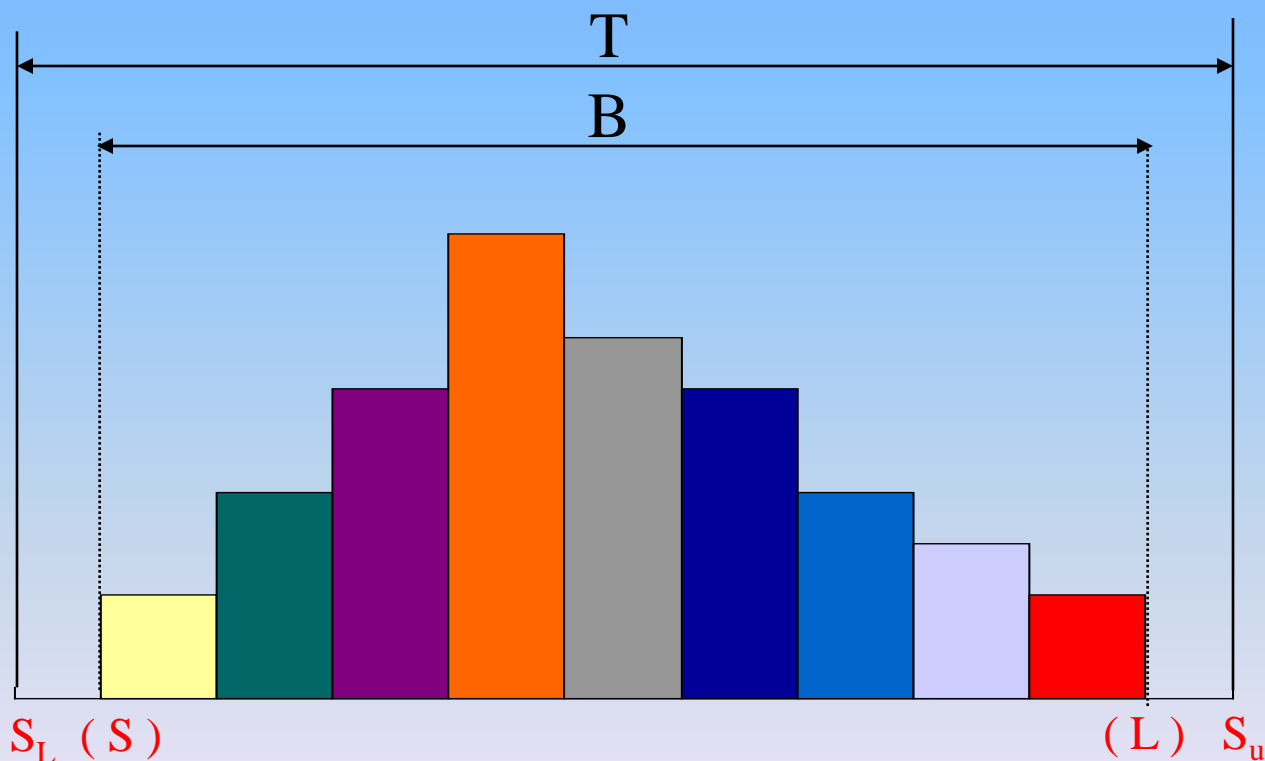
当教辅供应特性值符合规定标准时，其对应的直方图，必定在标准范围之内。

符合规定的直方图大致有下面四种类型：



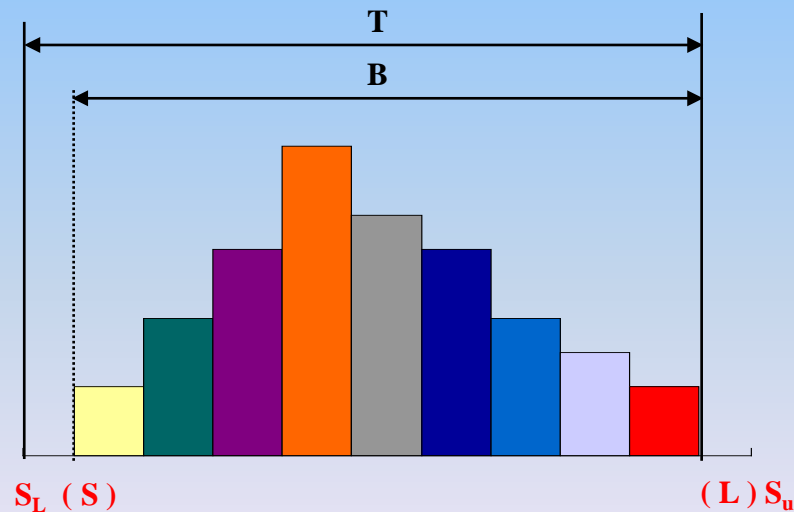
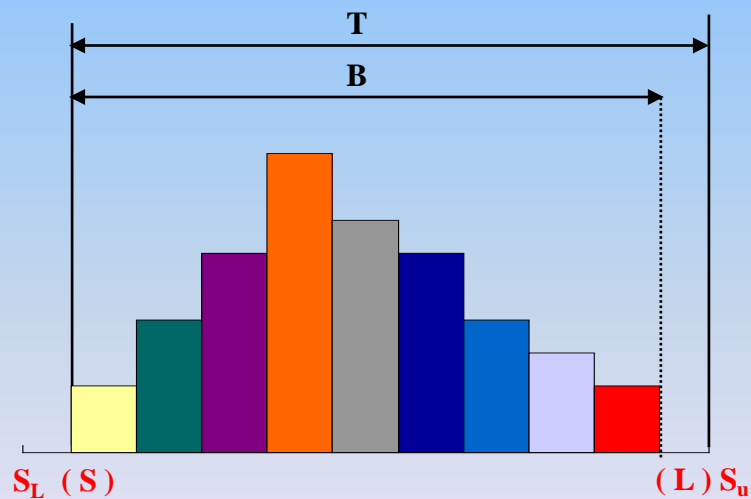
直方图在标准范围内的情况

直方图的分布范围B位于标准范围T内，但有余量;直方图的分布中心与标准中心近似重合，这是理想的直方图。此时，销售处于正常管理状态。



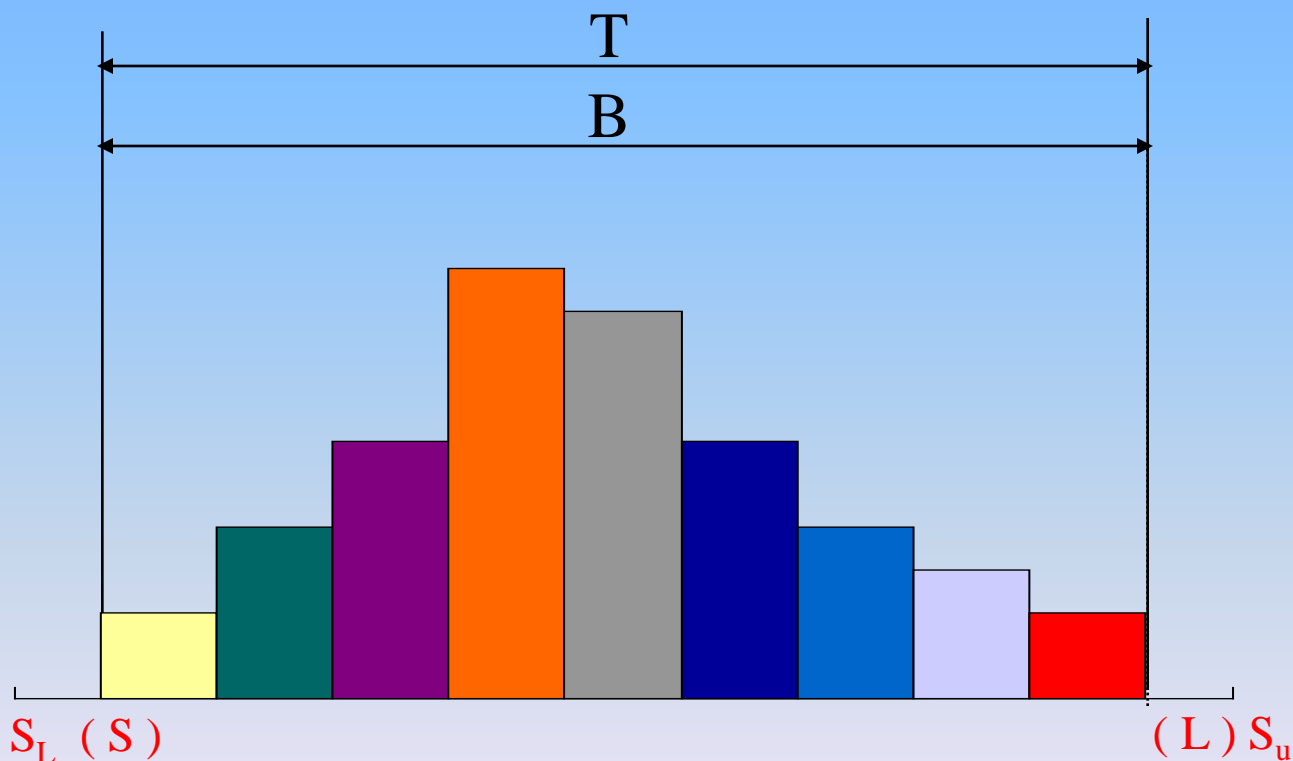
直方图在标准范围内的情况

直方图的分布范围B位于标准范围T内，数据变化仍比较集中，但分布中心偏移标准中心，并且直方图的一侧已达到标准界限，此时状态稍有变化，产品就可能超出标准，出现销量下滑。因此，需要采取措施，使得分布中心与标准中心重合。



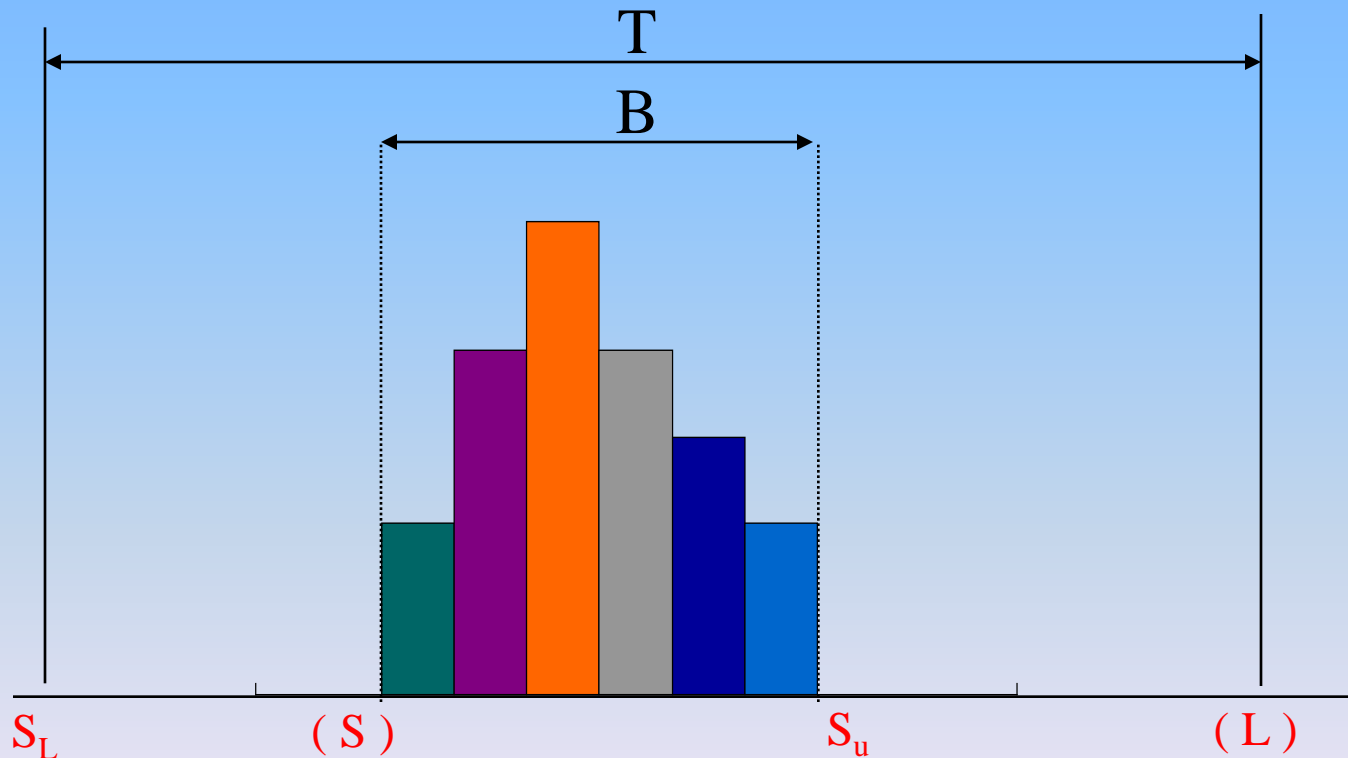
直方图在标准范围内的情况

直方图的分布范围B没有超出标准范围T，但没有余量。此时分布中心稍有偏移便会出现跑单，所以应及时采取措施，缩小订货特性值的分布范围。



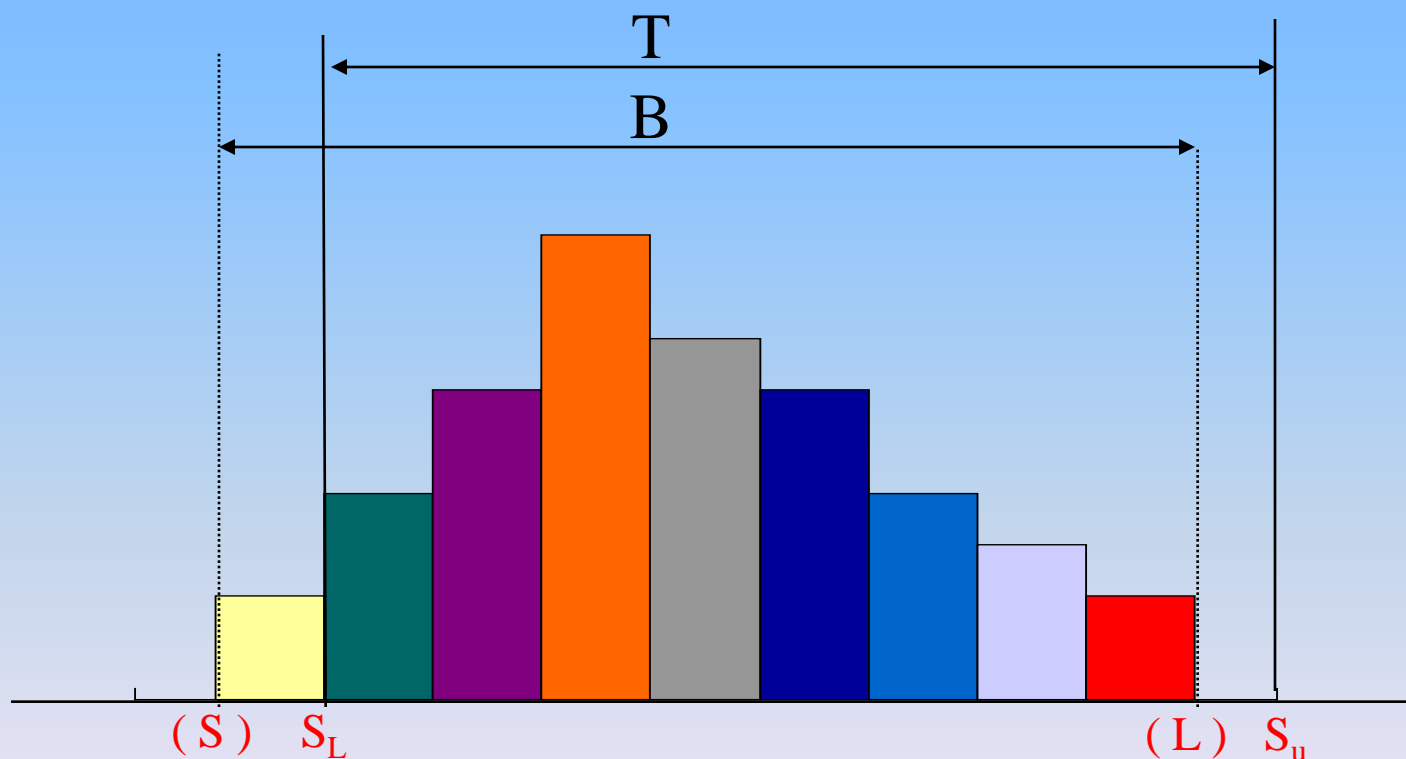
直方图在标准范围内的情况

客户订货特性值的分布非常集中，致使直方图的分布范围B与标准范围T之间的余量过大。此时，可对客户订货管理等适当放宽要求，从而降低渠道库存积压



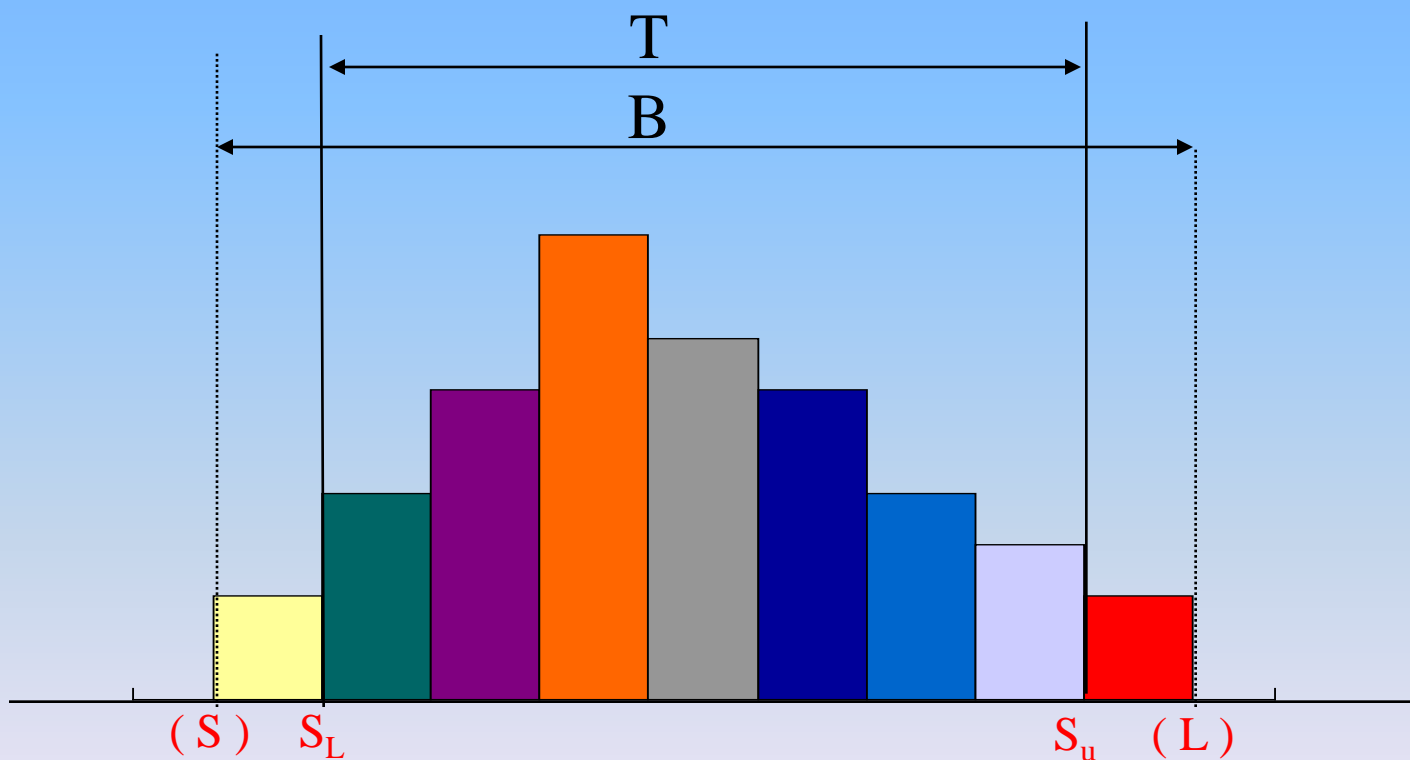
直方图在标准范围内的情况

客户订货特性值的分布中心向左(或向右) 偏离标准中心, 致使直方图分布范围B的下界限(上界限)超出标准范围T的下界限(或上界限), 因而在下界限(或上界限)出现客户后期从第三方进货或者由于订数据不退货害怕库存慎重销售。



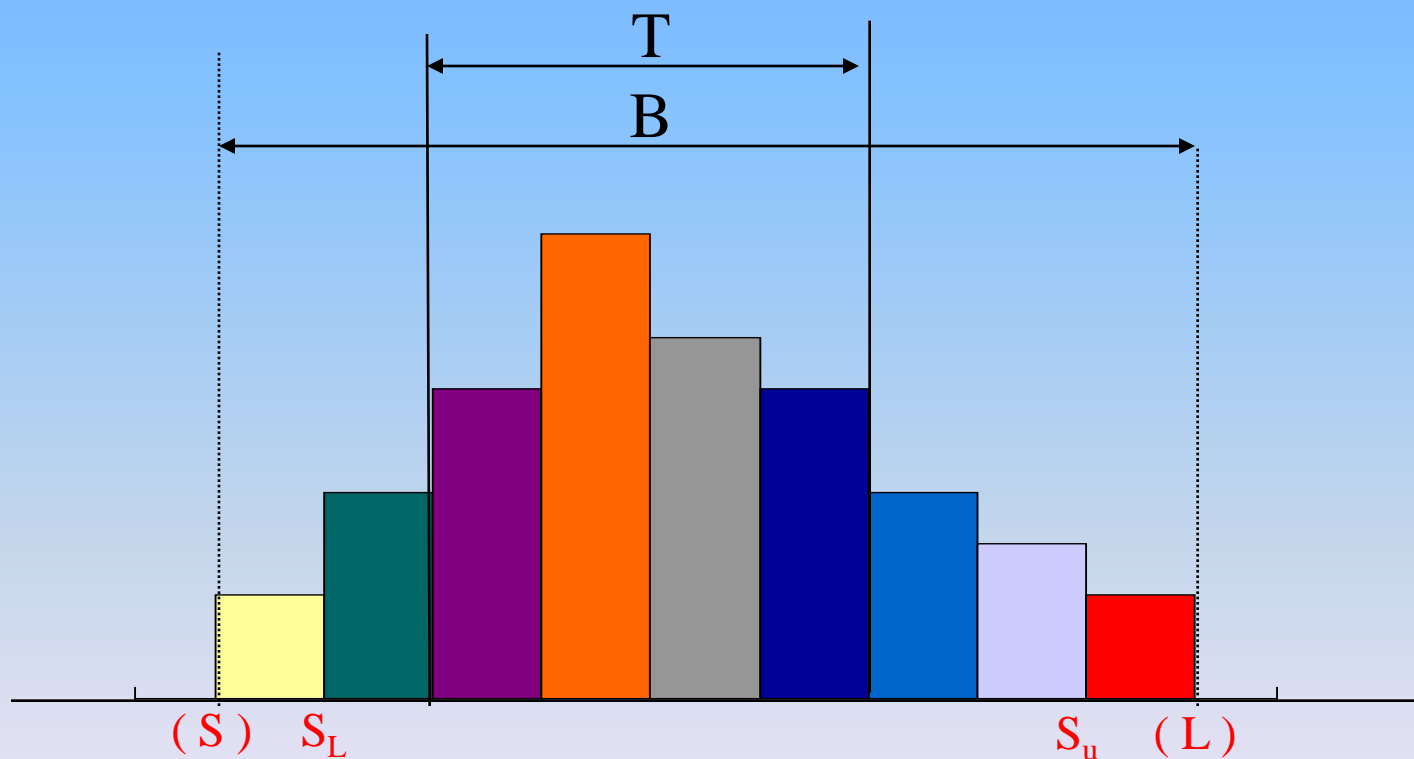
直方图超出标准范围内的情况

直方图的分布范围B超出标准范围T，此时，在标准上界限和下界限都出现不合格品。这种情况通常是由于产品销售疲软，这时，应及时采取促销措施，降低分布的标准差。如果属于标准定得不合理，可以放宽标准范围。

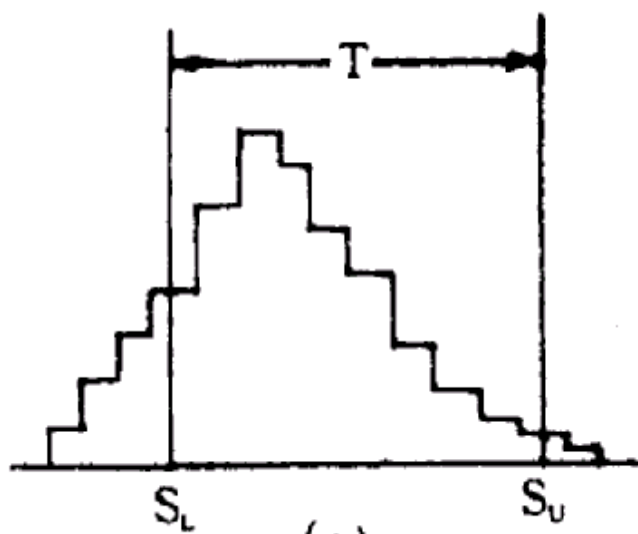


直方图超出标准范围内的情况

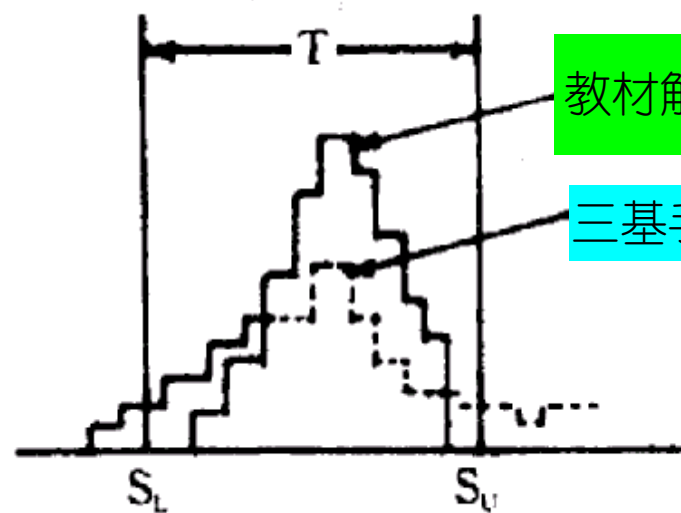
直方图的分布范围B大大超出标准范围T，此时已出现市场大面积少货，必须立即分析原因，采取紧急措施；如果营销政策允许改变，就重新修订营销政策。



直方图超出标准范围内的情况



(a)



(b)

教材解读

三基手册

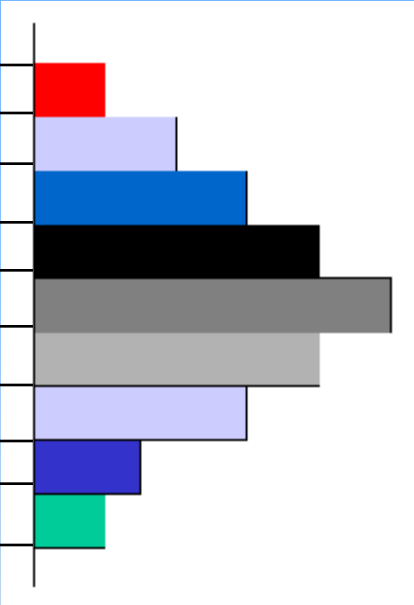
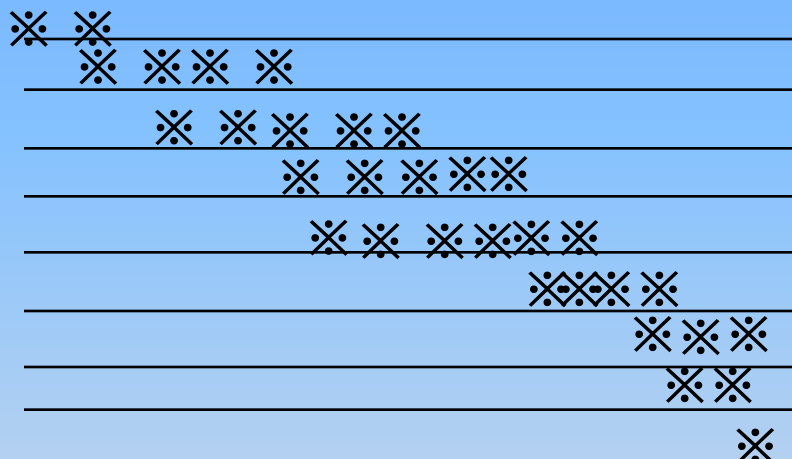
直方图的分层比较

波动图

直方图

0.09
0.08
0.07
0.06

0.01



时间

直方图的缺点

下线客户不同品种订货统计表

i \ j	1	2	3	4	5	6	7	8	9	10
1	15.0	15.8	15.2	15.1	15.9	14.7	14.8	15.5	15.6	15.3
2	15.1	15.3	15.0	15.6	15.7	14.8	14.5	14.2	14.9	14.9
3	15.2	15.0	15.3	15.6	15.1	14.9	14.2	14.6	15.8	15.2
4	15.9	15.2	15.0	14.9	14.8	14.5	15.1	15.5	15.6	15.1
5	15.1	15.0	15.3	14.7	14.5	15.5	15.0	14.7	14.2	14.2

直 方 图(练习)

1、从数据中找出最小值S和最大值L。

$$S = 14.2$$

$$L = 15.9$$

2、决定组数。

$$m = 1 + 3.3 \lg n = 6$$

3、计算组距。

$$\text{组距 } h = 0.3$$

4、求界限值。

$$\text{下限值 } S - h/2 = 14.15$$

5、计算组中值。

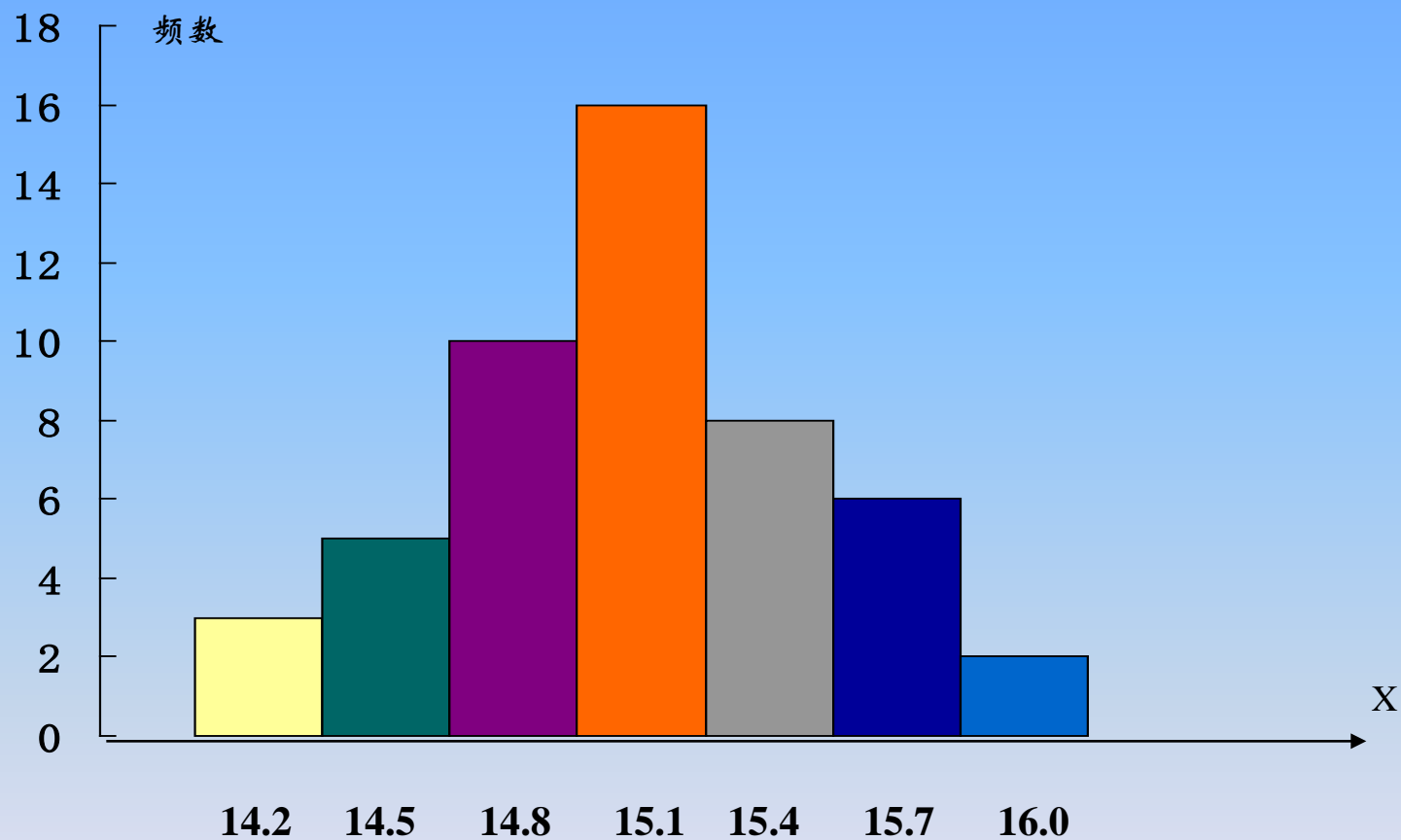
6、统计频数。

7、列频数分布表。

直方图(练习)

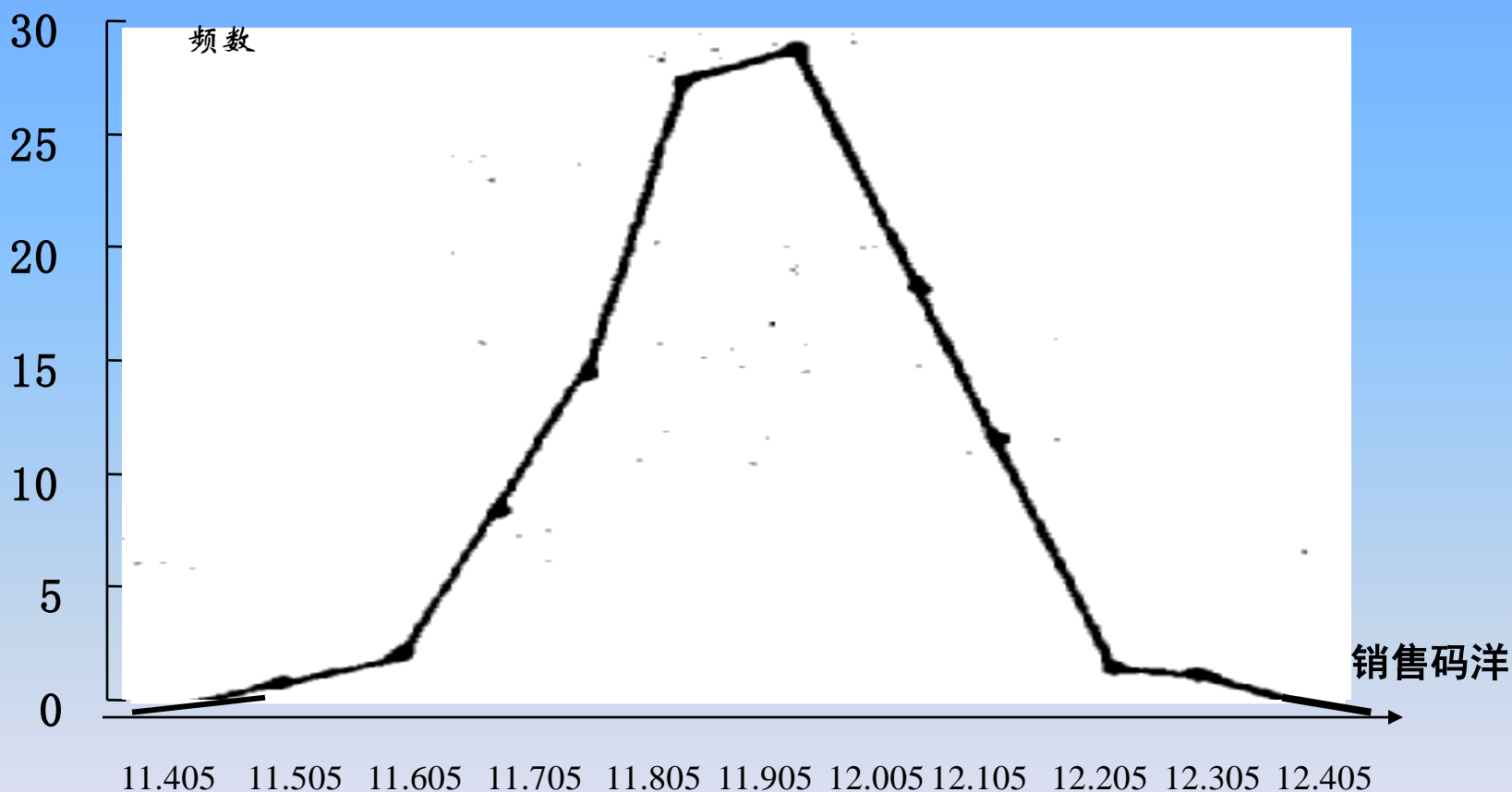
组号	下界限~上界限	组中值	频数fi	频率Pi
1	14.05~14.35	14. 2	3	0. 06
2	14.35~14.65	14. 5	5	0. 10
3	14.65~14.95	14. 8	10	0. 20
4	14.95~15.25	15. 1	16	0. 32
5	15.25~15.55	15. 4	8	0. 16
6	15.55~15.85	15. 7	6	0. 12
7	15.85~16.15	16. 0	2	0. 04
Σ			50	100%

直方图(练习)



直方图(练习)

以坐标横轴表示组中值, 坐标纵轴表示频数, 所画出的多边形图称为频数多边形, 简称多边形图。多边形图的作法与直方图类似, 不同的只是多边形图以组中值为横坐标, 频数为纵坐标, 在坐标平面上依次标出各点的位置, 然后把相邻各点用直线段连接起来, 由此得到频数多边形。



频数多边形

请大家参与讨论