

# 测量（上）

----

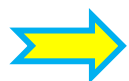
## 过程分析与文档

2022

六西格玛黑带课程培训



# 课程内容



1

过程分析概述

P133

2

流程图

3

因果图

4

因果矩阵图

5

其他过程分析工具和文档

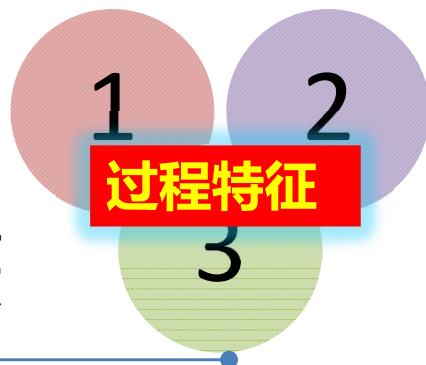
# 过程分析概述

## 过程分析

是六西格玛管理活动中的重要内容，也是测量阶段的重要工作。  
测量阶段从过程分析开始。

将输入转化为输出

是一个或一组活动，接受输入并增加价值，然后向内部或外部顾客提供输出

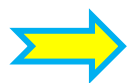


是一组相互关联的活动

它将一个概念、一项要求或一个订单转化为一种可交付的产品或服务

# 课程内容

1 过程分析概述



2 流程图

P134

3 因果图

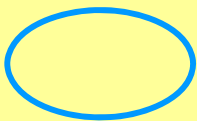

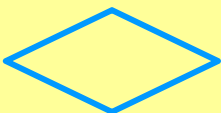


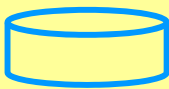
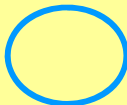
4 因果矩阵图

5 其他过程分析工具和文档

# 流程图

## 流程图概念

**流程图**是展现过程步骤和决策点顺序的图形文档，是用来了解、分析和归档公司的流程和活动的一种工具，帮助企业确定流程设计。流程图可以显示出将特定输入转化为所需输出结果的一系列步骤。

开始与结束	活动	判定	流线	文件	数据库	连接
						

# 流程图

## 流程图的作用

“ 一张图胜过千言万语 ”

质量改进首先应当着眼于**分析和评价**过程结构的合理性，通过全面细致的分析进而逐步确定流程中的**评价或失控**的薄弱环节。

认识过程

流程图的结构有助于团队全体成员对**过程全面的了解**。

改进过程

分析过程

通过对细化流程的分析，可能揭示出**改造这一过程的机会**。

# 流程图



## 流程图的作用

## 如何分析？

### 调查每个菱形符号

菱形判定符号用于表示一次检查活动。检查包括**检验、试验、鉴定、评审**等。检查是过程控制的重要活动，**遗漏必要的检查或者检查的内容不充分**，往往是导致过程失控、质量问题高发的原因。

### 调查每个循环

循环代表了**由于不合格而引起的返工**，是资源和时间的浪费。团队要探讨循环内的活动，首先需要关心循环的“长度”---返工活动的数量、时间、成本等。质量改进团队应该尽最大努力以**寻求缩短循环回路，降低成本**。

### 调查每个活动符号

现在团队可以把注意力转向过程主流程的**每项活动**中去。譬如，它是增值的还是辅助的活动？它是必须的吗？能省略吗？该项活动的成本和效率如何？顾客对它是否关注？这些活动本身是否具备防差错能力？

### 调查每个文件

正确的**文件和数据**是过程的重要信息来源。错误的信息则会对过程造成严重的损害。团队应对文件和数据结构进行调查，考虑输入和输出的信息是否充分？同时还应当关注文件与数据的更新频次，保证信息的真实性。

# 流程图

## 概要流程图

团队准备开展质量改进工作时，采用一个概要流程图是很有用的。概要流程图一般包括4-12个步骤。简明概括地描述了过程的主要结构，有助于团队对过程的总体达成共识、界定项目范围和选择团队成员。



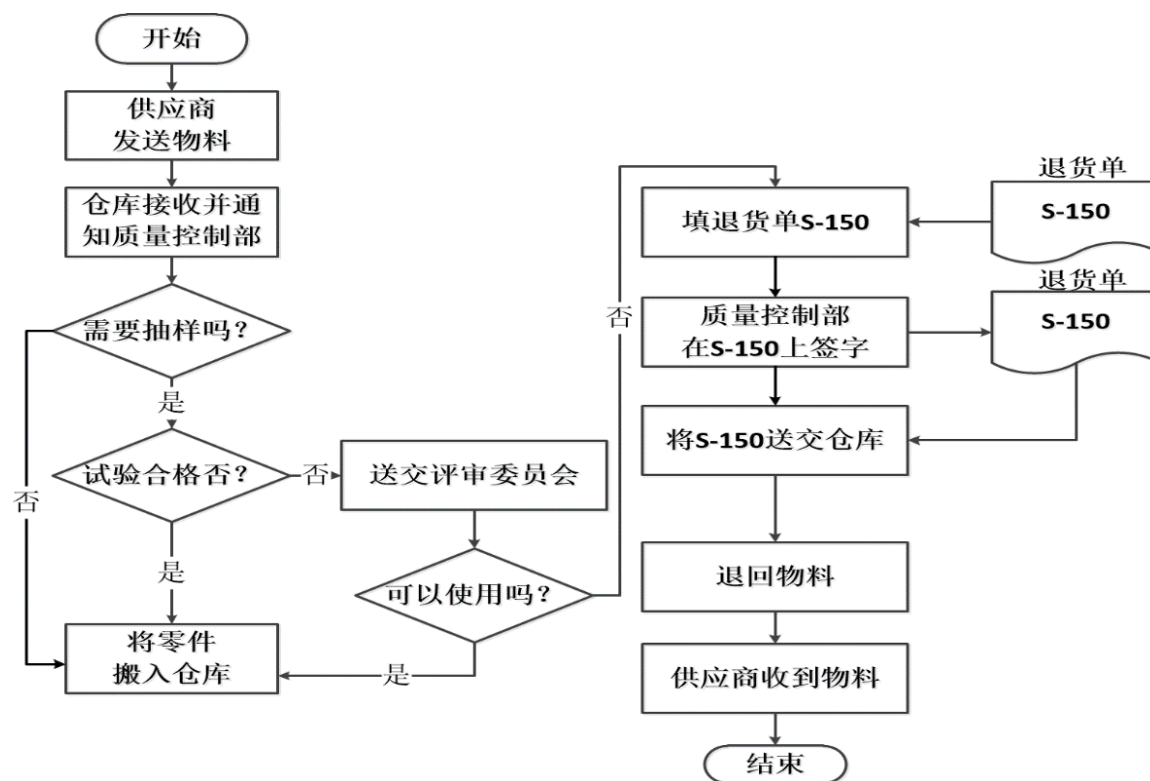
概要流程图—采购物料进货过程



# 流程图

## 细化流程图

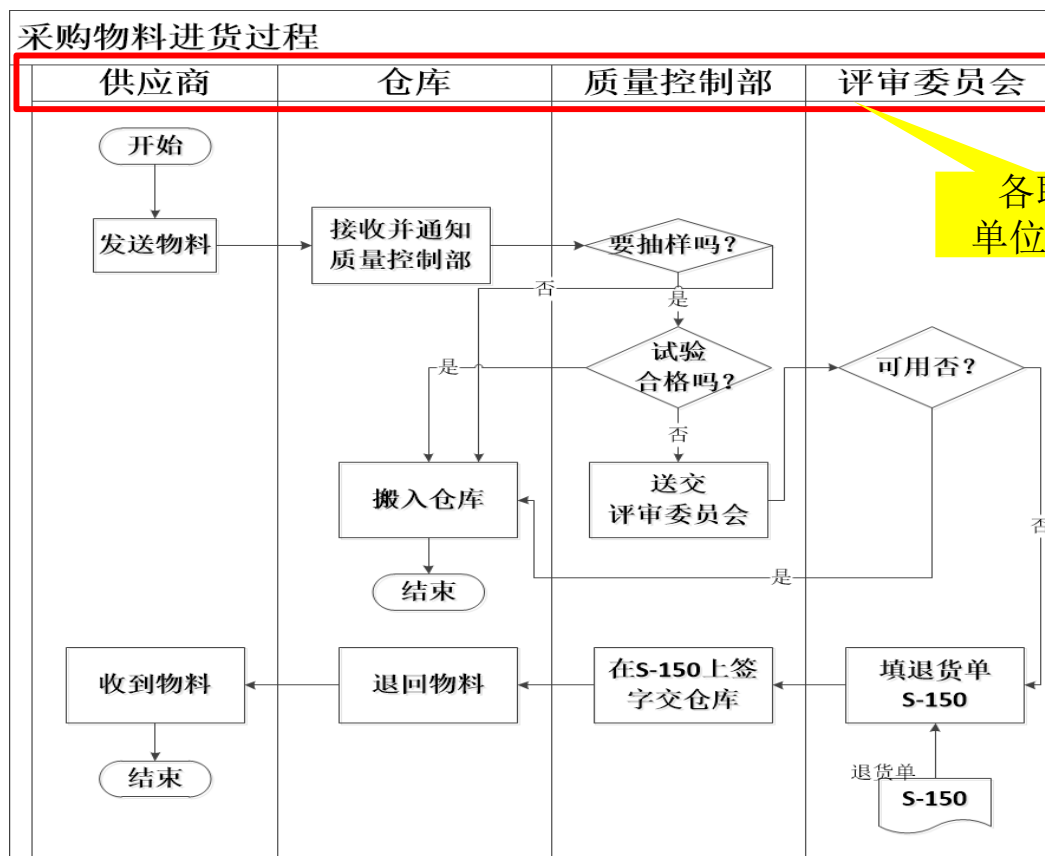
在上述概要认识的基础上，团队需要详细了解过程中每一步骤的具体活动及实施细节。通过细化流程图，可以将概要流程图进行详细的分解和展开。



# 流程图

## 矩阵流程图

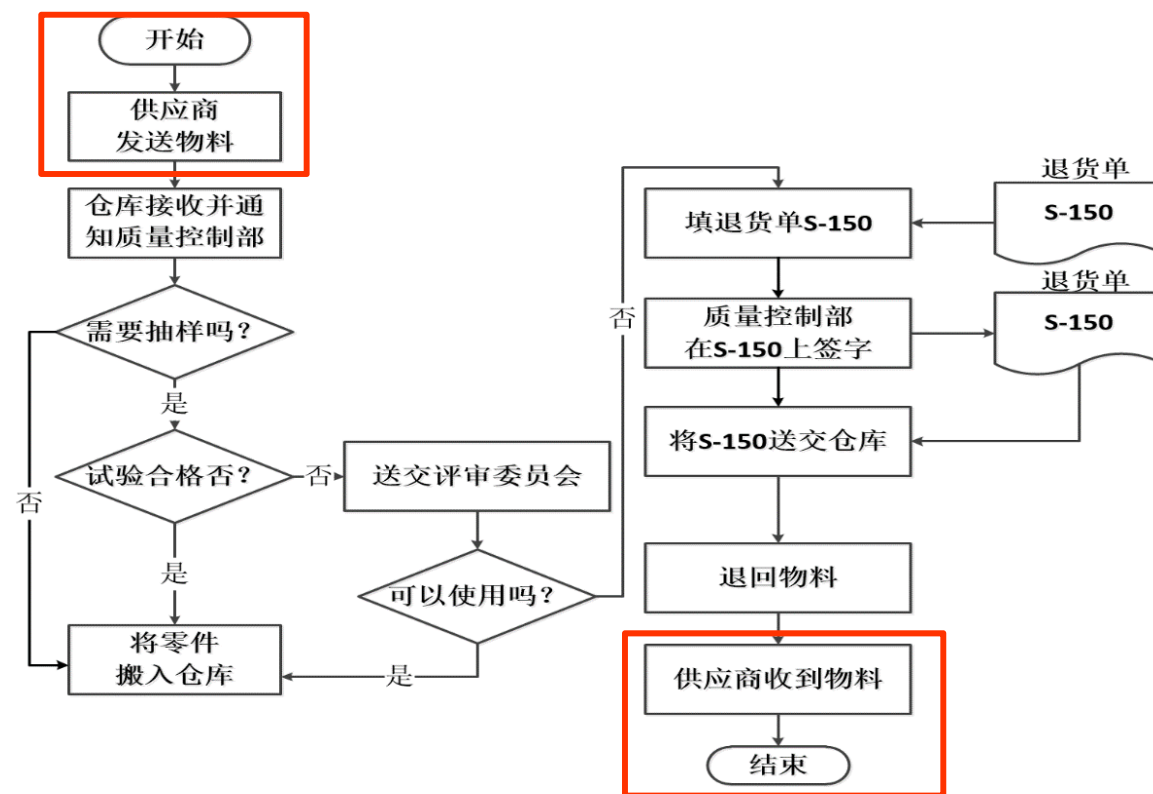
如果过程涉及的活动和参与的单位较多，常规的细化流程图会比较复杂，尤其是各相关方在过程中的职能作用及其相互衔接关系（接口）往往难以识别。右图采用的矩阵流程图是在细化流程图基础上的一种改进形式，它能够更清晰地表达相关方在过程中的关系



# 流程图

## 绘制流程图

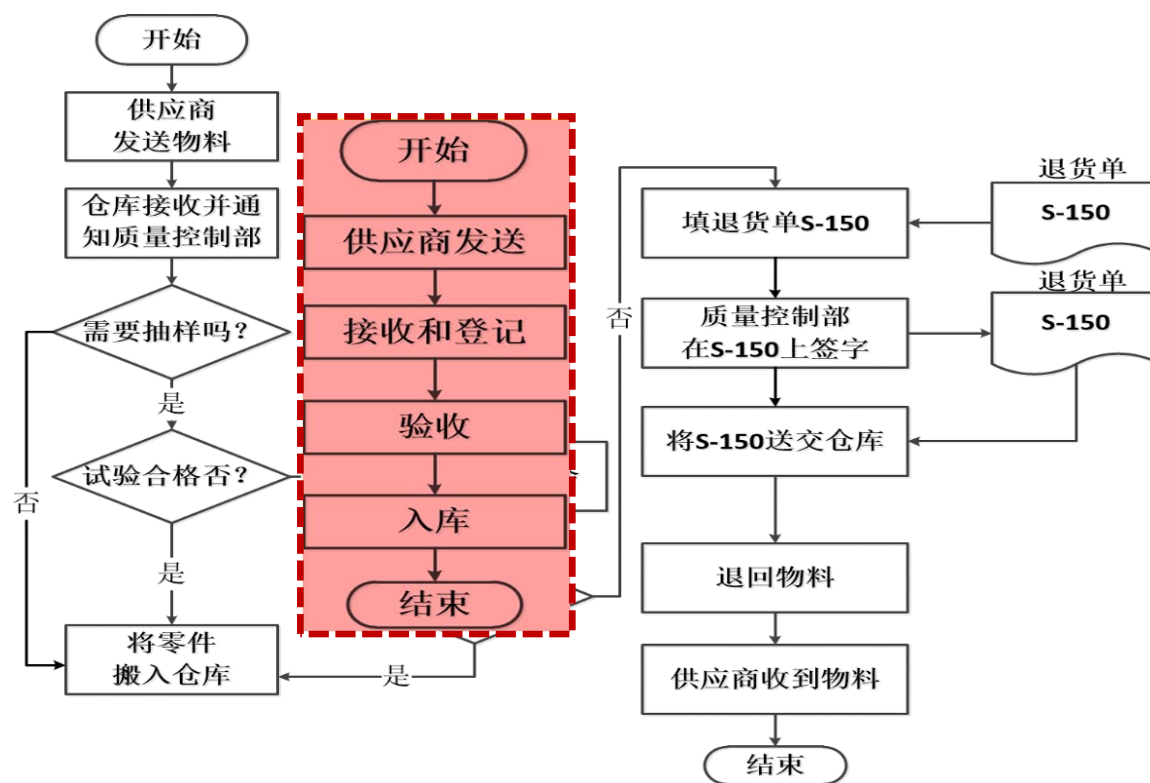
第一步: 判定过程的开始点和结束点;



# 流程图

## 绘制流程图

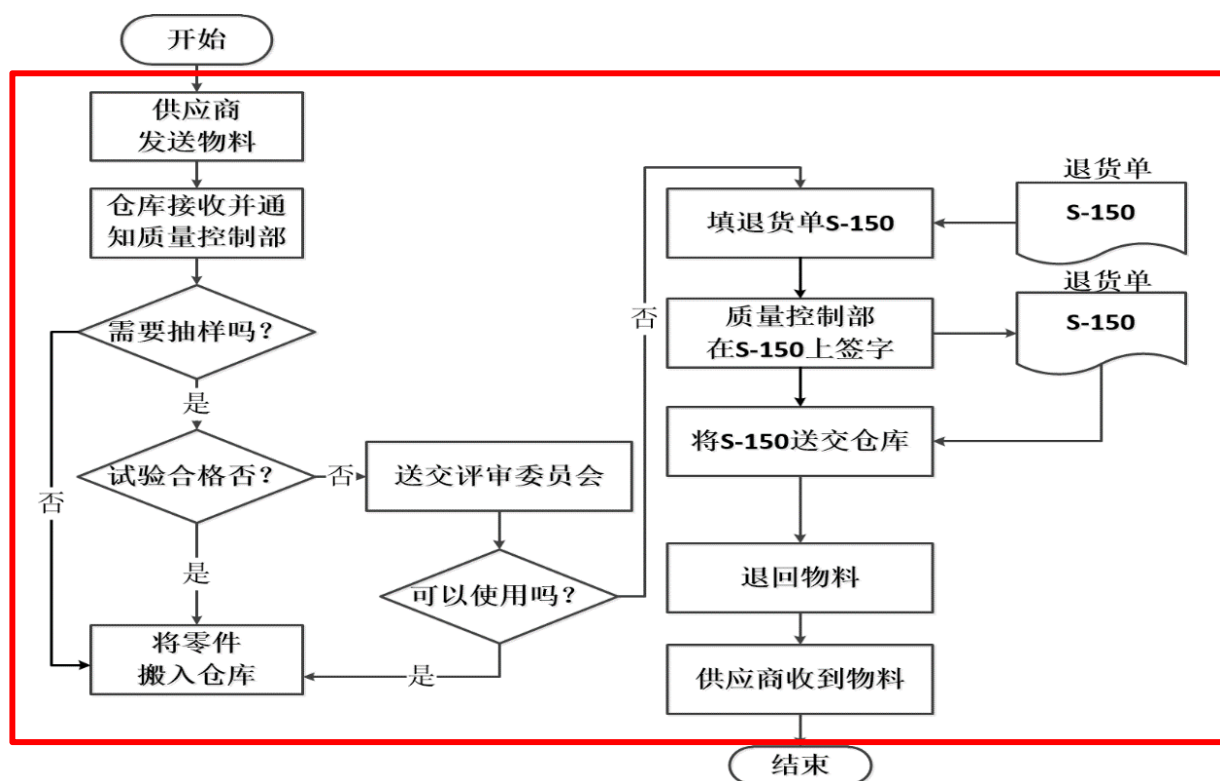
第二步: 观察从开始到结束的整个过程;



# 流程图

## 绘制流程图

第三步: 识别过程中的步骤（包括主要活动、判断和决策点等），以及各个步骤或活动的流向和相互关系；



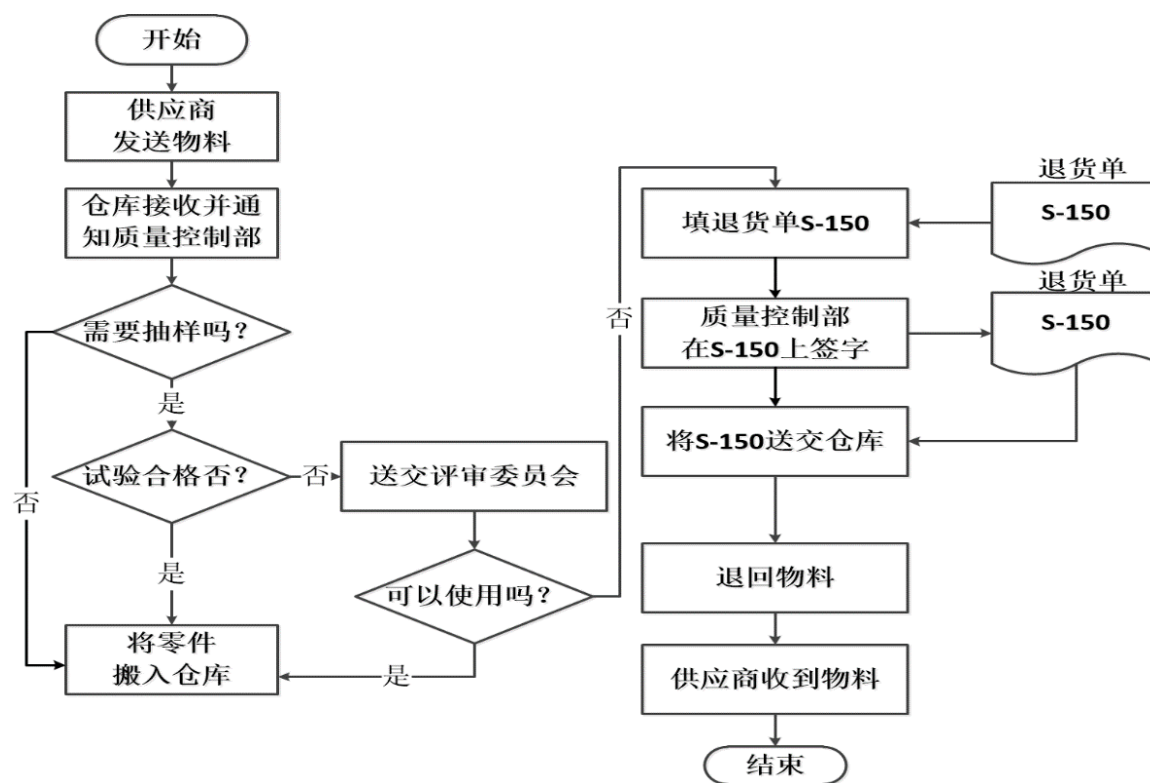
# 流程图

## 绘制流程图

第四步: 将上述步骤绘制出来, 形成流程图的草图;

第五步: 团队成员就这张草图进行充分沟通并达成一致;

第六步: 形成正式文档。



# 流程图

## 好的流程图所具备的特点

- 流程名称
- 清楚的开始和结束点
- 由上而下，从左到右都有清楚的顺序流向，禁止死循环
- 符合作业、突出重点

在绘制流程图的过程中，**应避免闭门造车**、想当然绘制过程的流程图。对过程分析来说，最重要的是根据过程的实际情况绘制流程图。

# 流程图

**绘制完流程图之后，应对流程图进行充分的分析，应关注以下几个方面：**

- 产生过程输出**缺陷**或问题的**重点关注区域**在那些环节或步骤上；
- 流程中的**非增值**步骤或环节在何处，如返工/返修环节或步骤等；
- 流程中是否有**“瓶颈”**存在，这里的“瓶颈”是指某点的工作负荷大大超过此处的处理能力，从而延缓了整个工作进度。“瓶颈”现象也指准时把产品或服务按足够的数量传递给顾客的薄弱环节。
- 流程中是否有**缺失、冗余或者错误**的步骤等。

流程图在**DAMIC**项目工作的各个阶段都将用到，它是**过程分析**的**重要工具**，也是项目工作中产生的**重要文档**。



# 习题

题目：在进行管理流程改进时，经常需要绘制跨职能流程图，有关跨职能流程图的说法，下面错误的是：

C

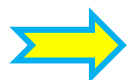
- A. 跨职能流程图的优点是能够清晰地分析流程在各职能部门的流动情况
- B. 绘制跨职能流程图时，职能部门可以水平摆放，也可以垂直轴摆放
- C. 跨职能流程图不能分析流程中的活动是否是增值活动**
- D. 跨职能流程图上可以标注活动所需时间



# 课程内容

1 过程分析概述

2 流程图



3 因果图

P136

4 因果矩阵图

5 其他过程分析工具和文档

# 鱼骨图（因果图）

P136

## 鱼骨图的概念

1953年**日本**管理大师**石川馨**发明了一种极方便又有效的原因分析法，叫“鱼骨图”又叫“石川图”。鱼骨图是一种发现问题“根本原因”的方法，它也可称为“因果图”。鱼骨图广泛应用于质量管理。

鱼骨图分析法倡导头脑风暴法，它是一种通过集思广益、发挥团体智慧，从各种不同角度找出问题所有原因或构成要素的会议方法。

# 鱼骨图（因果图）

- ◆ 鱼骨图是一个**非定量的工具**，可以帮助我们找出引起问题潜在的**根本原因**。
- ◆ 它使我们问自己：问题为什么会发生？**使项目小组聚焦于问题的原因，而不是问题的症状**。
- ◆ 能够集中于问题的实质内容，而不是问题的历史或不同的个人观点。
- ◆ 以**团队**努力，聚集并攻克复杂难题。
- ◆ 辨识导致问题或情况的所有原因，并从中找到根本原因。
- ◆ 分析导致问题的**各原因之间相互的关系**。
- ◆ 采取补救措施，正确行动。

# 鱼骨图（因果图）

## 鱼骨图的分类

整理问题型



各要素与特性值间不存在原因关系，而是结构构成关系。

原因型



鱼头在右，特性值通常以“为什么……”来写。

对策型



鱼头在左，特性值通常以“如何提高/改善……”来写。

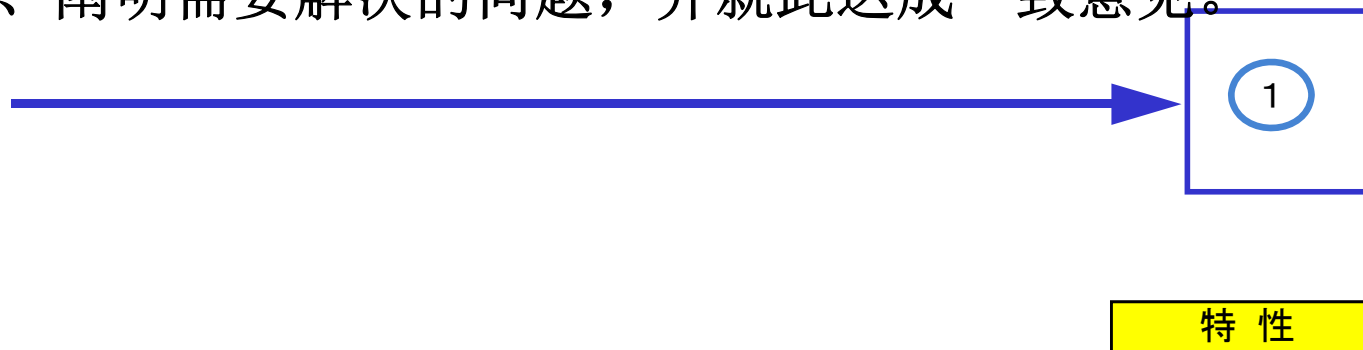
# 鱼骨图（因果图）

## 鱼骨图的分析步骤

第一步，决定问题的特性，简单的说特性就是“工作的结果”。

首先，对团队成员讲解会议目的；

然后，认清、阐明需要解决的问题，并就此达成一致意见。

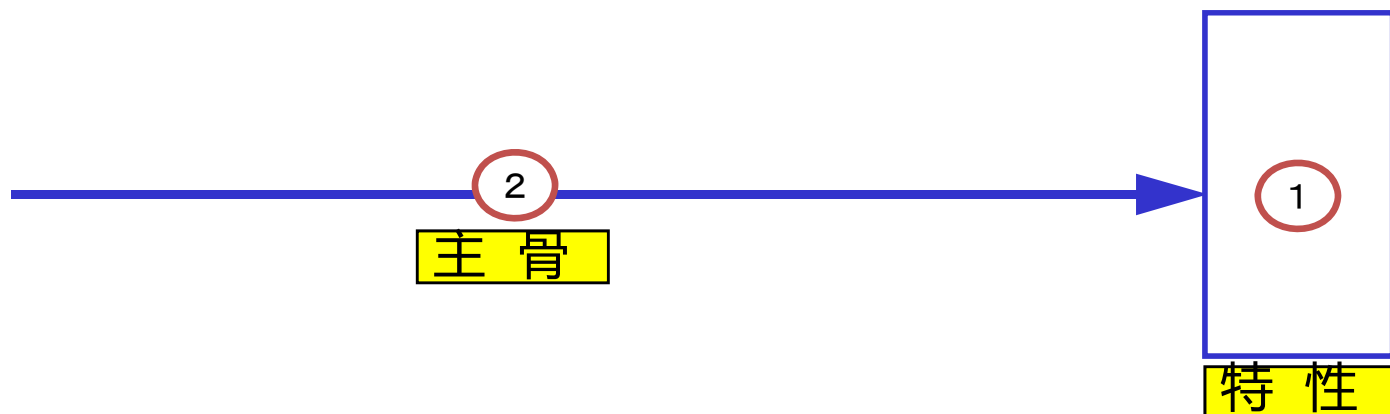


# 鱼骨图（因果图）

## 鱼骨图的分析步骤

第二步，确定特性和主骨。

特性写在右端，用四方框圈起来。主骨用粗线画，加箭头标志。

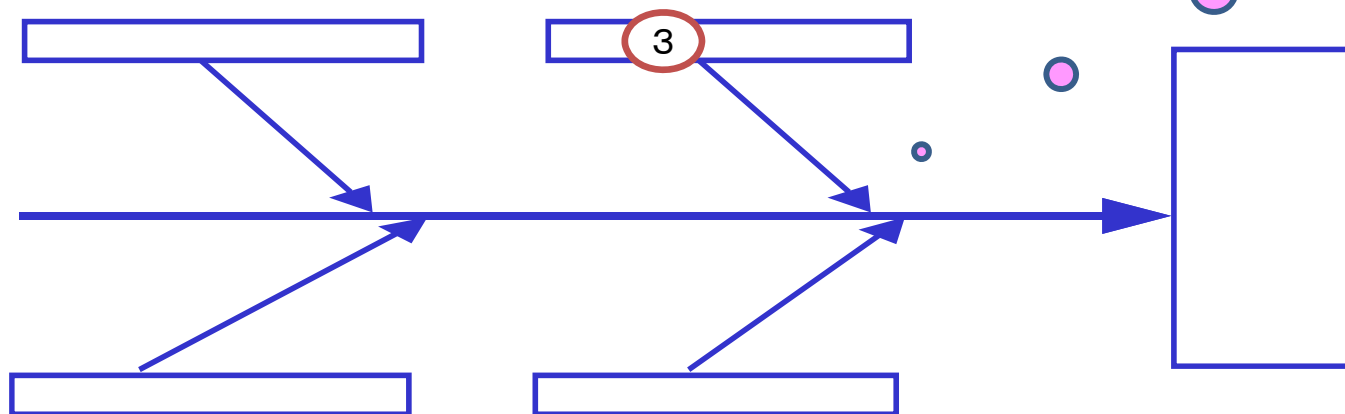


# 鱼骨图（因果图）

## 鱼骨图的分析步骤

第三步，确定大骨和要因。

大骨上分类书写3~6个要因，用四方框圈起来。



要点：绘图时，应保证大骨与主骨成60度夹角，中骨与主骨平行。



# 鱼骨图（因果图）

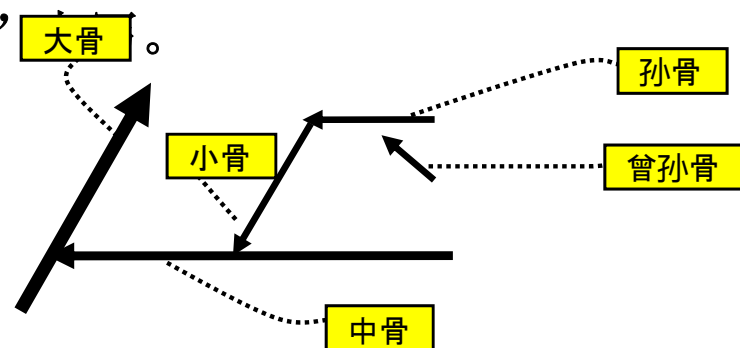
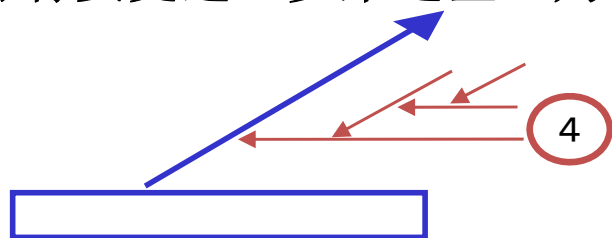
## 鱼骨图的分析步骤

第四步，确定中骨、小骨、孙骨。

中骨「事实」。（不从事实开始的话，要做出对策的要因的真实味就淡了。）

小骨要围绕“为什么会那样？”来写。

孙骨要更进一步来追查“为什么会那样？”



# 鱼骨图（因果图）

## 鱼骨图的分析步骤

第五步，中骨、小骨、孙骨的“要点”。

反复寻找为什么，摸索探询的话比较好；

要因是「主语＋谓语」的形式比较好。

例： 「软管 → 软管长」

「涂料 → 涂料飞溅」

现地现物前要围绕事实系统地整理要因。

# 鱼骨图（因果图）

## 鱼骨图的分析步骤

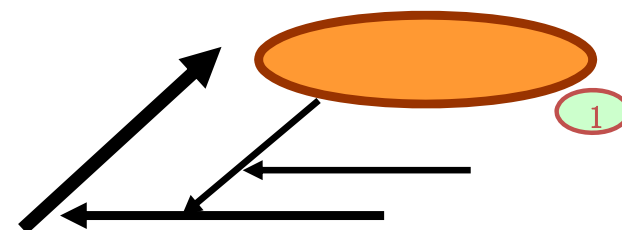
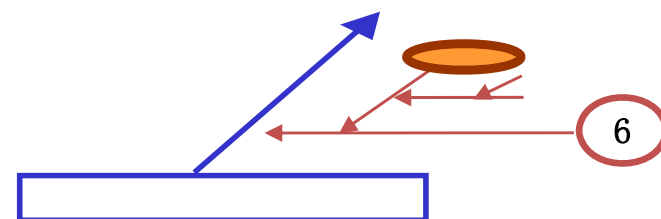
### 第六步，深究要因。

考虑对特性影响的大小和对策的可能性，深究要因（不一定是最后的要因）。

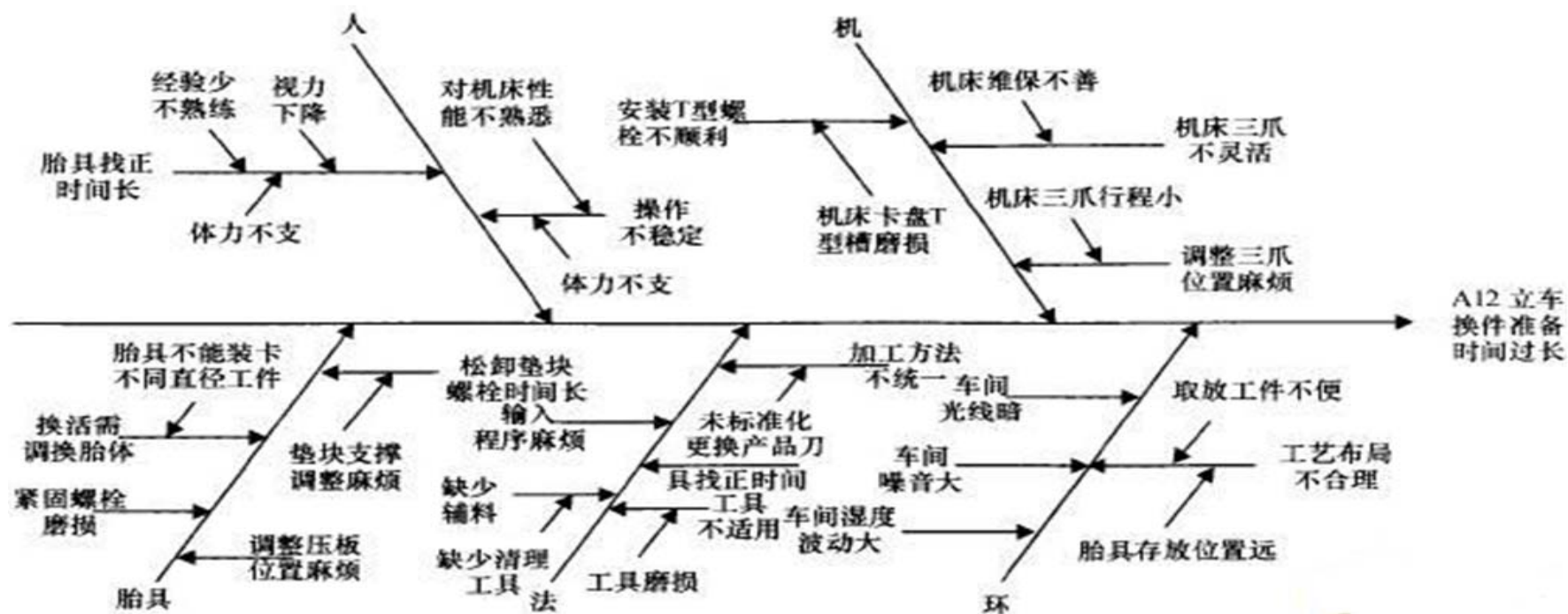
追查要因的时候，要由**全员讨论**决定。

将深究的要因称为主要因，用○标记。

主要因是对策的内容，用事实和数据确认。



# 鱼骨图（因果图）

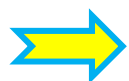


# 课程内容

1 过程分析概述

2 流程图

3 因果图



4 因果矩阵图

P137

5 其他过程分析工具和文档

# 因果矩阵

## 因果矩阵的概念

因果矩阵，是当预期解决的问题比较复杂，有多种缺陷形式，且它们的影响互相关联，无法将它们分开来考察和解决时，可以帮助管理团队选择重点关注的过程输入或影响因素，以便于有针对性地收集数据进行分析的一种有效地分析工具。

- 一个简化版的QFD(质量功能展开)，强调了解客户需求的重要性；
- 用流程图作为主要工具来联系关键的输入和输出(客户需求):
  - a) 关键输出根据对客户的重要性来评分
  - b) 关键输入根据与输出的关系来评分

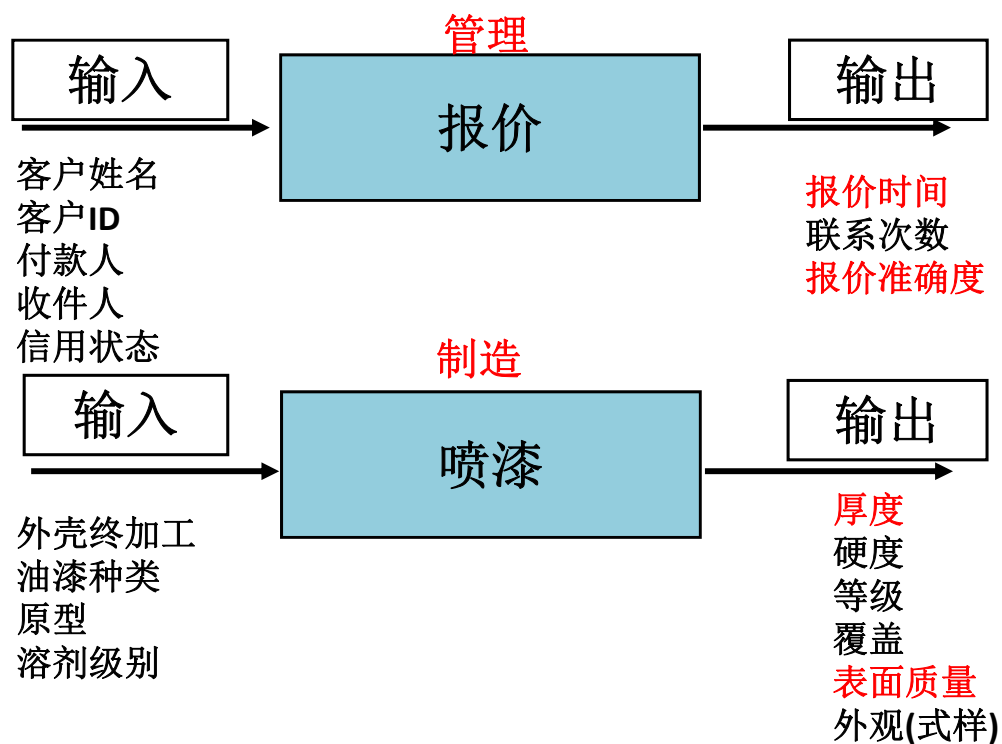
# 因果矩阵

## 因果矩阵分析步骤

- ① 在矩阵图的上方填入过程**输出缺陷的形式或关键过程输出变量**；
- ② 确定每一输出特性或缺陷形式的重要度，并给定其**权重**（1~10,10代表的重要度最高）；
- ③ 在矩阵图的左侧，列出输入变量或**所有可能的影响因素**；
- ④ **评价**每一输入变量或影响因素对各个输出变量或缺陷的相关关系，矩阵图中的单元格用于表明该行对应的输入变量的相关程度，一般将这种相关程度分为四类，并分别赋予0,1,3,9的分支，表明其不同的相关程度；
- ⑤ **计算**.评价过程输入变量或影响因素的重要程度，将每一单元的相关程度得分值乘以该列对应的输出变量的权重数，然后将每一行的乘积加起来，这个结果代表了该输入变量或影响因素的权重；
- ⑥ 考察每个输入变量或影响因素的权重数，**权重较高的将是项目重点关注的对象**。

# 因果矩阵

## 因果矩阵示例





# 因果矩阵（表）

## 1.列出关键输出

客户：-----		对客户权重 (低 1- 高 10)											
流程步骤	流程输入	流程输出		1	2	3	4	5	6	7	8	合计	
				客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求		
1												0	
2												0	
3												0	
4												0	
5												0	
6												0	
7												0	
8												0	
9												0	

对大多数项目来说，使用**宏观的Y**

# 因果矩阵（表）

客户：-----		对客户权重 (低 1- 高 10)									合 计
		流 程 输 出	1	2	3	4	5	6	7	8	
流 程 步 骤	流 程 输 入		客户要求	客户要求	客户要求	客户要求	客户要求	要求	要求	要求	
1											0
2											0
3											0
4											0
5											0
6											0
7											0
8											0
9											0

2.按照对客户的重要性定出输出等级



## 因果矩阵（表）

[illegible]



# 因果矩阵（表）

## 联系输入与客户要求

- 相关分数：不超过4个等级：0，1，3和9或0，1，3和5
- 打分很浪费时间，因此特别要注意详细说明评分的标准
  - 0 = 不相关
  - 1 = 流程输入对客户需要影响很小
  - 3 = 流程输入对客户需要有一定影响
  - 9（5） = 流程输入对客户需要有直接，重要的影响

# 因果矩阵（表）

客户：-----		对客户的权重 (低 1- 高 10)									
流程步骤	流程输入	流程输出		1	2	3	4	5	6	7	8
		客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求	客户要求
1											
2											
3											
4											
5											
6											
7											
8											
9											

5.相交乘积和定  
优先级

0  
0  
0  
0  
0  
0  
0  
0  
0

对所有要求将权重评分和相关性分值相乘并加和

# 试题

13. 下表表示某企业拟选的六西格玛项目及项目选择需要考虑的因素和权重，请根据下表，选出项目优先顺序（从高到低）的正确排序。

	对质量的影响 (0.5)	对周期的影响 (0.2)	对成本的影响 (0.2)	对管理水平的影响 (0.1)
项目 1	◎	△		△
项目 2		◎	△	○
项目 3	○	△	△	◎
项目 4	△	○	○	

◎强相关 (9 分)    ○较强相关 (3 分)    △弱相关 (1 分)

◎强相关 (9 分)    ○较强相关 (3 分)    △弱相关 (1 分)

- A. 项目 1、项目 3、项目 4、项目 2
- B. 项目 2、项目 4、项目 1、项目 3
- C. 项目 1、项目 3、项目 2、项目 4**
- D. 项目 1、项目 2、项目 3、项目 4

项目1:  $9*0.5+1*0.2+1*0.1=4.8$   
 项目3:  $3*0.5+1*0.2+1*0.2+9*0.1=2.8$   
 项目2:  $9*0.2+1*0.2+3*0.1=2.3$   
 项目4:  $1*0.5+3*0.2+3*0.2=1.4$



# 试题

22. 关于因果矩阵，下列陈述正确的是： D

- A. 因果矩阵可筛选出对 CTQ 最为重要的输入变量
- B. 因果矩阵是精益六西格玛中一种倾听各方意见的方法
- C. 因果矩阵是一种找出满足顾客需求方案的方法
- D. 以上说法均正确**

题目：关于因果图和因果矩阵，下列说法错误的是： B

- A. 因果图一般只分析影响一个结果（Y）的多个可能原因
- B. 因果矩阵一般只分析影响一个结果（Y）的多个可能原因**
- C. 因果矩阵可以帮助方便地确定关键的影响因素
- D. 对于同一个问题，可以先绘制因果图，再绘制因果矩阵





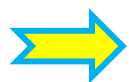
# 课程内容

1 过程分析概述

2 流程图

3 因果矩阵图

4 因果图



5 其他过程分析工具和文档

P137

# 其他过程分析工具和文档

- 1) **程序文件**：通过了ISO9000标准认证的组织都建立了这样的文件体系，用以规范生产、服务和管理活动；
- 2) **检查单**：用来显示过程的每一步或检查的每一项内容等。例如，飞行员使用的检查单、建筑作业需要的材料检查单，用来确保每一间客房准备就绪的酒店清洁检查单等；
- 3) **照片**：显示产品或服务在过程的关键阶段的外观性状。例如，电信装置中带有色标的导线在塑封前的终端位置、装配操作的工作站布局、零件组装的照片等；
- 4) **图表**：用来显示多个部件之间的关系。例如，一个“部件展开”或“安装”图表显示了各个部件是如何安装成一个更大部件的。
- 5) **录像**：用来显示过程的某一阶段的操作过程；
- 6) **计算机数据库**：操作者在工作站用来“提取”工作所需的图纸、程序/作业指导书和规范；
- 7) **交互的计算机程序**：在工作现场的计算机终端，允许操作者直接输入数据或从过程中接收数据等。

# 测量（上） 数据收集和整理

---

中国商飞六西格玛黑带培训课程

主讲人： 宋丽静（上飞公司）

联系方式： 20853895

# 课程内容

1

数据类型和测量尺度

P162

2

收集数据的方法

3

抽样方法

4

描述性统计方法

5

数据的图示方法

# 数据的类型和测量尺度

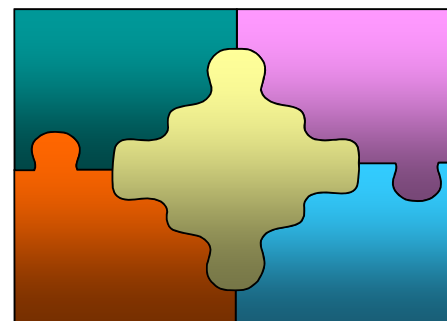
测量数据类型：分为**连续型数据**和非连续或**离散型**数据。

类型	连续型数据	非连续或离散型数据
别称	计量值数据	计数数据
敏感性	比较敏感地反映过程的变化，包含的信息丰富	在反映过程变化方面不如连续数据敏感
样本量	较少	较大或较长周期
测量手段和精度	要求较高	要求较低
成本	高	低
举例	长度、温度或湿度	合格/不合格、通过/失败，是/否、接受/不接受、好/坏以及缺陷个数

# 数据的类型和测量尺度

四类测量尺度：**定类（名义）**、定序、定距、定比。

1. 计量层次最低
2. 对事物进行平行的分类
3. 各类别可以指定数字代码表示
4. 使用时必须符合类别**穷尽和互斥**的要求
5. **数据表现为“类别”**
6. 具有=或 $\neq$ 的数学特性



# 数据的类型和测量尺度

四类测量尺度：**定类（名义）**、定序、定距、定比。

- 按现象性质差异进行的辨别与区分。测量结果形成**定类变量或定类指标**。
- 定类变量或指标确切的值是**以文字表述的**，可以用数值标识，但仅起标签作用。
- 定类变量或指标的各类别间是**平等**的，没有高低、大小、优劣之分。



性别



种族

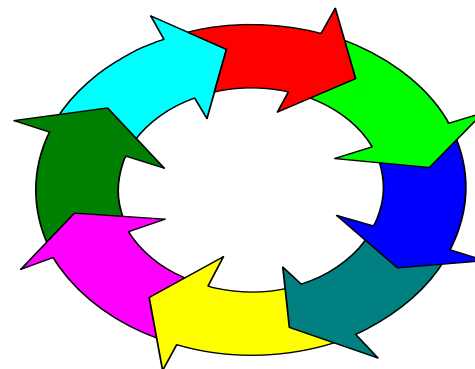


运动项目

# 数据的类型和测量尺度

四类测量尺度：定类（名义）、**定序**、定距、定比。

1. 对事物分类的同时给出各类别的顺序
- 2. 比定类尺度精确**
3. 未测量出类别之间的准确差值
4. 数据表现为“类别”，但有序
- 5. 具有>或<的数学特性**





# 数据的类型和测量尺度

四类测量尺度：定类（名义）、**定序**、定距、定比。

- 按现象顺序差异进行的辨别与区分。测量结果形成**定序变量**或**定序指标**。
- 定序变量或指标确切的**值是以文字表述**的，也可以用数值标识，也仅起标签作用。
- 定序变量或指标**各类别间有高低优劣之分**，不能随意排列。



>

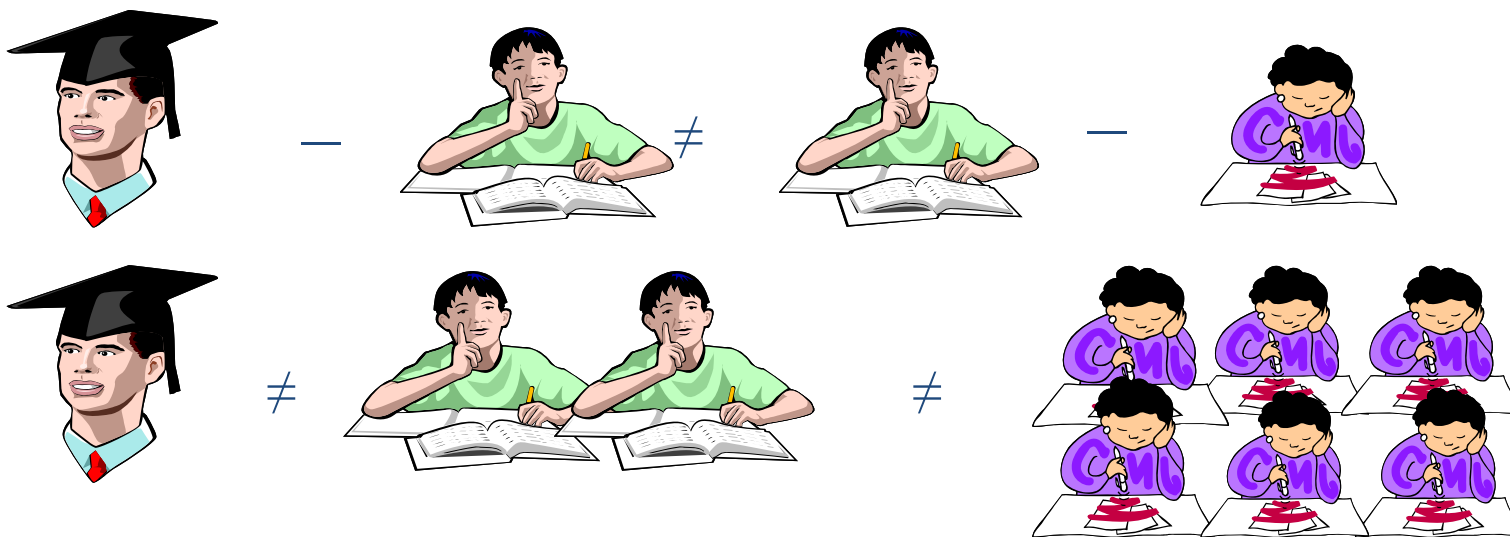


>



# 数据的类型和测量尺度

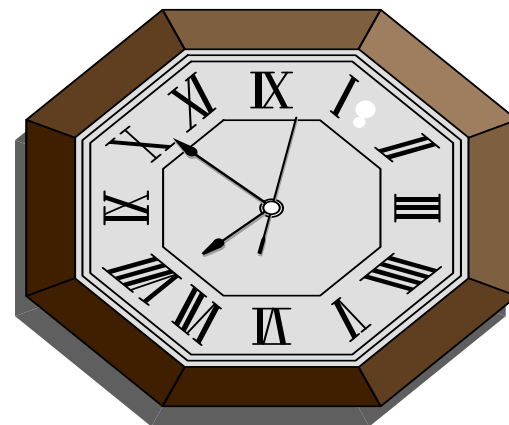
四类测量尺度：定类（名义）、**定序**、定距、定比。



# 数据的类型和测量尺度

四类测量尺度：定类（名义）、定序、**定距**、定比。

1. 对事物的准确测度
2. 比定序尺度精确
3. 数据表现为“数值”
4. 没有绝对零点
5. 具有 + 或 - 的数学特性



# 数据的类型和测量尺度

四类测量尺度：定类（名义）、定序、**定距**、定比。

- 按现象绝对数量差异进行的辨别与区分。测量结果形成**定距变量或定距指标**。
- 定距变量或指标的值**以数字表述**，有**计量单位**，可以进行**加减运算**。
- 定距变量或指标各类别间自然有大小之分，但**没有绝对的零点**，**不能乘除计算**。

天气预报：沈阳：最高温度 $3^{\circ}\text{C}$ ，最低 $-7^{\circ}\text{C}$

上海：最高温度 $15^{\circ}\text{C}$ ，最低 $-1^{\circ}\text{C}$

两地**最高温度**相差 $12^{\circ}\text{C}$  ✓

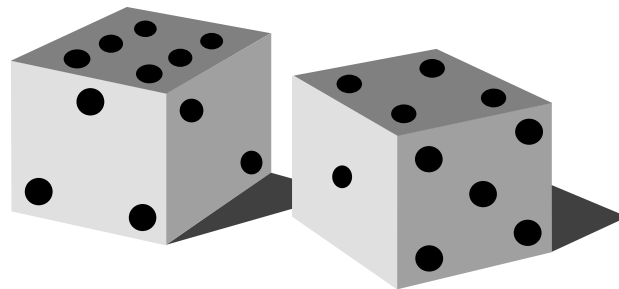
沈阳**最低温度**较上海**最低温度**低 $6^{\circ}\text{C}$  ✓

~~上海最高温度是沈阳最高温度的5倍~~ ✗

# 数据的类型和测量尺度

四类测量尺度：定类（名义）、定序、定距、**定比。**

1. 对事物的准确测度
2. 与定距尺度处于同一层次
3. 数据表现为“数值”
4. 有绝对零点
5. **具有  $\times$  或  $\div$  的数学特性**



# 数据的类型和测量尺度

## 四类测量尺度：定类（名义）、定序、定距、**定比。**

- 按现象绝对差异与相对差异进行的辨别与区分。测量结果形成定比变量或定比指标。
- 定比变量或指标确切的值也以数字表述，有计量单位，可以加减。
- 定比变量或指标有绝对意义上的零点，既可以加减，可以乘除运算。

**某公司产品2015年产量为5000辆，2016年年产量为为10000辆，2016年产量为2015年的两倍。**

# 数据的类型和测量尺度

四种计量尺度的比较				
	定类尺度	定序尺度	定距尺度	定比尺度
分类 ( $=$ , $\neq$ )	✓	✓	✓	✓
排序 ( $<$ , $>$ )		✓	✓	✓
间距 ( $+$ , $-$ )			✓	✓
比值 ( $\times$ , $\div$ )				✓

# 试题

题目：黑带在改进某一业务流程中确定了以下测量量，请问哪个测量指标是定序测量尺度的指标：

- A. 准时交付率      B. 人工成本      C. 人员技能等级      D. 客户抱怨数量



将降雨量按“旱”、“偏少”、“中等”、“偏多”、“涝”分为5类，这种分类方法属于下列哪个测量尺度？ B

- A. 定类      B. 定序      C. 定距      D. 定比



# 课程内容

1 数据类型和测量尺度

2 收集数据的方法

3 抽样方法

4 描述性统计方法

5 数据的图示方法

P163

# 收集数据的方法

## □ 检查表

- 检查表是收集过程数据时最常用的工具，常用于DMAIC改进方法的测量阶段。用于记录缺陷的频次，失效的类型、时间等，是分析原因时信息的重要来源。

一般而言检查表可依其工作的目的或种类分为下述两种。

**点检用检查表：**在设计时就已定义，使用时，只做是非或选择的标记，其主要功能在于确认作业执行、设备仪器保养维护的实施状况或预防事故发生，以确保使用时安全用。

**记录用检查表：**此类检查表是用来收集计划资料，用于不合格原因和不合格项目的记录，由于常用于作业缺点，质量差异等记录，所以也称为改善用检查表。

# 收集数据的方法

## ◆ 点检用检查表

10000 KM 时定期保养	
顾 客 名:	日 期:
车牌号码:	费 用:
车 种:	行驶公里:
作 业 者:	
<input type="checkbox"/> 电瓶液量	<input type="checkbox"/> 空气滤清器
<input type="checkbox"/> 水箱	<input type="checkbox"/> 机油
<input type="checkbox"/> 胎压	<input type="checkbox"/> 分电盘盖
<input type="checkbox"/> 火星塞	<input type="checkbox"/> 化油器
<input type="checkbox"/> 风扇皮带	

注: √检查 ○调整 X更换

# 收集数据的方法

## ◆ 记录用检查表

No	组 界	组中心点	检 查	次数
1	119.5~126	122.75	####	5
2	126~132.5	129.25	#### ##	10
3	132.5~139	135.75	#### ## //	12
4	139~145.5	142.25	#### ## ## ## ## ## /	31
5	145.5~152	148.75	#### ## ## ## ///	23
6	152~158.5	155.25	#### ##	10
7	158.5~165	161.75	####	5
8	165~171.5	168.25	/	1
合 计				97

# 课程内容

1 数据类型和测量尺度

2 收集数据的方法

3 抽样方法

4 描述性统计方法

5 数据的图示方法

P164

# 抽样方法

## □ 简单随机抽样

- ◆ 按随机原则直接从总体中抽取样本，构造统计量，并对总体相应的指标做出统计推断。

## □ 分层抽样

- ◆ 又称为类型抽样或分类抽样。是先将总体按一定的标识加以分层，例如**按不同的设备、材料、供应商、版次、年龄组等分层**，然后在各层中按简单随机抽样抽取若干个样本，再由各层的样本组成一个总体样本。

# 抽样方法

- 简单随机抽样

- ✓ 必须满足两个基本条件：

- 等可能性。若总体有 $N$ 个个体，每个个体被抽到的概率为 $1/N$ ；
    - 独立性。是指每次抽样是相互独立的。

- ✓ 随机抽样的方法有：

- 抽签法：当总体个数不多时，适宜采用这种方法；
    - 滚球法（即摇奖方法）；
    - 计算机模拟法；
    - 随机数表法。

# 抽样方法（简单随机抽样）

## ◆抽签法

将总体全部单位编号，并写在**物理形状相同**的签条（小球或卡片）上。混合均匀，随机抽取（重复/不重复），技术简单。





# 抽样方法（简单随机抽样）

## ◆滚球法

即为人们最熟悉的摇奖法。



# 抽样方法（简单随机抽样）

## ◆计算机模拟法

在中大型计算机中，一般都带有随机数字发生器。

例：某射手中靶的概率为**0.8**，每次射击**10**发子弹，射击**10**次，模拟每次中靶的次数。

“数据 | 分析 | 数据分析 | 随机数发生器”，选择“分布”为“二项”，设置对话框如下：



随机数发生器

变量个数 (V): 10

随机数个数 (R): 1

分布 (D): 二项式

参数

P (A) = 0.8

试验次数 (N) = 10

随机数基数 (B):

输出选项

☒ 输出区域 (O): \$A\$1

☐ 新工作表组 (P):

☐ 新工作簿 (W)

确定 取消 帮助 (H)

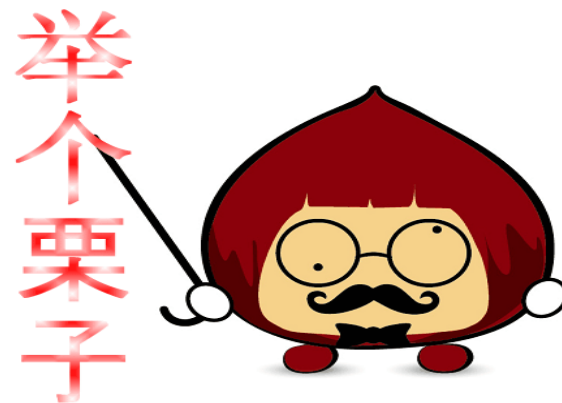
	A	B	C	D	E	F	G	H	I	J
1	8	7	8	9	9	5	7	6	8	10

# 抽样方法（简单随机抽样）

## ◆随机数表法

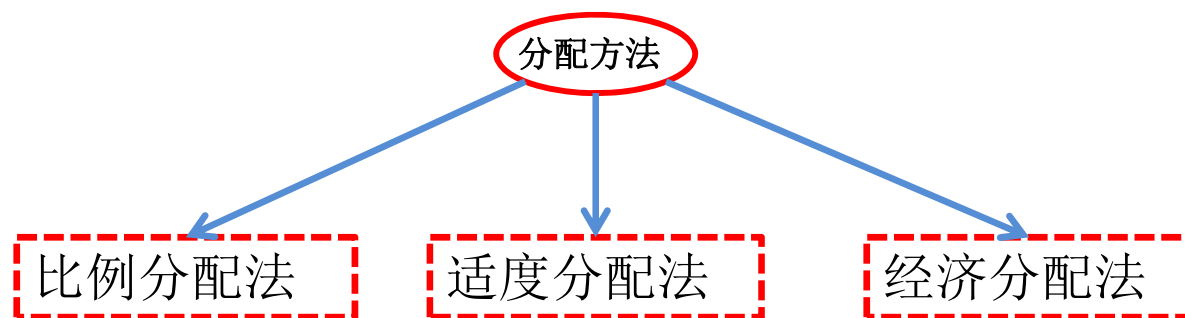
例如：某系有920名学生，抽取50个学生进行家庭收入情况调查。运用随机数表过程如下：

1. 将920名学生，由001-920号码编号。
2. 确定起始页，起始行，起始列。例如：1，16，31。
3. 抽取样本。



# 抽样方法（分层抽样）

- 分层抽样，又称为**类型抽样或分类抽样**。这种抽样形式是先将总体按一定的标识加以**分层**，例如按不同的设备、材料、供应商、班次、年龄组等分层。然后**在各层中按简单随机抽样取若干样本**，再由各层的样本组成一个总体样本。



若设总体含有 $N$ 个样本，将总体划分成 $k$ 层，每层总体样本为 $N_i$  ( $i = 1, 2, \dots, k$ )，则有 $\sum_i^k N_i = N$ ，从各层总体 $N_i$ 中随机抽取 $n_i$ 个样本，则样本 $n = \sum_i^k n_i$ 。

# 课程内容

1 数据类型和测量尺度

2 收集数据的方法

3 抽样方法

4 描述性统计方法

5 数据的图示方法

P166

# 描述性统计方法

P166

## 常见的描述性统计

样本统计量	说明	公式/记号
分布的位置或者中心趋势的度量		
样本均值	取自分布的样本的重心或质心	
中位数	有序样本中，排在中间位置的数据。数据集中的一半数据高于中位数，一半低于中位数	
众数	出现最频繁的数值。如果将数据分组，众数是频数最大的组	Mode

# 描述性统计方法

P166

常见的描述性统计		
样本统计量	说明	公式/记号
散程度的度量		
极差（全距）	样本极值之间的距离	$R = \text{最大值} - \text{最小值}$
样本方差	对围绕均值波动的度量	
样本标准差	对围绕均值波动的度量	
形状的度量		
偏度	对称性分布的 $\beta_3$ 为零，若 $\beta_3 > 0$ ，表示右偏态；若 $\beta_3 < 0$ ，表示左偏态	
峰度	峰度是分布平坦程度的度量。正态分布的峰度为 $\beta_4 = 0$	

# 课程内容

1 数据类型和测量尺度

2 收集数据的方法

3 抽样方法

4 描述性统计方法

5 数据的图示方法

P169



# 数据的图示方法

- 直方图

将一个变量的不同等级的相对频数用矩形块标绘的图表(每一矩形的面积对应于频数)

- 直方图作用

数据和规格的关系

数据的分布形状

数据分布的中心位置

数据的分散程度

# 数据的图示方法

P169

- 直方图（绘制）

第一步：收集数据，求极差R

在原始数据中找出最大值 $X_{\max}$ 和最小值 $X_{\min}$ ，计算二者差值，即级差。

第二步：确定分组的组数和组距

一批数据究竟分多少组，通常根据数据的多少而定，可参考下表。

样本量	组数 (k)
40~100	5~8
100~200	8~10
201~500	9~11
501~1000	10~13
1000以上	12~15

# 数据的图示方法

- 直方图（绘制）

## 第三步：确定各组界限

为避免数据落在组界上，组界值的末位数应取测量值单位的 $1/2$ 。

为了计算的需要，往往要决定各组的中心值。每组上下界限相加除以2，所得的数据即为组中值。组中值为各组数据的代表值。

## 第四步：作频数分布表

将测得的原始数据分别归入相应组中，统计各组数据个数，即频数 $f_1$ ，填好各组频数后，检验总数是否与数据总数相符，避免重复或遗漏。

## 第五步：画直方图

横坐标表示质量特性，纵坐标为频率/组距（或频率），在横轴上标明各组组界，以组距为底，频数为高画出一系列的直方柱，就成了直方图。

# 试题

41. 下列说法正确的是：

- A. 过程处于受控状态，则该过程生产出的产品一定能满足规格要求；
- B. 根据直方图能够判断过程是否处于受控状态；
- ☒ C. 直方图不能反映过程随时间的变化规律；
- D. 以上都不对

16. 关于直方图的说法，下列表述不正确的是：

- A. 直方图可以直观地表示数据的分布特征；
- B. 通过直方图，可以初步分析数据分布是否存在异常；
- C. 样本数据越多，直方图所显示的数据分布与实际分布越接近；
- ☒ D. 直方图的形状与数据分组的个数无关



使用直方图应注意的问题：

- (1) 如果直方图上画有规格限，那么直方图可用来比较过程与要求。要达到标，应确认直方图具有合适的比例。
- (2) 直方图最好能与其他图，如趋势图或控制图结合使用。这是因为直方图不是按时间顺序给出的，无法反映过程随时间变化的情况。
- (3) 根据数据来源的不同，应对数据分层，分别绘制直方图。

## 2. 茎叶图

茎叶图 (stem-and-leaf plots) 是直方图的变种，适用于较小的数据集 ( $n < 200$ )。与直方图相比，茎叶图的一个主要优点是全部或者部分地保留了原始数据的信息，使我们掌握有关总体的更多信息。

# 数据的图示方法

P<sub>170</sub>

## ● 茎叶图

茎叶图是直方图的变种，适用于较小的数据集（ $n < 200$ ）。

优点：全部或部分地保留了**原始数据的信息**，从而可以使我们掌握有关总体的更多信息。

高位部分称为“茎”

低位部分称为“叶”

1. 将每个数据分为茎（高位）和叶（低位）两部分；
2. 将最小茎和最大茎之间的数按大小次序排成一行，写在左（右）侧；
3. 将各个数据的叶按大小次序写在其茎右（左）侧。

# 数据的图示方法

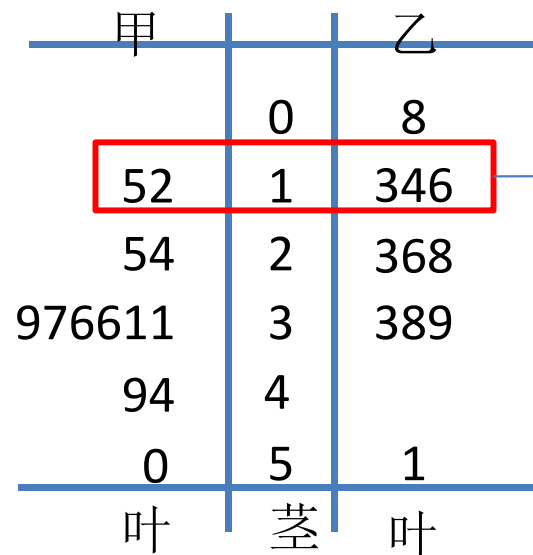
- 茎叶图

例：甲乙两人比赛得分记录如下：

甲：49, 24, 12, 31, 50, 31, 44, 36, 15, 37, 25, 36, 39

乙：13, 51, 23, 8, 26, 38, 16, 33, 14, 28, 39

用茎叶图表示两人成绩，说明哪一个成绩好。



表示甲得分为15分、12分，乙得分为13分，14分、16分。其他各行与此类同。

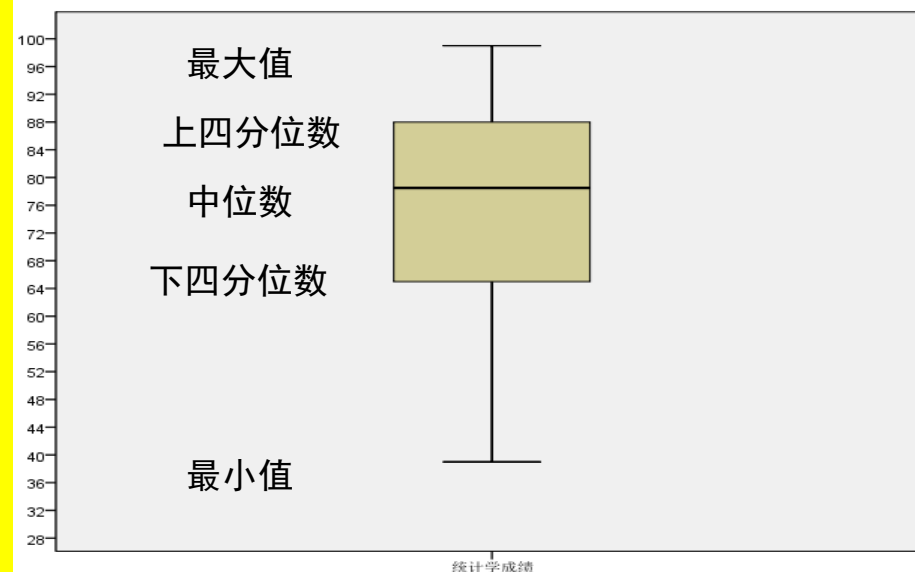
从这个茎叶图可以看出：甲运动员的得分大致对称，平均得分及中位数、众数都是30多分；乙运动员的得分除一个51外，也大致对称，平均得分及中位数、众数都是20多分，因此，甲运动员发挥比较稳定，总体得分情况比乙好。

# 数据的图示方法

箱线图由一组数据的最大值、最小值、中位数、两个四分位数这5个值绘制而成的。

## ● 箱线图

1. 先按从小到大的顺序将数排列，再依次找出最大、最小值、中位数Me、下四分位数QL、上四分位数QU，中位数位置 =  $(n+1)/2$ ，下四分位数位置 =  $n/4$ ，上四分位数位置 =  $3n/4$ 。
2. 连接两个四分位数画出箱子。
3. 再将两个极限点与箱子相连接，中位数在箱子中间。





## 数据的图示方法

从某大学经济管理专业二年级学生中随机抽取11人，对8门主要课程的考试成绩进行调查，所得结果如表。试绘制各科考试成绩的多批比较箱线图，并分析各科考试成绩的分布特征。

。

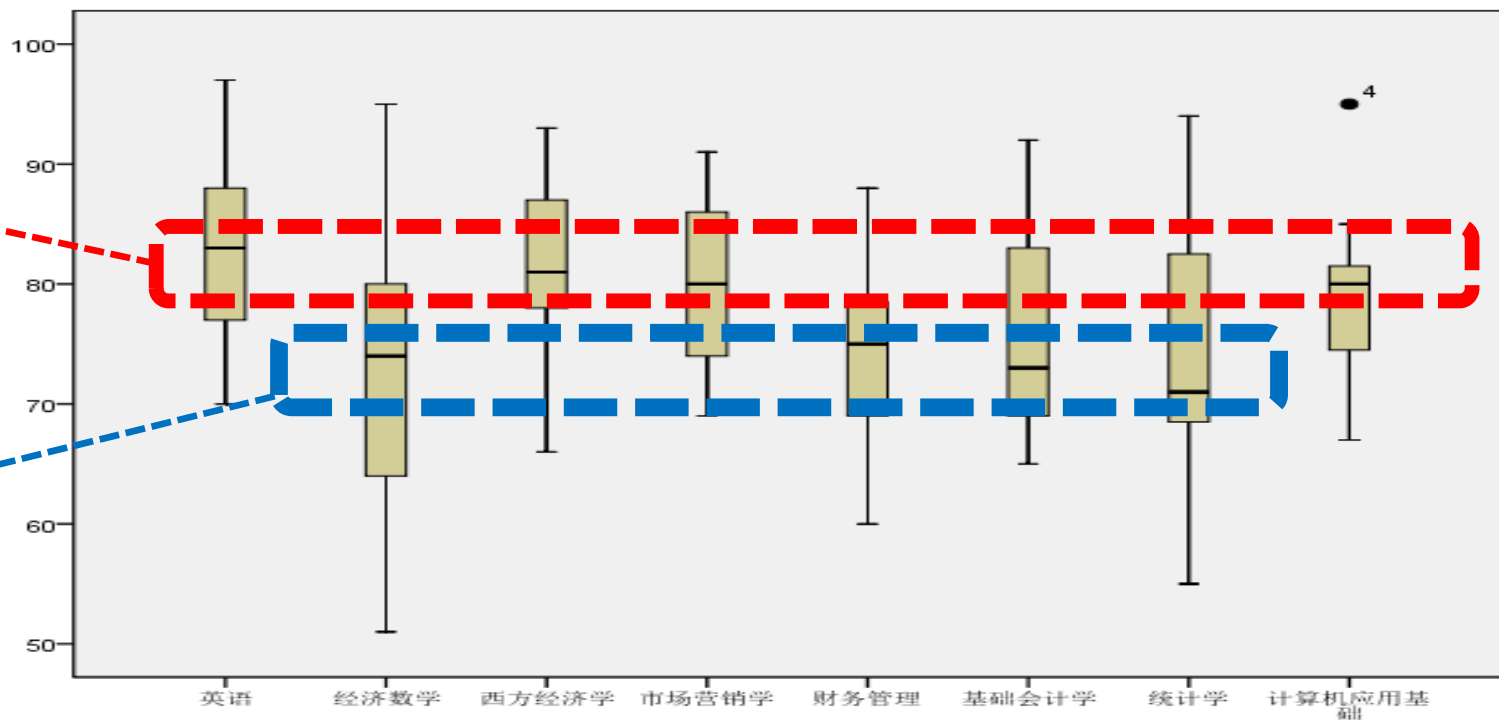
课程名称	学生编号										
	1	2	3	4	5	6	7	8	9	10	11
英语	76	90	97	71	70	93	86	83	78	85	81
经济数学	65	95	51	74	78	63	91	82	75	71	55
西方经济学	93	81	76	88	66	79	83	92	78	86	78
市场营销学	74	87	85	69	90	80	77	84	91	74	70
财务管理	68	75	70	84	73	60	76	81	88	68	75
基础会计学	70	73	92	65	78	87	90	70	66	79	68
统计学	55	91	68	73	84	81	70	69	94	62	71
计算机	85	78	81	95	70	67	82	72	80	81	77

# 数据的图示方法

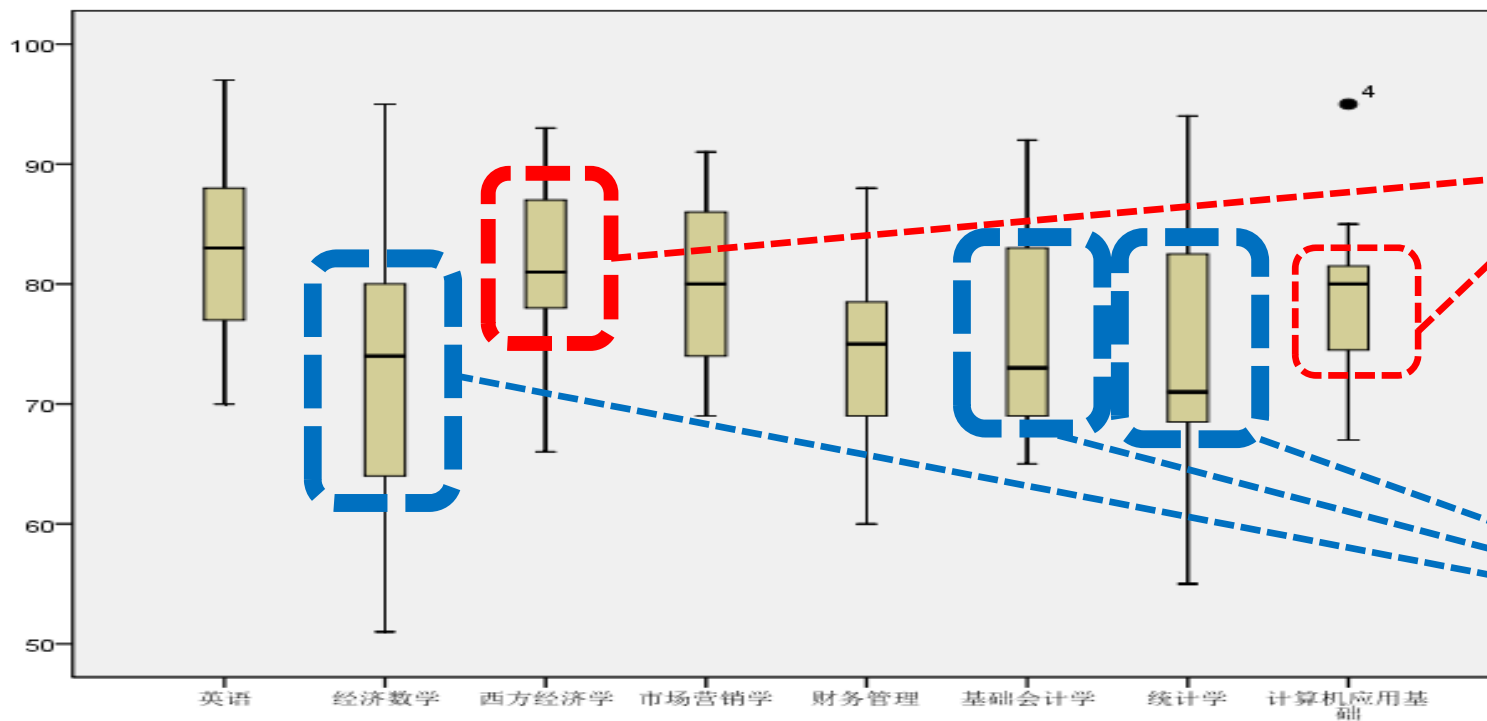
## 中位数的高低

英语、计算机、西方经济学和市场营销学平均成绩较高

统计学、经济数学、财务管理和基础会计学平均成绩较低



# 数据的图示方法



## 箱子的长短

计算机、西方经济学等  
课程成绩相对集中

经济数学、统计学、基  
础会计学等课程成绩相  
对分散

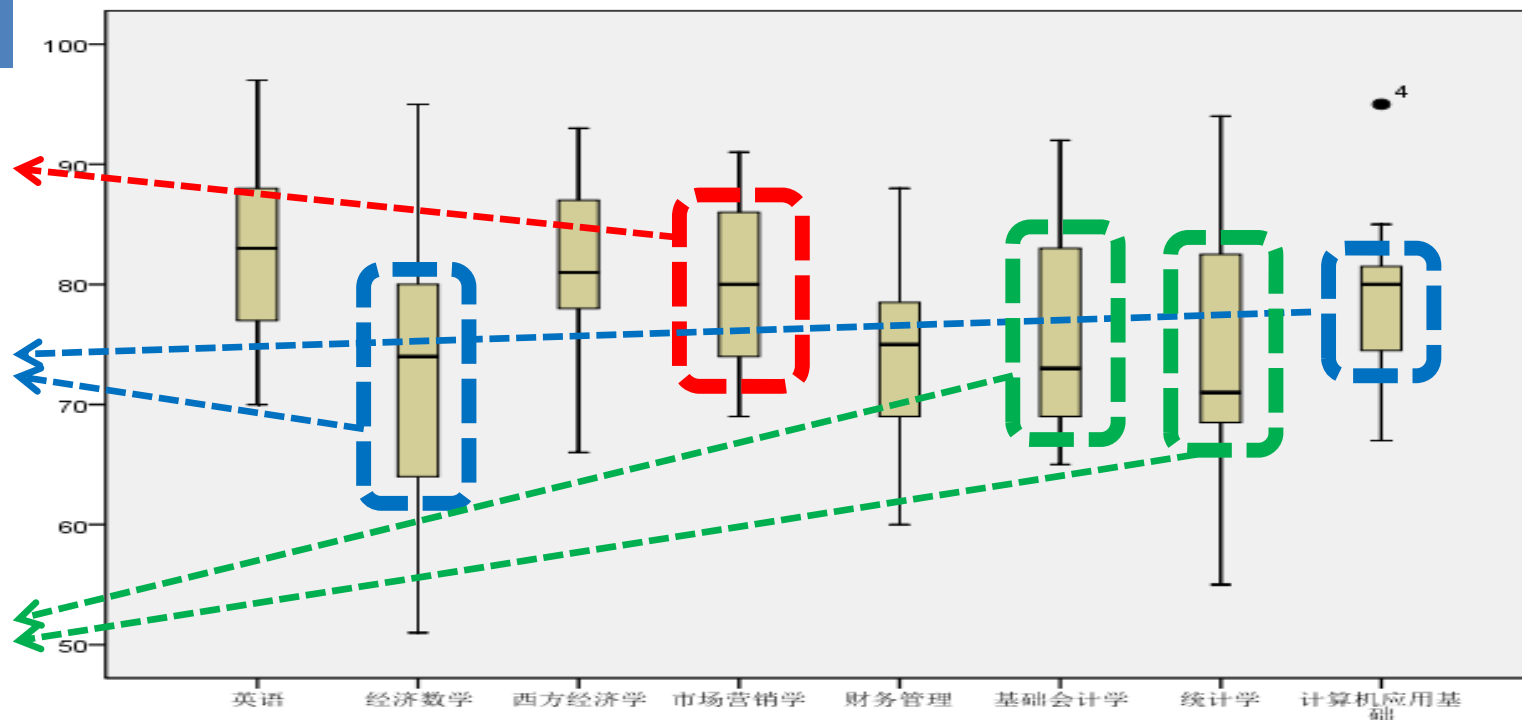
# 数据的图示方法

## 中位数与箱子位置

市场营销学等课程为正常分布

经济数学、计算机等课程呈左偏分布

统计学、基础会计学等课程呈右偏分布



2. 在箱线图 (Box-Plot) 分析中, 已知最小值=-4;  $Q_1=1$ ;  $Q_3=4$ ; 最大值=7; 则正确的说法是: **A**
- A. 上须触线终点为: 7; 下须触线终点为: -3.5
  - B. 上须触线终点为: 8.5; 下须触线终点为: -3.5
  - C. 上须触线终点为: 7; 下须触线终点为: -4
  - D. 上须触线终点为: 8.5; 下须触线终点为: -4

说明:

**下限**= $\max[Q_1-1.5(Q_3-Q_1), \text{最小观测值}] = \max[-3.5, -4] = -3.5$

**上限**= $\min[Q_3+1.5(Q_3-Q_1), \text{最大观测值}] = \min[8.5, 7] = 7$

若此公式的计算结果已超出观测值的极端值, 则须触线的终点分别到极端观测值为止;

若上下限的计算结果未超出观测值的极端值, 则箱线图上就会出现一些游离点, 这些点是数据中超出上下限的观测值, 有异常值的嫌疑。

# 数据的图示方法

- 链图（趋势图）

链图也称为趋势图，是一种特殊的散布图。它是显示任何测量特性随时间变化的图表。分析链图的目的是为了确认所出现的波动模式是由普通因素引起的，还是由特殊因素引起的。**链图可用于任何按时间序列组织的、连续尺度测量的数据图形分析。**

- 正态概率图

正态概率图用于检查一组数据是否服从正态分布。是实数与正态分布数据之间函数关系的散点图。如果这组实数服从正态分布，正态概率图将是一条直线。



THANK YOU !