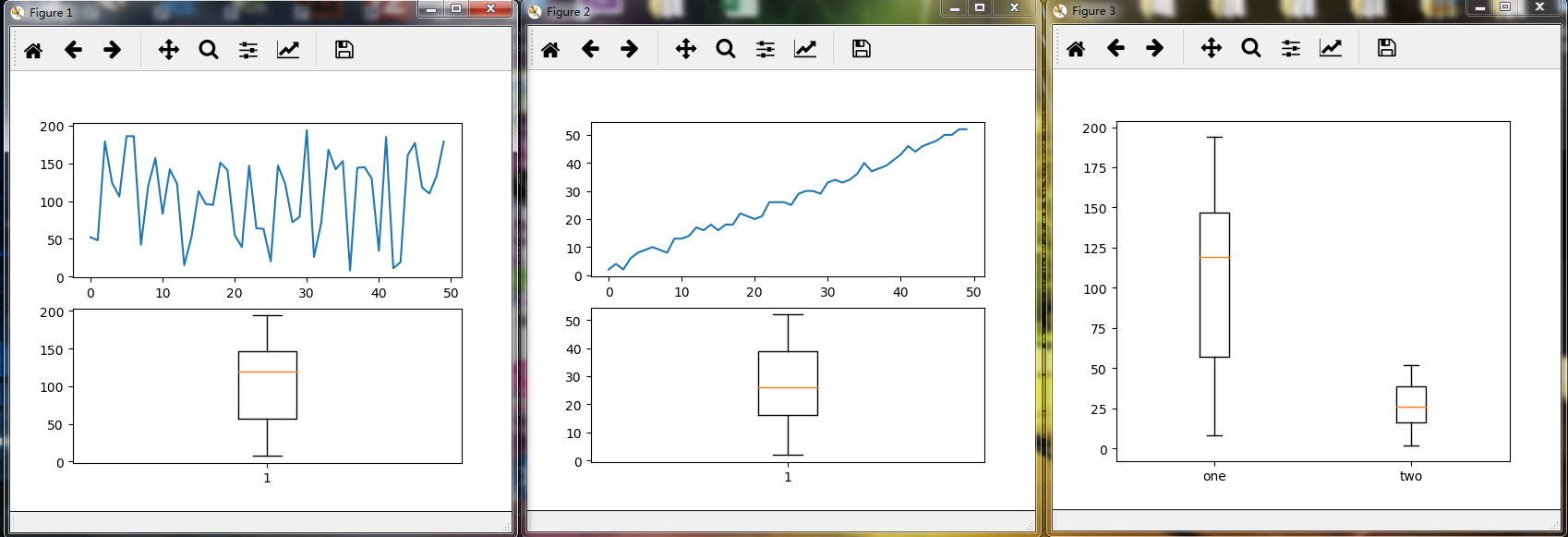
# 箱图

## 代码

**import** random  
**import** matplotlib.pyplot **as** plt  
**import** pylab **as** pl  
x = list()  
y = list()  
**for** i **in** range(50):  
 a = random.randint(0, 200)  
 x.append(a)  
 b = random.randint(0,5)  
 y.append(i+b)  
one = pl.figure(1)  
one1 = pl.subplot(211)  
one2 = pl.subplot(212)  
pl.sca(one1)  
pl.plot(x)  
pl.sca(one2)  
pl.boxplot(x)  
two = pl.figure(2)  
ax1 = pl.subplot(211)  
ax2 = pl.subplot(212)  
pl.sca(ax1)  
pl.plot(y)  
pl.sca(ax2)  
pl.boxplot(y)  
three = pl.figure(3)  
pl.boxplot((x,y),labels=(**'one'**,**'two'**))  
pl.show()

## 运行结果



## 结论

### 箱图意义

箱形图有5个参数：

下边缘（Q1），表示最小值；

下四分位数（Q2），又称“第一四分位数”，等于该样本中所有数值由小到大排列后第25%的数字；

中位数（Q3），又称“第二四分位数”等于该样本中所有数值由小到大排列后第50%的数字；

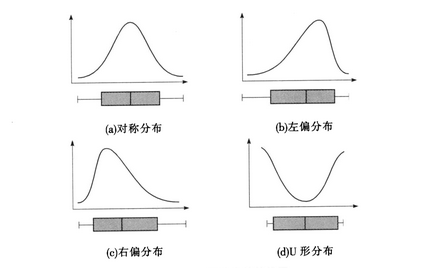
上四分位数（Q4），又称“第三四分位数”等于该样本中所有数值由小到大排列后第75%的数字；

上边缘（Q5），表述最大值。

第三四分位数与第一四分位数的差距又称四分位间距。

### 分析

1. 为了反映原始数据的分布情况，比如数据的聚散情况和偏态

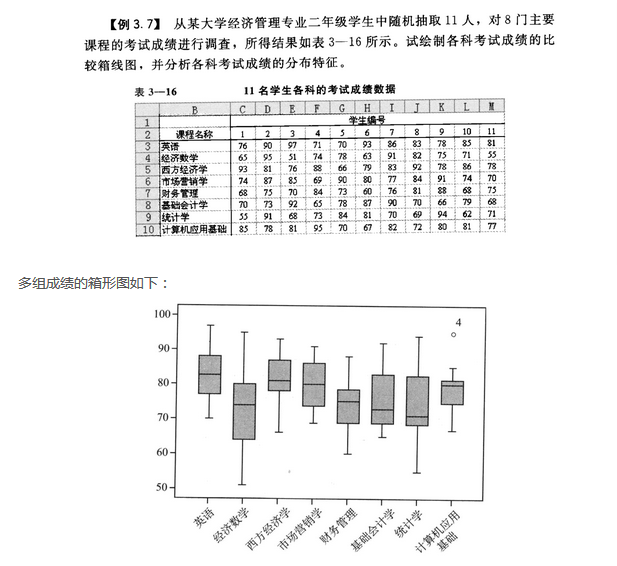


从图中我们可以直观地看出，箱形图的中位数和上四分位数的间距比较窄的话，对应曲线图，这个间距内的数据比较集中，还有就是箱形图的上（下）边缘比较长的话，对应曲线图，尾巴就比较长。

1. 箱型图有个功能就是可以检测这组数据是否存在异常值。异常值在哪里呢？就是在上边缘和下边缘的范围之外

如图上方就可以看出是异常值

1. 可以直观地比较多组数据的情况



1) 各科成绩中，英语和西方经济学的平均成绩比较高，而统计学和基础会计学的平均成绩比较低。（用中位数来衡量整体情况比较稳定）

2) 英语、市场营销学、西方经济学、计算机应用基础和财务管理成绩分布比较集中，因为箱子比较短。而经济数学、基础会计学和统计学成绩比较分散，我们可以对照考试成绩数据看看也可以证实。

3) 从各个箱形图的中位数和上下四位数的间距也可以看出，英语和市场营销学的成绩分布是非常的对称，而统计学呢？非常的不平衡，大部分数据都分布在70到85(中位数到上四分位数)分以上。同样，也可以从成绩单里的数据证实

4) 在计算机应用基础对应的箱形图出现了个异常点，我们回去看看成绩单，计算机那一栏，出现了个计算机大牛，考了95分，比第二名多了10分。而其他同学的成绩整体在80分左右。

5) 其实我们也可以从中得知，用平均值去衡量整体的情况有时很不合理，用中位数比较稳定，因为中位数不太会收到极值的影响，而平均值则受极值的影响很大。