单元 10：基本统计学的量度及其释义

10.1 集中趋势的量度：平均数、加权平均数、中位数和众数

本部分基本上是温习从前学过的课题。教师应与学生重温各种数据的图示法，以及各种集中趋势的量度方法。为要引起学生兴趣，教师应多给与一些实际生活例子。

日常生活例子：(摘录自《生活与统计》，由政府统计处制作)

(i) 一班学生的平均高度是 1.73 米。
(ii) 1994 年每天的平均最高气温是 25.9 °C。
(iii) 1994 年第三季家庭收入的中位数是 $15,000。
(iv) 1993 年男性结婚年龄的中位数是 30 岁 (只计算首次结婚的年岁)
(v) 1991 年人口普查的资料显示，家庭人数的泉数值是 4 人。
(vi) 货品生产商需知道每件产品的众数值。如恤衫制造商会制造较多是数尺码的恤衫。

以下是另外一些例子：

<u>例一</u>
试利用每天在报章上刊登的数据 (如：股价、世界各城市气温等)，比较其集中趋势 (平均数、中位数和众数)。

<u>例二</u>
记录每位学生每星期所得的零用钱数目，数额以 $10 为单位。试找出零用钱的平均数、中位数和众数。哪一个数据较能代表学生每月所得的零用钱？学生可从这个例子察觉数据有时会有多于一个众数。

<u>例三</u>
一公司员工的每年收入如下：

董事	$ 2,400,000	经理	$ 2,000,000
营业员	$ 300,000	总服务员	$ 150,000
3 名服务员	每位 $ 100,000	2 名秘书	每位 $ 90,000
1 名见习员	$ 80,000		

求这 10 名员工薪金的众数、中位数及平均数。哪个数值最能代表员工的平均薪金？试解释答案。

加权平均数

日常生活常会用到加权平均数的概念。分组数据的平均值实际上是将各组中值乘以各组频数后得出的平均数。考试分数的计算亦运用了加权平均数的概念。

加权平均数广泛应用在各种指数的计算上，如物价指数等。

指数

指数是一个用来表示某些数据相对于某时期的变化和走势的数字。作为比较的基准时期叫做基期，其指数通常定为 100。指数是以基期的百分数表示。

香港有各种各样的指数，例如：

(i) 甲类消费物价指数
(ii) 类消费物价指数
(iii) 恒生物价指数

各采用不同的加权数值去反映不同家庭的不同消费模式。

例一

甲类消费物价指数覆盖了全港市区家庭的 50%，这些家庭在 1989 年 9 月至 1990 年 9 月期间的每月支出是在 $ 2,500 至 $ 9,999 之间。为了显示每一时期各种食物及服务的一般水准，我们对每个时期用一个百分数(即指数) 表示，以观察其总体上的变化。

例二

恒生指数在 1964 年 7 月 31 日为 100，在 1997 年 9 月 29 日，已上升至 14,864，显示总市值有超过 100 倍的增长。

物价指数

商品	基期价格	时期 t 的价格	权数
1	c_{01}	c_{t1}	w_1
2	c_{02}	c_{t2}	w_2
⋮	⋮	⋮	⋮
k	c_{0k}	c_{tk}	w_k

t 期物价加权平均指数

$$= \frac{\text{基期价格的加权平均数}}{\text{基期价格的加权平均数}} \times 100$$

$$= \frac{\dfrac{w_1c_{t1} + w_2c_{t2} + w_3c_{t3} + \cdots + w_kc_{tk}}{w_1 + w_2 + w_3 + \cdots + w_k}}{\dfrac{w_1c_{01} + w_2c_{02} + w_3c_{01}3 + \cdots + w_kc_{0k}}{w_1 + w_2 + w_3 + \cdots + w_k}} \times 100$$

$$=\frac{w_1c_{t1}+w_2c_{t2}+w_3c_{t3}+\cdots+w_kc_{tk}}{w_1c_{01}+w_2c_{02}+w_3c_{03}+\cdots+w_kc_{0k}}\times 100$$

<u>例三</u>

假设以 1992 年为基期，下表列出了 1992 年和 1995 年几种主要食物及零食的价格，求 1995 年的物价指数。

商品	1992年的价格	1995年的价格	权数
米	28	30	5
猪肉	21	26	4
鱼	30	36	4
蔬菜	6	8	3
零食	4	6	2

物价指数

$$=\frac{5\times 30+4\times 26+4\times 36+3\times 8+2\times 6}{5\times 28+4\times 21+4\times 30+3\times 6+2\times 4}\times 100$$

= 117.3

建议指数研究课题：

(i) 个人的消闲物品消费指数。
(ii) 学生交通消费指数。

集中趋势量度的应用

我们用以下例子说明：

陈先生以更换破烂窗子为业。根据过往数月的工作记录，他的女儿替他计算出更换一个窗子所需时间的结果如下：

平均数 = 2.0 小时
中位数 = 1. 8 小时
众数 = 1. 5 小时

这些数字对陈先生有甚么作用呢 ?

(i) 台风过后，一所学校的 100 只窗给击碎了。陈先生大约需要 $100 \times 2.0 = 200$ 小时去更换那些窗子。这情况所用的时间采用平均数来计算。

(ii) 若问陈先生一般需要多少时间去更换一个窗子，他通常会利用众数作为答案。

(iii) 陈先生日常更换窗户是以件工计算。现在公司要求他在 2 小时内更换一个窗子，陈先生应否接纳这份工作 ? 为甚么 ?

教师可利用其它例子带出平均数、中位数及众数的优点及缺点。

10.2 离差的量度

教师可提出两组平均数及中位数相同但分布状态不一样的数据，让学生看出，单以集中趋势的量度去描述数据很明显并不足够。如何量度散布在平均数四周的数据呢？这种量度的方法称为离差的量度。

离差量度的种类

以下是常用的离差量度：

(i) 分布域

(ii) 四分位数间距或半四分位数间距

(iii) 方差和标准差

以上各项的定义可用生活化实例及图像说明。跟很多数学概念一样，用图像显示统计数据比用符号或数字较易令学生明白和掌握。

(i) 分布域

日常生活例子：天气报告；气温相差；股市波动；潮水涨退等。

两组有相同的分布域但分布不同的数据，可以用图像来显示。

(ii) 四分位数间距或半四分位数间距

日常生活例子：薪金；考生入学试的成绩等。

我们可根据数据的性质，用插值法来估计四分位数之值。

注：除可运用四分位数，也可采用百分位数来计算。

(iii) 方差及标准差

以上提及的分布域、四分位数间距、半四分位数间距，并没有尽用数据内的所有数值。当某些数据的数值改变时，离差值并没有因应改变。用来描述所有数据分布状态的一个重要计算方法是方差，或称为标准差。

例一

找出不同年龄的男孩及女孩高度的标准差，并作出适当的结论。

例二

以下是两位学生做实验时所量度得到的两组数据：

	酸的体积 (cm^3)									
学生 A	8	12	7	9	3	10	12	11	12	14
学生 B	7	6	7	15	12	11	9	9	13	11

(i) 求每一学生所量度酸的体积之平均值。

(ii) 求酸的体积之标准差。

(iii) 哪组数据较可靠？

比较不同离差量度

教师应详细说明每种离差量度的利弊。

10.3 频数分布、累积频数分布及其图像额示和释义

教师可用生活实例去复习这课题。

分组数据和不分组数据的异同，宜作重点说明。每种统计图均有其特色，教师可透过各种不同的图像作示范，从而说明某类图像是否适用于表达某些资料与结论。重点应放在理解图表和图表在预测方面的功用。

可利用计算机软件将已知数据绘画成不同图像。当学生已经掌握用各种图像表达数据的方式后，可尝试利用计算机分析各种形式的异同，以及优劣的地方。教师可运用计算机作为教学工具，使学生便于明白。

以下是一些课题建议，作为学生研究的项目：

1. 题目：人体体温
   问题：是否所有人皆有相同体温？男性与女性的体温是否相约？是否女性的体温较男性高？女性的体温波幅是否较大？
   数据搜集：用人体温度计量度 50 – 100 人的体温。大部份温度计以摄氏 (°C) 单位表示，量度温度须准确至 0.1 °C。

2. 题目：路口交通情况
   目的：研究某繁忙路口于一分钟内汽车通过的数目。
   数据搜集：选择一交通流量大的路口。决定以何种车辆作为统计数据及统计的地点所在。数算该类车辆于一分钟内在同一方向通过该路口的数目。统计时间为一小时。若时间容许，可选择一交通流量受阻的地方，例如在红绿灯附近，或十字路口。重复以上步骤并作记录。(若作为学生的研究项目，可选择不同地点给不同组别的学生。)

3. 题目：中六生的零用钱数目
   问题：中六男生或女生每星期所得的零用钱是否相约？
   数据搜集：问 50 位中六男生及女生每星期的零用钱数目。若他们没有零用钱，记录那些答案为零。分别记录女生和男生的零用钱的数目。

4. 设计及搜集关于一些课题的意见 (例如：流行歌曲、政治团体、教育学历等。)

5. 究竟在某一快餐店购买“中薯条”，会有多少条薯条呢？学生搜集这些数据时，会发觉需要处理碎薯条的情况。很快地，班内会有不同的处理方法。我们可以怎样描述那些答案呢？你发现甚么情况呢？有经验的学生可能会讨论薯条之长度或形状分布。薯条的形状大小，要怎样才不会被人投诉？最好购买那一类包装的薯条：大、中或小？这快餐店与其它店铺的比较又如何呢？

10.4 箱形图

箱形图 (亦称框线图) 可用作展示一组数据的特征，包括最小值、下四分位数、中位数、上四分位数、最大值。

在水平图尺上绘制箱形圆的步骤如下：

步骤 1. 在图尺上绘画一个窄长的方箱，由下四分位数伸延至上四分位数；
步骤 2. 在方箱内中位数的位置上画一垂直线段；
步骤 3. 从方箱的左边画一线段到最小值；
步骤 4. 从方箱的右边画一线段到最大值。

步骤 3 与步骤 4 所画的两线段称为框线。

图 10.4(i) 显示一个箱形图的大致形状。

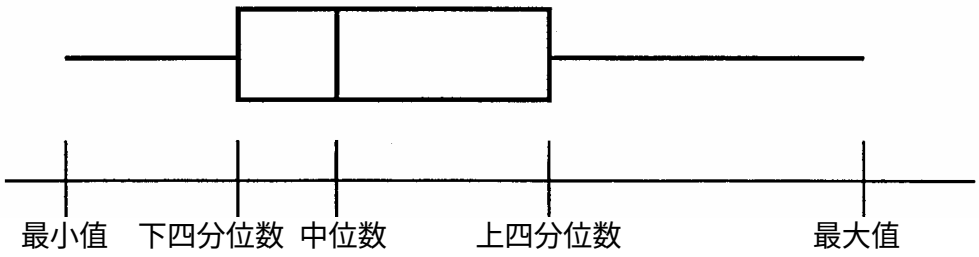


图 10.4 (i)

用者可随意选定箱形图的方箱阔度，丝毫不会影响该图所表达的意思。

下面例子说明箱形图的制作。

<u>例一</u>
以下是 20 位足球员每星期用在购买汽水的支出：

20 位足球员用在
购买汽水的支出 (元)

26	34	31	36	16
27	32	21	43	41
30	6	15	38	27
58	35	28	21	20

从这些数字得知
最大值 = 58
上四分位数 = 35.5
中位数 = 29
下四分位数 = 21
最小值 = 6

绘成箱形图如下 (图 10.4(ii))：

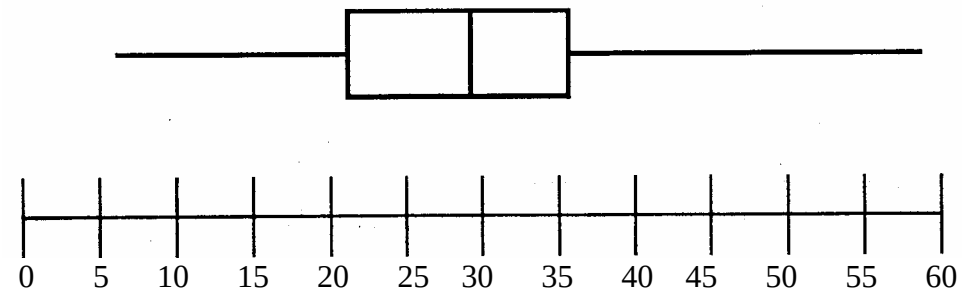


图 10.4(ii) 购买汽水费用箱形图

把上述制图的步骤作出相应的调整后，箱形图也可以在垂直图尺上绘制 (参照图 10.4(i))

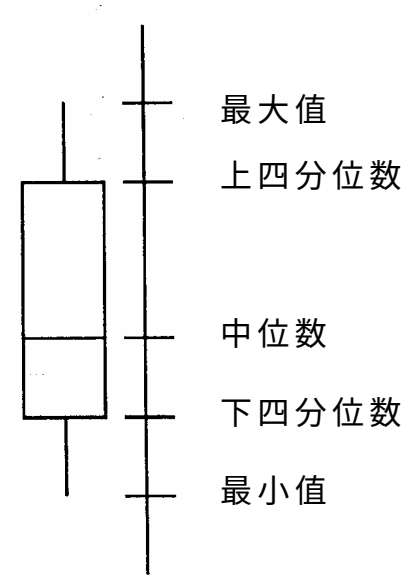


图 10.4(iii)

箱形图可以表示一组数据的中心位置、散布情况及对称性。

1. 中位数上的线段表示数据的中心位置。
2. 方箱的长度显示中间一半数据的散布情况。
3. 两线段的长度显示数据的第一个及第四个四分之一部分的散布情况。
4. 从图的形状即可看出数据分布的对称程度。
   如果图是对称的，有关数据很可能是对称的。偏离对称的图可反映出有关的数据也有类似的偏离 (见图 10.4(iv) 及 10.4(v))。

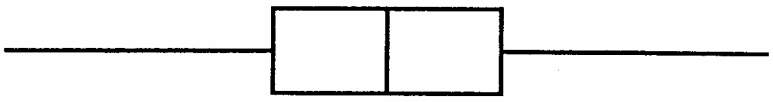


图 10.4(iv) 对称于中位数的数据分布

图 10.4(v) 非对称于中位数的数据分布

箱形图用来比较多组数据，也很有效。可用作比较不同数据组的中心位置、散布情况及对称性。

图 10.4(vi) 的箱形图展示三个城市的工资分布。

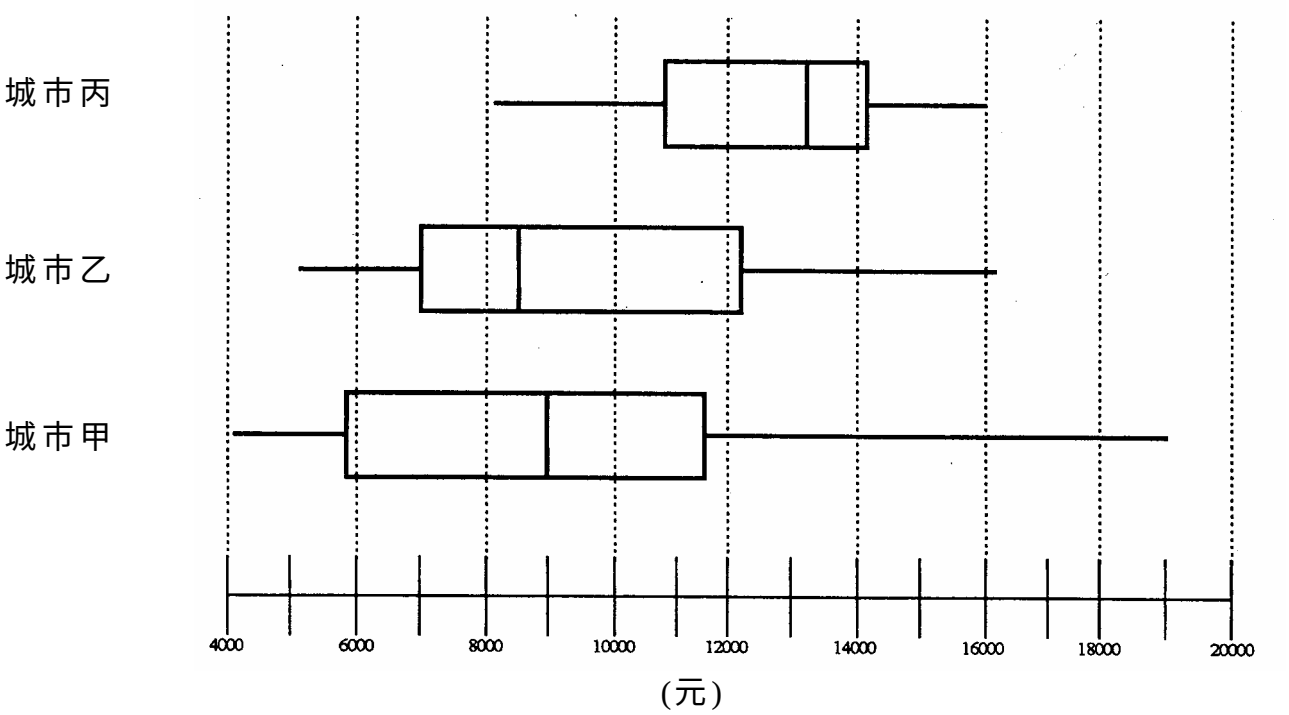


图 10.4(vi) 三个城市的工资分布

由图 10.4(vi) 可知，城市丙的工资中位数最高，城市甲的工资离差最大，城市乙近半数在职人士的工资低于城市丙的最低工资。

10.5 干叶图

干叶图是一种结合图与表以表达数据的方法，以数据来替代直方图中的直方形，基本上是一已作改良的分组频数分布。我们采用数据的最后单位(称为「叶」) 来代替画记，用来显示观察所得的数值。这些观察值是根据「叶」之前的数字 (称为「干」) 作排序，由小至大依次排列在叶的左方。

以下是利用 40 位学生的数学测验成绩，来显示干叶图的制作：

49	73	58	59	40	61	45	61
51	43	80	55	81	52	54	45
58	53	55	38	59	72	46	68
42	65	65	64	29	48	39	67
47	56	66	48	42	56	51	53

步骤 1. 画一垂直线。

步骤 2. 选择数据的前置数字为干。

步骤 3. 将干从小至大排列，并列于该垂直线的左方。

步骤 4. 检视这些数据，将最后的数字 (叶) 由小至大排列在对应干的右方。

学生应能掌握描绘及解释干叶图的方法。

干 (十位值)	叶 (个位值)
2	9
3	89
4	02235567889
5	1123345566889
	9
6	11455678
7	23
8	01

图 10.5 (i)

例一

干 (百位值)	叶 (十位值)
1	11 23 47
	48
2	23 27
3	55 63 80
4	21 23
5	00

图 10.5(ii)

在上图，干是百位数及叶是十位以下数字。

干叶图所显示的数据为：

111, 123, 147, 148, 223, 227, 355, 363, 380, 421, 423, 500

例二

以下为某家庭于冬天的九个星期里每星期的耗电量 (单位为 KWh)：
338，354，341，353，351，341，353，346，341。

根据数据，耗电量的百位数字每个都是 3。再观察数据的十位值，数值为 3、4 或 5 。所以，我们选取数字 33、34 及 35 为干，放在垂直线的左方，个位值为叶，放在垂直线的右方。最后把数据依递增次序排列，制成干叶图，如图 10.5(iii) 所示。

干 (十位值)	叶 (个位值)
33	8
34	1 1 1 6
35	1 3 3 4

图 10.5 (iii)

例三

11 位学生的中、英文科考试成绩如下：

英文科： 23， 39， 40， 45， 51， 55， 61， 64， 65， 72， 78
中文科： 37， 41， 44， 48， 58， 61， 63， 69， 75， 83， 89

可绘制并列干叶图来比较两科成绩的表现。

英文科

干 (十位值)	叶 (个位值)
2	3
3	9
4	0 5
5	1 5
6	1 4 5
7	2 8
8	

中文科

干 (十位值)	叶 (个位值)
2	
3	7
4	1 4 8
5	8
6	1 3 9
7	5
8	3 9

图 10.5(iv)

利用背靠背的干叶图可以更清楚反映两者的比较。

英文科 叶 (个位值)	干 (十位值)	中文科 叶 (个位值)
3	2	
9	3	7
5 0	4	1 4 8
5 1	5	8
5 4 1	6	1 3 9
8 2	7	5
	8	3 9

图 10.5(v)

对于某些数值特别小或特别大的数据，为要令到干叶围看来较美观，可以另外写出这些数据。用以下的数据为例子，如图 l 0.5(vi) 或图 10.5(vii) 所示。

12.0	21.2	23.1	24.8
22.0	25.9	29.8	31.9
21.2	21.6	22.8	22.5
26.3	25.1	21.7	21.6
23.5	27.0	38.5	28.5

干 (个位值)	叶 (十分位值)
LO = 12.0	
21	2 2 6 6 7
22	0 5 8
23	1 5
24	8
25	1 9
26	3
27	0
28	5
29	8
HI = 31.9, 38.5	

图 10.5(vi)

干 (个位值)	叶 (十分位值)
LO = 12.0	
21	2 2 6 6 7
22	0 5 8
23	1 5
24	8
25	1 9
26	3
27	0
28	5
29	8
30	
31	9
HI = 38.5	

图 10.5(vii)

10.6 总体参数和样本统计量的初步认识

本节介绍**总体**、**总体参数**，**样本**及**样本统计量**的概念。从总体抽取一些样本，这些样本可以适当地代表总体，从而取得估计量的总体参数及作出推论。可用例子介绍总体、样本、样本方差及样本平均数这几个名词。

总体及样本

例一

年龄分布：　　某中学学生的年龄分布。
总体：　　该校的所有学生。
样本：　　随机在该校抽取 100 名学生。

例二

电撞见观众：进行一项调查于下午六时至晚上十时最多人观看的电视节目。
总体：　　下午六时至晚上十时期间，所有电视观众。
样本：　　从总体电视观众中随机抽取 800 人作为访问对象。

例三

晶质控制：工厂所生产的电灯泡寿命 o
总体：　　全部电灯泡产品。
样本：　　随机抽取 30 个电灯泡加以测试。

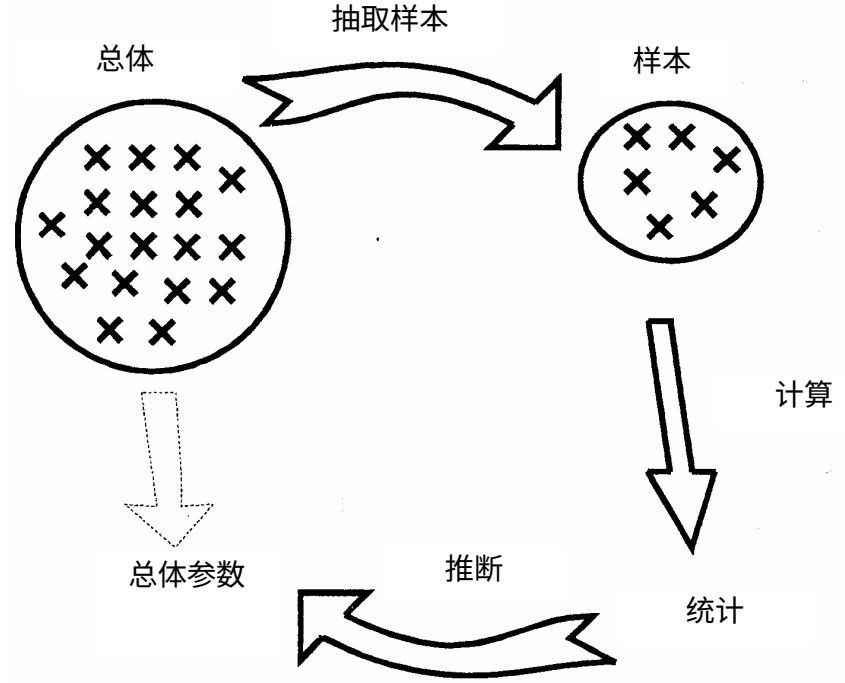
例四

某生产录像机公司的客户服务部经理有意作一项调查，看看在过去 12 个月内购买录像机的顾客对产品是否满意。他计划根据 1,425 名顾客所交回的保养证去进行调查。

i)　试描述经理所想要了解的调查总体及样本。
ii)　试描述该经理有兴趣搜集的数据。

总体参数是描述总体的特色，而样本统计量则是由总体抽取样本得出特色的一些计算值。

教师可以利用下图说明各项的关系：



样本平均数及样本方差

学生应知道：

(i) 样本平均数 $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}f_i x_i$， 其中 $n=\sum_{i=1}^{k}f_i$

(ii) 样本方差 $s^2=\frac{1}{n-1}\sum_{i=1}^{k}f_i\left(x_i-\bar{x}\right)^2$，其中 $n=\sum_{i=1}^{k}f_i$

(iii) 当样本数目增至无限大时，样本平均数 $\bar{x}$ 及样本方差 s^2 分别趋向总体平均数 μ 及总体方差 σ^2。

(iv) 对于有限的总体来说，总体方差

$\sigma^2=\frac{1}{N}\sum_{i=1}^{k}f_i\left(x_i-\mu\right)^2$，其中 N 是总体数目。

学生并不需要明白**样本平均数的平均值**及**样本平均数的方差**的意义。

<u>例一</u>

一硬币的两面各刻上数字 3 及 4 。这硬币是有偏差的，平均来说，每 5 次投掷中有 4 次出现数字 3 及 1 次出现数字 4 。因此可假设总体只有这几个数：3，3，3，3，4。

总体平均数 $\mu=\frac{3+3+3+3+4}{5}=3.2$

总体方差 $\sigma^2=\frac{4\times(3-3.2)^2+(4-3.2)^2}{5}=0.16$

若学生不知道这硬币有偏差，他们投掷这硬币 3 次，所有可能出现的样本如下：(3, 3, 3), (3, 3, 4), (3, 4, 3), (4, 3, 3), (3, 4, 4), (4, 3, 4), (4, 4, 3), (4, 4, 4).

以样本 (3, 3, 3) 为例：

样本平均数 $\bar{x}=\frac{3+3+3}{3}=3$

样本方差 $s^2=\frac{(3-3)^2+(3-3)^2+(3-3)^2}{3-1}=0$

以另一样本 (3, 4, 4) 为例：

样本平均数 $\bar{x}=\frac{3+4+4}{3}=3\frac{2}{3}$

样本方差 $s^2=\frac{(3-3\frac{2}{3})^2+(4-3\frac{2}{3})^2+(4-3\frac{2}{3})^2}{3-1}=\frac{1}{3}$

教师凭此可指出：样本平均数和样本方差与总体平均数和总体方差并不一定相同。