符号与注记

为了公式恐惧者：代数的符号与公式是表达整族个别数值量的方便形式。数学是这些形式的语言。形式，作为符号和符号组，是作为等式和不等式的句子中的名词、主语和宾语。推导可变革、发展和揭示形式间的联系。把族的形式看作它们自身的实体，并仔细地识别它们；按照它们的内容给予它们以自己的名字。于是句子能被读懂，而且它们的真相能被认识。一些通常的统计学形式如下：

希腊字母(发音)：α(阿尔发)；β(培塔)；δ(代尔塔)；ε(依伯西隆) χ(盖)；μ(弥伏)；π(派爱)；ρ(罗)；Σ(西格玛)；σ(西格玛)；ζ(兹依塔)。

有用的惯例：(i)罗马字母用于随机变量（如估计量），希腊字母用于固定的总体参数。例： x和x估计参数 μ_x；r是参数 ρ 的估计；s^2 估计参数 σ^2。(ii)一个音调符号‘ $\wedge$ ’或‘ 带帽 ’表示被带帽的参数的估计； μ_x 的一个估计 x 是一个 μ_x，以及 s_x^2 是一个 σ_x^2。

记号：$|x|=x$ 的绝对值，(6.2)；$\sum_{i=1}^{n} x_i=(x_1+x_2+\cdots+x_n)$ (2.1)；$\sum' xx=\sum_1^n(x_i-\bar{x})^2$，(2.5)；$\sum' xy=\sum_1^n(x_i-x)(y_i-y)$，(48.1)；$\leqslant$，“不大于”，$\mu[y]=\mu_y=E(y)=$ 随机变量 y 的总体均值，不管 y是什么，(§5)。$\mathrm{Var}(y)=V(y)=\sigma_y{}^2$ 是总体中一随机成员 y 的总体方差，(5.1)。因此，关于 $y=(x-\mu_x)^2$，其总体方差的另一个符号是 $\mu[(x-\mu_x)^2]=\mathrm{Var}(x)=V(x)=\sigma_x{}^2$。

“ $\sim$ ’和‘ ，…，’分别被读成‘ 按什么型分布 ’和‘ 继续下去直到 ”。因此，在 §13 中 $x_i\sim NI(\mu_x,\sigma_x{}^2)$ ， $i=1,2,\cdots,n$，被读成“ 随机变量 x_1,x_2，继续下去直到 x_n都是服从正态型分布，并且都相互独立。’它们的正态分布的总体均值为 μ_x 总体方差为 $\sigma_x{}^2$。

$P[A]=p$ 意思是"[ ]中的陈述 A 为真的概率等于正数 p，其中 $0\leqslant p\leqslant 1$"陈述了有关随机变量（不是有关常量）的概率。因此 $P[|t_{10}|>2.228]=0.05$ 是一个随机地选定的自由度为 10 的 t- 变量的绝对值超过 2.228 的概率；参阅 §4 和(12.3)，随机变量名称的整数下标表示自由度数。然后，下标可能小于 1，例如 v_α，$0<\alpha<1$，表示 $P[v>v_\alpha]=\alpha$ 时的随机变量 v 的值，因而下标 α 是点 v_α 的右侧 v- 分布的面积。所以对于 §15 中 χ_{10}-变量，$\chi^2_{0.025}=20.843$ 意味着 P[自由度为 10的 χ^2 随机变量 $>20.843]=0.025$ 。当自由度和右侧面积两者都需要表出时，诸如 $\chi^2(10;0.025)=20.843$ 这样的符号是有用的。

垂直一竖"|"表示条件性；也即一般情况的某种限制。因而，在简单线性回归 §47)中 $\sigma^2_{y|x}$ 表示在任一设想为固定的 x- 值处概念上可观察到的 y-值的总体方差。一竖有助于 §127 中多元回归的符号，那里譬如 β_1 是当 $x_2,\cdots,x_p$ 为常数时 y 在 x_1 上回归系数 $\beta_{y1|2\cdots p}$ 的缩写。

R- 符号：例如 $R(\mu,\beta_1,\beta_2)$ 表示 $P=2$ 的多元回归模型(127.6)中"属于"三个参数 μ、β_1 和 β_2、自由度为 3 的平方和。$R(\beta_3,\beta_4|\mu,\beta_1,\beta_2)$ 表示在已经包含在"|"后面的参数 μ μ_1 和 μ_2 的模型中附加"属于"|"前面包含参数 μ_3 和 β_4 另外的平方和。

参照分布

随机变量	分析用于	定义所在章节表	
Z	均值，当 δ^2已知	§11	*A2*
$t(f)$	均值，当 δ^2未知	§12	*A3*
$\chi^2(f)$	方差和离散数据	§14	*A4*
$F(f_1,f_2)$	方差比	§26	*A5*
二项	特殊情况	§31	—
普哇松	特殊情况	§44	—
$q(a,f)$	多元极差检验	§100	*A6*

几位小数？：罗马皇帝 Agrippa（公元 63－12 年)做了一个很早的区间估计，不列颠的十分惊人地不规则和随机的周界长的

区间估计。关于此 Lloyd(1966)报告如下："Agrippa 早先极其重视计算它的周线，在以精密的精度装饰出一条十分宽广的边界，公开宣告它不大于 28.104 司达堤 (Stadia)，不小于 20.526 司达堤 (Stadia)——一个司达堤 (Stadium)大约为 202 码。"袖珍电动计算器给出很多位小数，因此在中间的计算中没有必要作四舍五入的进位。然而作为常识，建议最后提出的估计和区间四舍五入到二位有效数字，即两个不同于 0 的数码[67]，在一般实践中是方便的——这与 Ehrenburg(1981) 所提倡的一样。

基本概念和方法

§1．定义

在相同状况下取得的重复观察值并展示或常发生变异称为一个统计随机变量的特殊值、测定值或实现的样本。

连续随机变量是一种概念上可以取其自然值域中任何一个值的随机变量。例如，电池的寿命，血压以及胆固醇浓度。

离散随机变量是一种其可能值被有限区间分隔开的随机变量。例如，心率，家庭大小以及辐射计数。

样本中的个别随机变量值称为从概念上的一个统计总体中抽取的特殊元。

统计总体是在特定的试验状况下，可以想像为单个观察所能取得的一切可能的随机变量值的全体，不论这些值是同还是不同。在这一概念中，重复的、恒等的随机变量值都被当成是总体中的不同的成员。

样本只包含从上述全体或总体中，真正被取为观察值的那些值的集合。

随机样本　是按照相互独立原则抽取的观察值，抽取时要确保总体的每一个元都有相等的机会被选中，即在样本中出现。

独立性：观察值或随机变量值是两两独立的，如果任一个值不能与另一个值有关，受其限制，或由它来预测。

在任何调查研究中，重要的是明确一个适宜的总体，并保证观察值只是这一总体的随机独立元。

在科学上引起兴趣的总体特征的标志，通称为参数。其中重要的有：

（a）位置参数，它们是依据度量原点到总体某一规定特征的

距离来给总体确定位置。例如：

（ i ）总体均值：所有随机变量值的平均值；

（ii）总体中位数：所有随机变量值的一半大于，一半小于中位数；

（iii）总体众数：最经常出现的随机变量值的大小。

（b）提供一个观察值到另一个观察值的变异性的数值度量或标志的参数。例如：

（iv）总体方差：随机变量值距离其总体均值的平均平方偏差；

（ v ）总体标准差：总体方差的平方根。

总体可能在它们的位置和变异性方面有差别，这些差别将由有关参数的不同数值表示出来。在搜索总体的信息中，我们总是搜索有关其参数的信息。这些参数仅在总体的每一个元都知道下才能完全确定。这种知识是很难取得的。因而，有必要利用从随机样本计算出的估计值对总体参数作出推断。

计算点估计是为了得出未知参数的最优单值。

区间估计(例如,正文第 3 页中的Agrippa)是一个未知参数的估计会落在其中的一个数值的范围。

置信区间是以指定信度断言参数的值将落在其中的区间估计。

统计方法利用相对小的样本以

（ i）获得总体参数的点估计和置信区间估计；

（ii）估计样本的计算结果与未知总体或其参数“真值”的假定值在某种程度上相符的概率，亦即统计一致的概率；

（iii 推断有关总体参数的差和比。

§2. 样本均值和方差：变异系数

样本均值：若 $x_1, x_2, \cdots x_n$ 是几个观察的样本中的值，样本均值为统计量① x 的大小，定义为

①统计量是样本观察值 $\cdots x_1, x_2, x_n$ 的函数,它不含总体的未知参数。——译注

$$\bar{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\frac{\sum_{i=1}^{n}x_i}{n} \tag{2.1}$$

其中和式 $\sum_{i=1}^{n}x_i=(x_1+x_2+\cdots+x_n)$，读成“希格玛 x_i 从 $i=1$ 到 $i=n$”。关于样本均值的偏差的和 $\sum_{i=1}^{n}(x_i-\bar{x})$，等于零，因为

$$\sum_{1}^{n}(x_i-\bar{x})=\sum_{1}^{n}x_i-\sum_{1}^{n}\bar{x}=\sum_{1}^{n}x_i-n\bar{x}$$

$$=(x_1+x_2+\cdots+x_n)-n\bar{x}=0,\text{利用}(2.1).$$

若 n_j 表示样本中具有相同值 x_j 的观察值的个数，则观察值的和是

$$\sum_{1}^{n}x_i=\sum_{j}n_jx_j。$$

这时，右边乘积的和是对 x 的不同的值，记为 x 的序列 x_j。由(2.1)，我们有

$$\bar{x}=\frac{\sum_j n_jx_j}{n}=\sum x_j\left(\frac{n_j}{n}\right)。 \tag{2.2}$$

分式 n_j/n 是观察值 x_j 在样本中的相对频率，记 $f(x_j)=n_j/n$，(2.2)变为

$$\bar{x}=\sum x_jf(x_j)。 \tag{2.3}$$

另外，由于 $\sum n_j=n$，故

$$\sum f(x_j)=\sum\left(\frac{n_j}{n}\right)=\frac{\sum n_j}{n}=1,$$

即全体相对频率的和为 1。

例 设样本观察值(随机变量的值)为

$$x_1=3.2,\ x_2=2.9,\ x_3=4.6,\ x_4=2.8,$$

$$x_5=2.9,\ x_6=3.2,\ x_7=3.2,\ x_8=2.8。$$

由(2.1)，在 $n=8$ 下，

$$\bar{x}=\frac{3.2+2.9+\cdots+2.8}{8}=3.2。$$

对 $i=1,2,\cdots,8$ 计算偏差 $(x_i-\bar{x})$：

x_i	3.2	2.9	4.6	2.8	2.9	3.2	3.2	2.8
$\bar{x}$	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2
$(x_i-\bar{x})$	0.0	−0.3	1.4	−0.4	−0.3	0.0	0.0	−0.4

故

$$\sum_{i=1}^{8}(x_i-\bar{x})=0.0-0.3+1.4+\cdots-0.4=0.$$

现将样本改写成

$$2.8,\ 2.8,\ 2.9,\ 2.9,\ 3.2,\ 3.2,\ 3.2,\ 4.6,$$

则在(2.2)中取 x_j 序列为2.8，2.9，3.2和4.6，

$$\bar{x}=\frac{2(2.8)+2(2.9)+3(3.2)+1(4.6)}{8}$$

$$=2.8\left(\frac{2}{8}\right)+2.9\left(\frac{2}{8}\right)+3.2\left(\frac{3}{8}\right)+4.6\left(\frac{1}{8}\right)=3.2。$$

这里 $f(x_j)=n_j/n$ 的逐次取值为 2/8，2/8，3/8 和 1/8。同样，由于

$$\Sigma n_j=2+2+3+1=8=n,$$

证实了

$$\Sigma f(x_j)=\Sigma\left(\frac{n_j}{n}\right)=\frac{2}{8}+\frac{2}{8}+\frac{3}{8}+\frac{1}{8}=1。$$

样本方差：这个描述性统计量规定了在观察值的样本中变异性的数值度量。若样本元为 $x_1,x_2,\cdots,x_n$，则样本方差是由

$$s^2=\frac{1}{n-1}\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\} \qquad (2.4)$$

定义的量或统计量，其中的‘…’读作‘等等直到’。将距离均值的偏差平方和缩写成

$$\sum_{1}^{n}(x_i-\bar{x})^2=(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2=\Sigma' xx \qquad (2.5)$$

是方便的，故(2.4)化为

$$s^2=\frac{\Sigma' xx}{n-1}。 \qquad (2.6)$$

就象稍后(§5)指出的那样，$\Sigma' xx$ 是除以$(n-1)$，不是 n，因为 s^2 扮演的角色是总体方差的估计量。

样本方差的计算．由于

$$(x-\bar{x})^2=x(x-\bar{x})-\bar{x}(x-\bar{x}),$$

故

$$\Sigma' xx=\Sigma x(x-\bar{x})-\Sigma\bar{x}(x-\bar{x})。$$

另外，由于

$$\sum \bar{x}(x-\bar{x}) = \bar{x}\sum(x-\bar{x}) = 0,$$

我们求得

$$\sum{}'xx = \sum x(x-\bar{x}) = \sum x^2 - \bar{x}\sum x = \sum x^2 - \frac{(\sum x)^2}{n}. \qquad (2.7)$$

利用 (2.1) 可得量 $(\sum x)^2/n = x\sum x = nx^2$，并通称它为“校正项”，因此 $\sum{}'xx$ 通称为校正（对于均值）平方和。(2.7)和 (2.6) 合起来则得 s^2 的便于计算的形式为

$$s^2 = \frac{1}{n-1}\left\{\sum x^2 - \frac{(\sum x)^2}{n}\right\} = \frac{\sum x^2 - \bar{x}\sum x}{n-1}。 \qquad (2.8)$$

例 利用前面例子中的数据，这时 $n=8$，(2.4)直接给出

$$s^2 = \frac{1}{7}\{0.0^2 + (-0.3)^2 + 1.4^2 + \cdots + (-0.4)^2\} = \frac{2.46}{7}$$

$$= 0.35$$

或由(2.8)

$$s^2 = \frac{1}{7}\left\{(3.2^2 + 2.9^2 + \cdots + 2.8^2) - \frac{(3.2+2.9+\cdots+2.8)^2}{8}\right\}$$

$$= 0.35,$$

再由 (2.8)，由于 $\bar{x} = 3.2$，$\sum x = 25.6$

$$s^2 = \frac{1}{7}\{(3.2^2 + 2.9^2 + \cdots + 2.8^2) - (3.2)(25.6)\} = 0.35。$$

样本标准差：这仅为

$$s = \sqrt{s^2}。 \qquad (2.9)$$

样本标准差是具有与观察值本身相同单位的样本变异性的度量。

变异系数：这是变异性的一个无量纲（没有单位）的指数，它是通过将样本标准差表成样本均值的百分比求得。因此，当 $\bar{x}$ 为正时，

$$\text{变异系数} = 100\left(\frac{s}{\bar{x}}\right)\% \qquad (2.10)$$

$$= 18.5\%,\text{对上述例子。}$$

在 s 随 x 增长的范围中，这个系数作为一种描述性统计量是有用的。

极差：观察值样本的极差是最大的和最小的观察值之间的差(在上面的例子中为 $x_3 - x_4 = x_3 - x_8 = 4.6 - 2.8 = 1.8$)。对于容量小① 的样本，它们的极差提供了样本变异性的一个快速度量，虽然它不是最优的($n = 2$ 除外)。

样本和总体相对频率分布

对于连续随机变量的观察值，一个样本相对频率分布是按下面方式得到的一个数组：

(i) 适当地把随机变量的近似的值域分成若干个区间或类，一般是等间隔的；

(ii) 对每一个区间显示出它的相对频率，即落在这个区间里的样本观察值所占比例。

注意分布着的是变量而不是频率。

直方图：样本频率分布的直方图是依次在每一个区间上绘出面积与各自的样本相对频率成比例的矩形图形。当且仅当区间的宽度都相同时，所有矩形的高就方便地与各自的相对频率成比例，如图 3

例：表 3.1 给出了 200 只山羊妊娠期（按天为单位）样本的频数与相对频率分布。例如，妊娠期在 148.5 到 149.5 天的有 28 只

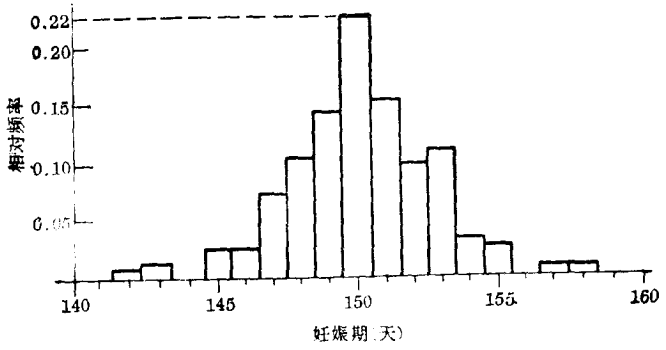

图 3　200 只山羊妊娠期的直方图

①原书“容量相同”系“容量小”之误。——译注

山羊，其相对频率为 $28/200=0.14$。图 3 给出了表 3.1 中数据的直方图。

表 3.1　200 只山羊妊娠期的样本频数和相对频率分布

妊娠期(天)	频数	相对频率
<141.5[a]	0	0
142	1	0.005
143	2	0.010
144	0	0
145	4	0.020
146	4	0.020
147	13	0.065
148	20	0.100
149	28	0.140
150	44	0.220
151	31	0.155
152	18	0.090
153	22	0.110
154	6	0.030
155	5	0.025
156	0	0
157	1	0.005
158	1	0.005
>158.5[a]	0	0
	总数 200	1

a 符号<表示"小于"；>表示"大于"。

相对频率分布与相应的直方图可能令人联想起总体的有趣特征。两个共同的特点是：(i) 聚集倾向——例如，大部分观察值聚集在一个小区间中，例如靠近分布中心的；(ii) 对称性——在距中心相等的区间中有相同比例的观察值。在数学上已证实，当影响观察值的变异性是由许多原因产生的小的成分综合而成时，则上述两个特点是可预期的。在这类情况下，观察值势必具有统计学的理论和应用上都至为重要的模型。它就是正态或高斯分布（§ 14)模型。

"茎与叶"(Stem and leaf) 展示法，其创始人 Tukey (1977)

这样称呼它，提供了展示数据特征的另一方法。表 3.2 给出第一个例子中 $n=8$ 个观察值的简单的茎与叶展示。

表 3.2 茎与叶展示

茎	叶
2	8899
3	222
4	6

例如，该表指出观察值 2.8,2.8,2.9 和 2.9 分在"茎"2 和四张叶为 0.8,0.8,0.9 和0.9上，因此它除了给出每一茎类中观察值的个数之外还给出了它们的实际值。

总体相对频率分布：如前所见，若将 n个观察值分布到各类之中，且 n_i 是第 i 类中的观察值个数，则比值 n_i/n是该类的相对频率。若在每个区间上作一矩形，则第 i 个矩形的面积为 n_i/n，直方图描绘了样本相对频率分布。此外，由于各类中的观察值个数的总和 $\sum n_i$ 等于 n，因此成立 $\sum(n_i/n)=1$；即直方图矩形的总面积等于 1。

对于逐渐增大的样本，我们可以逐次选用更小的区间，这样直方图的矩形就变得越来越细。在极限过程中，当样本变成无穷大，最后增长到与总体相等时，矩形就变得无限细，图形变成一条描绘总体相对频率分布的光滑曲线。这条曲线保存了直方图的以下两条重要性质：

（i）若在 x——轴上的任意两个随机变量值处引两条坐标线与曲线相交，则这两条坐标线、曲线和 x-轴围成的面积代表了这两个随机变量值间的区间上总体的相对频率。

（ii）曲线下的总面积，它对应于随机变量的整个值域，等于 1。

总体频率曲线常常用描述曲线的高度是如何按随机变量值变化的数学表达式来定义。由于相对频率与概率间的联系，总体相对频率曲线称为概率密度函数。

§4. 概率(连续随机变量)

若连续随机变量的一个单值是从某一总体随机地选取，则取值位于任一特定区间中的概率等于该区间的总体相对频率。若 a 和 b 是随机变量 x 的两个指定的可能值，我们记“随机选一单独的 x 大于 a 小于 b 的概率”为 $P[a<x<b]$。这个概率是用介于值 a 和 b 的纵轴之间，总体相对频率曲线之下的面积来度量，正如§10中图10所举例说明的那样。

我们要寻觅总体参数的优良估计，因为这些参数通过它们在定义概率密度函数的数学表达式中的出现，决定了曲线本身，从而也决定了度量所感兴趣的概率的面积。

更多建立在相对频率想法基础上的概率概念，并且适用于离散随机变量的场合，将在§20中讨论。

§5. 总体均值和方差：样本估计量

总体均值：若样本容量 n 小的话，增加另外的样本元将明显地改变样本均值，但是对于大样本，样本均值比较稳定是早已熟知的事实。当样本逐渐增大到构成可能观察值的整个总体时，样本均值的极限等于总体均值。常用的总体均值符号为 μ，μ_x，$\mu[x]$，$\mu(x)$以及 $E(x)$，后者读成“x 的期望值。”

总体方差：总体方差是总体元的变异性的一个常用数值度量。若 x 为随机变量，μ_x 为其总体均值，则 x 距离总体均值的偏差为 $(x-\mu_x)$，它的平方为 $(x-\mu_x)^2$。所有这些平方偏差的均值为总体方差，记为 σ_x^2 或 $V(x)$。因此，

$$V(x)=\sigma_x^2=\mu[(x-\mu_x)^2]=E(x-\mu_x)^2。\qquad(5.1)$$

在高变异性的情况中，当非常大或非常小的偏差占的比例高时，大的 $(x-\mu_x)^2$ 值占的比例也相应就高，从而 $(x-\mu_x)^2$ 值的平均，即 σ_x^2，相应就大。反之，在低变异性的情况中，当大部分 x 值接近其均值 μ_x 时，大部分 $(x-\mu_x)^2$ 的值，从而 σ_x^2，就小。

总体标准差：随机变量 x 的总体标准差，记为 σ_x，只不过是

总体方差的平方根。因此，标准差是变异性的另一可供选择的度量，且具有与 x 随机变量相同单位这一有用的性质。

样本估计量：样本均值、样本方差(2.6)和样本标准差(2.9)是总体相应量的自然点估计。样本的这些计算值模拟了，如果可能，总体参数该是怎么算的过程，此外，在方差的估计 (2.4)中，若 μ_x 为已知，则应以 μ_x 代替 x并用 n 取代 $(n-1)$ 去除 $\sum_1^n(x_i-\mu_x)^2$。

自由度（简记为df）：若 $x_1, x_2, \cdots, x_n$ 相互独立，则距离总体均值的 n 个偏差，$(x_1-\mu_x),(x_2-\mu_x),\cdots,(x_n-\mu_x)$也相互独立。但是，距离样本均值的 n 个偏差却不是相互独立的，因为 x 是一个变量且在各个偏差中有关联地出现，因为 $\sum_1^n(x-x)=0$ 表明任一偏差可由 0-(其余 $n-1$ 个偏差的和)求得。可以证明，距离 $\bar{x}$ 的偏差平方和，$\sum' xx$，在平均意义下等于距离 μ_x 的$(n-1)$个(不是 n 个）偏差平方和，独立偏差的这个等价的个数通称为自由度个数。

亦可证明 $\sum_1^n{}' xx$ 小于 $\sum_1^n(x-\mu_x)^2$，另一方面，除以自由度的个数，而不只是 n，平均地说，差别可得到补偿，并可保证 s^2 在严格意义下为 σ^2 的无偏估计，即

$$\mu\left[\frac{1}{n-1}\sum(x_i-\bar{x})^2\right]=\mu\left[\frac{1}{n}\sum(x_i-\mu_x)^2\right]=\sigma_x^2。\qquad(5.2)$$

注：虽然(2.6)中的 s^2 是 σ_x^2 的一个无偏估计，但是，统计量 $s=\sqrt{s^2}$ 并非总体标准差的无偏估计，不过通常这种偏度都小得是可保证$\sqrt{s^2}$估计量在统计实践中的使用。

§6. 随机变量的简单线性变换

假设我们有一个随机变量 x，其总体均值为 μ_x，总体方差为 σ_x^2。对每一个 x，我们按换算方程式或变换

$$u=cx+d\,,\qquad(6.1)$$

计算相应的新随机变量 u 的值，其中 c 和 d 是常数。那么，根据 x 的样本 $x_1, x_2, \cdots, x_n$，可以算得 u 的样本 $u_1, u_2, \cdots, u_n$，其中

$u_1 = cx_1 + d, u_2 = cx_2 + d, \cdots, u_n = cx_n + d$.

作为一种实际应用，通过选择 c 和 d，我们可以改变数据给出比诸 x 更便于计算的 u 值。下面的规则指明 u 随机变量与 x 随机变量的均值、方差和标准差是如何相互关联的。

规则 1：若 $u = cx + d$，其中 c 和 d 可以是任意常数，则

$$\mu_u = c\mu_x + d, V(cx) = \sigma_u^2 = c^2 V(x), \sigma_u = |c|\sigma_x. \qquad (6.2)$$

其中 $|c|$ 是 c 的绝对值，即 c 为正时，$|c| = c$，c 为负时，$|c| = -c$。

类似地，对应参数的样本估计量有关系

$$u = cx + d, \quad s_u^2 = c^2 s_x^2, \quad s_u = |c| s_x. \qquad (6.3)$$

例：

（i）若 $\mu_x = 10, \sigma_x^2 = 5$，且 $u = 4x - 3$，即 $c = 4, d = -3$，则 $\mu_u = 4(10) - 3 = 37$，$\sigma_u^2 = (4^2)(5) = 80$，和 $\sigma_u = 4(\sqrt{5}) = \sqrt{80}$.

（ii）若 $c = 1$，从而 $u = x + d$，我们有 $\mu_u = \mu_x + d, \sigma_u^2 = \sigma_x^2$，以及 $\sigma_u = \sigma_x$.

（iii）求 x 的样本 1960, 1910, 1940 和 1990 的均值和方差。减去 1900 再除以 10 的换算表达式为 $u = (x - 1900)/10 = \left(\frac{1}{10}\right)x - 190$。对应的 u 值为 6, 1, 4 和 9，给出 $\bar{u} = 5$ 以及 $s_u^2 = 34/3$。由 (6.3) 的逆运算得 $x = (u - d)/c = \{5 - (-190)\}/(1/10) = 1950$，以及 $S_x^2 = S_u^2/c^2 = 3400/3$，自由度为 3。

（iv）温度的华氏观察值的样本方差为 270。证明，若观察值用摄氏温度单位时，样本方差等于 83.3(℃)²。

规则 1 阐明，若每一观察值都加上(或减去)同一常数，如例(ii)，则总体和样本均值将移动相同常数值，而总体和样本方差和标准差保持不变；若每一观察值同乘以一个常数，则均值乘以这一常数，标准差乘以这常数的绝对值，而方差乘以这常数的平方。

§7. 独立随机变量的和与差

迄今我们考虑的样本其样本元都取自同一总体。常常更实用的是考虑更加一般的情况，这时，x_1 是取自均值 μ_1，方差 σ_1^2 的总

体的随机元，x_2 是取自均值 μ_2，方差 σ_2^2 的总体的随机元，等等，直到，比如说，x_k 为止，它是取自均值 μ_k，方差 σ_k^2 的总体的随机元，这里各均值可以，但没有必要是相等的，对于方差也一样。

独立随机变量是我们在一段时期内主要关心的，倘若所有的 x 随机变量是相互独立，第二条简单规则给出了下述随机变量的总体均值和方差，即这些 x 的线性组合的变量，比如，(x_1+x_2)，(x_1-x_2)和 $c_0+c_1x_1+c_2x_2+\cdots+c_kx_k$，其中 $c_0,c_1,\cdots,c_k$ 是常数，可以根据实际或理论的特殊应用作适当的选取。

规则 2 ：若 x_1 和 x_2 是两个独立分布的随机变量， 分别具有均值 μ_1 和 μ_2 以及方差 σ_1^2 和 σ_2^2，则

$$\begin{aligned}&\mu(x_1+x_2)=\mu_1+\mu_2,\\&V(x_1+x_2)=V(x_1)+V(x_2)=\sigma_1^2+\sigma_2^2.\end{aligned}\tag{7.1}$$

从而

$$\sigma(x_1+x_2)=\sqrt{\sigma_1^2+\sigma_2^2}。$$

这条规则有关方差的部分容易记忆如下：“两个独立随机变量的和的方差等于它们的方差的和。”

这条规则与规则 1 联合应用，我们可以容易地求出任一线性组合的均值和方差。

例（假定所有的 x -随机变量相互独立。）

(i) 若 $y=c_1x_1+c_2x_2$，则由规则 2，

$$\begin{aligned}&\mu_y=\mu(c_1x_1)+\mu(c_2x_2),\\&V(y)=V(c_1x_1)+V(c_2x_2);\end{aligned}$$

因此再利用规则 1，

$$\begin{aligned}&\mu_y=c_1\mu_1+c_2\mu_2,\\&V(y)=c\ \sigma_1^2+c\ \sigma^2。\end{aligned}$$

(ii) 若 $y=x_1-x_2$，则由上述结果，在 $c_1=1,c_2=-1$ 下得

$$\begin{aligned}&\mu_y=\mu_1-\mu_2,\\&V(y)=\sigma_1^2+\sigma_2^2=V(x_1+x_2\ )。\end{aligned}$$

因此，两个随机变量相互独立时，它们差的总体方差与它们和的总体方差相等。

(iii) 若 $y=c_1x_1+c_2x_2+c_3x_3$，c_1，c_2 和 c_3 为常数，则

$$\mu_y=\mu(c_1x_1+c_2x_2)+\mu(c_3x_3)=c_1\mu_1+c_2\mu_2+c_3\mu_3,$$

$$V(y)=V(c_1x_1+c_2x_2)V(c_3x_3)=c_1^2\sigma_1^2+c_2^2\sigma_2^2+c_3^2\sigma_3^2。$$

(iv) 若 $y=10+5x_1-2x_2-4x_3+x_4$，则反复运用规则 1 和 2，并注意到 10 是常数不是变量 $V(10)=0$，则得

$$\mu_y=10+5\mu_1-2\mu_2-4\mu_3+\mu_4,$$

$$\sigma_y^2=V(y)=25\sigma_1^2+4\sigma_2^2+16\sigma_3^2+\sigma_4^2.$$

因此

$$\sigma_y=\sqrt{25\sigma_1^2+4\sigma_2^2+16\sigma_3^2+\sigma_4^2}。$$

注：对应于上面的一些样本方差的结果将在稍后的实际问题中和 §68⁺ 中出现，并且在 §68⁺ 中还要给出相关，即不独立情况下的类似结果。

§8. 样本均值的分布

假设从某一无限总体或分布中，我们抽取一个 n 个观察值的随机样本并计算其样本均值，记之为 x_1，然后，再取另一随机样本并求其均值，记之为 x_2，等等。于是，我们应该累积第二个分布，样本均值的分布。这个导出分布的个体元是 $x_1,x_2,\cdots$。它的总体均值与原总体分布的相同，即 μ_x。但是，样本均值分布的总体方差等于 σ_x^2/n，因此，它将随样本容量的增加而减小。从而样本均值没有原来 x 的那么分散，而是更紧密地聚集在 μ_x 周围。相应地，如果我们取一个这种(有两个或两个以上观察值的)样本均值，则它比单个观察值更可能接近总体均值。这就是为什么要推荐用大样本去求精确估计的原因。

需要记住的结论是

$$\mu_{\bar{x}}=\mu[\bar{x}]=\mu_x, \tag{8.1}$$

$$\sigma_{\bar{x}}^2=V(\bar{x})=\mu[(\bar{x}-\mu_x)^2]=\sigma_{\bar{x}}^2, \tag{8.2}$$

$$\sigma_{\bar{x}}=sd(\bar{x})=\frac{\sigma_x}{\sqrt{n}}。 \tag{8.3}$$

这些结论容易用 §6与§7中的规则建立。首先，我们导出有关样本总和的结论。

样本总和—— 其总体均值与方差：若样本观察值为 $x_1, x_2, \cdots, x_n$，根据有关独立观察值总和的规则 2,我们得知

$$\mu[x_1 + x_2 + \cdots + x_n] = \mu_1 + \mu_2 + \cdots + \mu_n = n\mu_x。 \tag{8.4}$$

因为,这里所有的 n 个 x 来自同一总体,故 $\mu_1 = \mu_2 = \cdots = \mu_n = \mu_x$。同样,$V(x_1) = V(x_2) = \cdots = V(x_n) = \sigma_x^2$ ，从而，对于方差，规则 2 给出

$$V(\sum x) = V(x_1 + x_2 + \cdots + x_n) = V(x_1) + V(x_2) + \cdots + V(x_n)$$
$$= n\sigma_x^2。 \tag{8.5}$$

因此,样本容量 n越大，样本总和变化也越大。

样本均值—— 其总体均值和方差：样本均值 $\bar{x}$ 与样本总和 $\sum x$都是随机变量，各有一个建立在原来随机变量 x 的基础上的概率密度函数。由于 $\bar{x} = \sum x/n$,因此利用规则 1 (一个变量乘以一个常数),取常数 $c = 1/n$,则由(8.4)和(8.5)即得

$$\mu_{\bar{x}} = \left(\frac{1}{n}\right) n\mu_x = \mu_x,$$

如(8.1)中那样,另外又如(8.2)就有

$$V(\bar{x}) = \left(\frac{1}{n}\right)^2 V(\sum x) = \frac{n\sigma_x^2}{n^2} = \frac{\sigma_x^2}{n}。$$

估计量：由于(2.6)中的 $s_x^2 = \sum' xx/(n-1)$ 是 σ_x^2 的样本估计，样本均值分布的参数的估计结果为

x 是 $\mu_{\bar{x}} = \mu_x$ 的估计；

$s_{\bar{x}}^2 = \dfrac{s_x^2}{n}$，自由度等于$(n-1)$,是 $\sigma_{\bar{x}}^2$的估计； (8.6)

$s_{\bar{x}} = \dfrac{s_x}{\sqrt{n}}$是 σ_x的估计，

$s_{\bar{x}}$ 有时称为均值的‘标准误差’。

例：

(i) 若从 $\mu_x = 17$, $\sigma_x^2 = 1600$ 的 x-随机变量总体中抽取容量为 n 的重复随机样本，试求 $(a)n=4$, $(b)n=9$ 和 $(c)n=16$ 时，