第1部分

回归分析基础

本书的第1部分涉及计量经济学建模的最基本概念，这些概念形式上虽然简单，但是它们在商业和经济上的各种各样的应用却是相当广泛的。在这些回归模型里，被研究的变量被认为是若干解释变量的一个线性函数。单方程回归模型之所以重要，是因为它们不仅能够用来做假设检验和预测，而且构成联立方程模型和时间序列模型的基础。

第1章介绍初等曲线拟合以及最小二乘概念。第 2章是本书后面的分析所需要的基本统计概念的一个比较全面的综述。第 3章以一元回归模型为例，说明回归参数估计所需要的各种统计性质，重点在于假设检验和拟合优度的度量。第4章将一元回归模型推广到多元回归模型。回归模型中包含多个解释变量会导致包括多重共线在内的其他计量经济学方面的问题。多重共线问题的存在将会影响我们对回归系数的解释。我们将会讨论有助于解决这些问题的其他回归统计量。

第 1 章

回归模型介绍

在本章中我们从双变量的线性回归模型开始讨论计量经济学。在第 1节，我们以学生平均成绩为例，对曲线拟合进行讨论，讨论曲线拟合的最小二乘估计法，并将其与其他几种曲线拟合方法进行比较。在第 2节，我们将推导最小二乘估计法。在本章的最后是最小二乘回归估计的3个基本应用。

1.1 曲线的拟合

对变量进行度量可以从若干不同的角度并有各种不同的形式。描述变量关于时间变化的数据称为时间序列数据，可以是日数据、周数据、季度数据或年度数据。描述个人、企业或其他事物在某一给定时刻的数据称为截面数据。关于某一时刻家庭费用的市场研究很可能会利用截面数据；也可以运用截面数据对一组商业财务报表进行检验，以便了解一个行业中各企业的典型行为。同时包括时间序列数据和截面数据的数据集称为混合数据，混合数据可以用来研究不同企业的行为随时间变化的情况。

假设我们对变量X和Y之间的关系感兴趣。为了研究它们之间的统计关系，我们需要每一个变量的一组观测值，以及有关变量之间关系的数学表达式的假设。这组观测值称为一个样本[⊖]。我们主要讨论的是X和Y之间呈线性关系——即由直线描述——的情况。在X和Y之间为线性关系的假设下，我们的目的是确定一个规则，从而决定 X和Y之间的“ 最佳 ”直线。

例如，假设我们希望检验学生的平均成绩是否可以用学生父母亲的收入进行解释。表 1-1列出了8个(虚构的)样本点，图1-1是这8个样本点的散点图。很多直线都可以用来拟合这些样本点。可以将最小的 X值点和最大的X值点相连(直线l_1)，也可以靠眼睛画出适合所有散点的直线(直线l_2)。比较好的方法是选择一条直线使图上各点到该直线的垂直距离 (可正可负)的和为0(这些距离称为离差，如图1-2所示)。这种做法认为大小和符号完全相同的离差具有相同的重要性。不幸的是，它具有下面不受欢迎的特点：大小相等、符号相反的离差会互相抵消。因此，会有一条或多条离差和为零、但是拟合程度很差的直线。

如果对样本点到拟合直线的离差的绝对值进行极小化，这个方法就比前一种方法好。这里隐含着一种观点，即认为离差的重要性与其数量大小成正比。虽然极小化离差绝对值之和的做法很不错，但是它有几个缺陷。首先，这种方法不易计算。第二，有理由认为较大的离差应当比较小的离差得到更多的重视。例如，具有一个两个单位误差的预测或许应当被认为比具有两

⊖ 样本数据是从代表真实关系的总体中选取的观测值。

个一个单位误差的预测要差。

表1-1　学生的平均成绩和家庭收入

Y 平均成绩	*X* 父母的收入/1000美元
4.0	21.0
3.0	15.0
3.5	15.0
2.0	9.0
3.0	12.0
3.5	18.0
2.5	6.0
2.5	12.0

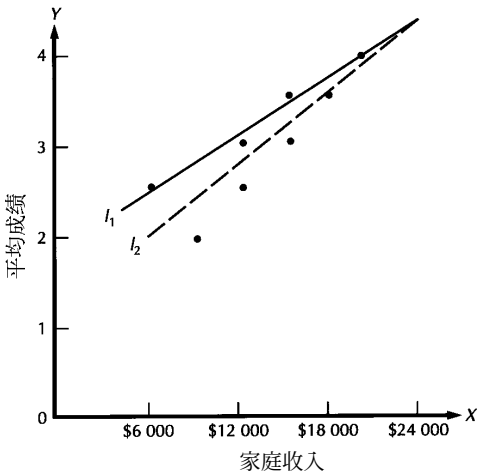


图1-1　散点图

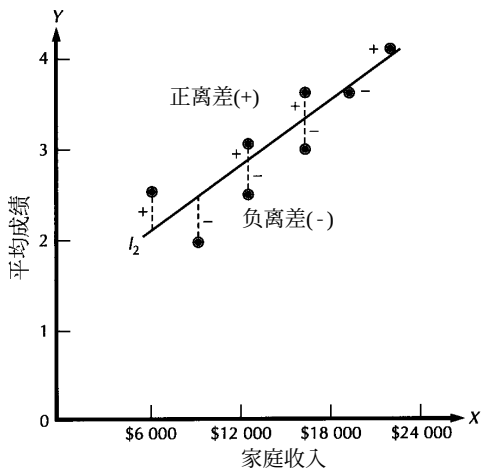


图1-2　离差

有一种方法计算上简便易行，且对较大误差的惩罚比对较小误差的惩罚更大，这就是最小二乘估计法。最小二乘法的做法如下：“**最佳拟合直线**”**是使样本点到该直线的离差平方和达到最小的直线**(**采用垂直距离**)。我们将在下面两章中看到，最小二乘法还能够用来做统计检验。

在本书中，我们将主要利用最小二乘估计法，但是其他估计法也是可行的，并且偶尔也会用到。通过图1-3a和1-3b我们可以看到最小二乘法和其他方法的关系。图 1-3a在横轴上表示数据点到直线的离差，并在纵轴上表示与该离差相关的“损失”。对于最小二乘估计法，每一个离差的损失是该离差的平方。对于最小绝对值估计法，损失是该离差的绝对值。最小二乘估计法和最小绝对值估计法的损失函数都是关于离差符号对称的函数，但是最小二乘法的损失函数对于较大离差的惩罚比最小绝对值法的损失函数大。

若存在一个或更多较大的离差时，最小二乘估计法存在一个问题。假设第一个学生的平均成绩写错了，本来应该是4分，但被写成1分。如果图1-1中的直线l_2是可能的最小二乘估计直线，则第一个数据点的离差将会非常大，其平方和就更大，最佳拟合直线将发生很大变化，其倾斜度就会变小。最小二乘法对大离差的惩罚使得该估计法充分强调拟合直线和第一数据点的关系，其结果是最小二乘估计直线的斜率 (和截距)对那些离真正回归直线较远的数据点非常敏感。我们称这些离回归直线距离很远的数据点为异常点。当然异常点也可能反映变量间关系的重要信

息。因此，不能不经过深入的分析就随便丢掉异常点。对异常点的仔细研究确实有助于我们发现错误，并做出修正

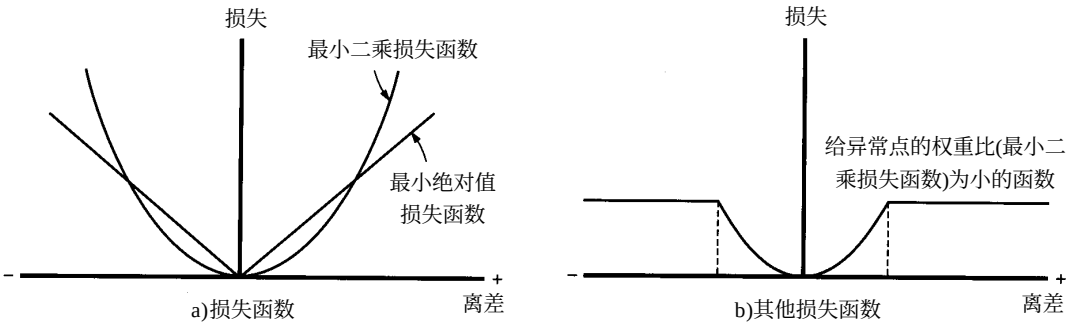


图 1-3

如何处理最小二乘估计法对异常点的敏感性呢？最简单的解决办法是去掉异常点，重新计算最小二乘直线，通过原来的和新的最小二乘直线的截距和斜率，我们就能够判断我们的结果对异常点的敏感程度。因为关于异常点产生原因的判断是多样的，因此更好的办法是给大的离差较小的权重，图 1-3b就是这样的一个例子，它表示的是与最小二乘法或最小绝对值法相比，对异常点不那么敏感的一种损失函数。

1.2 最小二乘估计法的推导

构造统计关系的目的通常是为了预测或说明一个或多个可以控制或可以解释的变量的变化对另一个变量的影响。对于图 1-1中的散点，我们可以写出 $Y=a+bX$的线性方程，其中左边的变量Y称为非独立变量，右边的变量 X称为独立变量。因为我们想解释或预测 Y的变化，所以我们的目的很自然地是要选择使垂直离差平方和最小的直线[⊖]。

为了获得计算a，b值的最小二乘公式，我们必须运用基本的数学工具。对于那些不熟悉求和算子性质的人，我们建议其简单地复习附录 1-1；对于那些对偏导数的使用没有把握的人，我们认为对这些细节的理解并不重要。

最小二乘估计的准则可以写成如下公式：

$$\text{Minimize} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \tag{1-1}$$

其中$\hat{Y}_i = a+bX_i$ 代表截距是a、斜率是b的直线方程，Y_i是第i个观测值的实际Y值，对应于该观测值的X值，N是观测值的个数。$\hat{Y}_i$称为Y_i的预测值或拟合值，是观测值X_i在拟合直线上相应的Y值。这在图 1-4中可以看得很清楚，这里的离差是 Y的实际值与Y_i的差。因此，对于 X上的每一个观测都有一个从实际值到拟合值的离差，我们希望使这些离差的平方和最小，这样就使我们能够对一条直线对数据的拟合程度进行度量 (见第3章)。

我们的问题是选择能使公式 (1-1)中的表达式最小的 a和b值，可以运用一些基本的微积分和代数方法得到这些值。为了使我们讨论直接扼要，我们把微积分推导的细节放在附录 1-1

⊖ 一般来说，我们选择采用方程$Y=a+bX$的形式，而不是$X=A+BY$的形式，意味着我们已经做出变量Y的变化是由变量X“引起”，而不是相反的判断。因此，在学生平均成绩的例子中，我们假设平均成绩依赖于家庭收入。如果改变关于因果关系的观点，认为家庭收入依赖于平均成绩的话，我们就会把方程写为$X=A+BY$，求拟合直线的做法也要做相应的改动。因为不同的方程会得出不同的回归直线，所以这一点很重要。

中[⊖]。正如附录中所显示的，斜率和截距的最小二乘解为：

$$b = \frac{N\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{N\Sigma X_i^2 - (\Sigma X_i)^2} \tag{1-2}$$

$$a = \frac{\Sigma Y_i}{N} - b\frac{\Sigma X_i}{N} = \overline{Y} - b\overline{X} \tag{1-3}$$

其中 $\overline{X}$和$\overline{Y}$分别代表Y和X的样本均值。

现在考虑X和Y的样本均值都是0的特殊情况如何简化公式(1-2)和(1-3)。首先改写等式(1-3)，我们注意到：

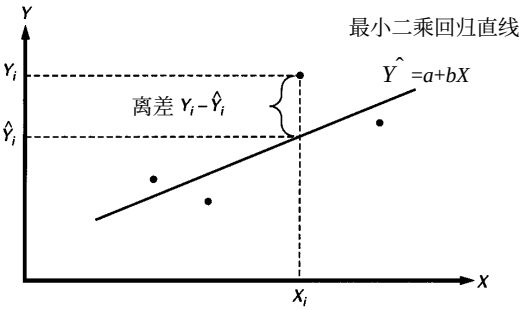


图1-4 拟合值

$$a = \overline{Y} - b\overline{X} = 0 \tag{1-4}$$

因此当X和Y的样本均值为0时，拟合回归直线的截距为 0。为了得到这种特殊情况下斜率的估计值，我们把(1-2)式的分子分母都除以N^2：

$$b = \frac{\Sigma X_i Y_i/N - (\Sigma X_i/N)(\Sigma Y_i/N)}{\Sigma X_i^2/N - (\Sigma X_i/N)^2}$$

代入$\overline{X}$和$\overline{Y}$得到

$$b = \frac{\Sigma X_i Y_i/N - \overline{X}\,\overline{Y}}{\Sigma X_i^2/N - \overline{X}^2}$$

但是，由假设$\overline{X}$=$\overline{Y}$=0，因此

$$b = \frac{\Sigma X_i Y_i/N}{\Sigma X_i^2/N} = \frac{\Sigma X_i Y_i}{\Sigma X_i^2} \tag{1-5}$$

等式(1-5)比(1-2)简单，这一事实说明不论均值是否等于零，如果我们在最小二乘估计式中采用变量关于样本均值的离差作为新的变量的话，就能使事情变得简单并有助于我们的理解。因此，我们把每一个X和Y的观测值变换为关于各自的样本均值的离差形式：

$$x_i = X_i - \overline{X} \qquad y_i = Y_i - \overline{Y}$$

在这样的定义下，由于变量 x和y的均值都是0[⊜]，因此最小二乘斜率的估计式可以由 (1-5)式得到。实际上，我们把 X和Y图的原点移到了样本均值点，从而使数据得到中心化。在这种情况下，小写的变量是大写变量的中心化结果。

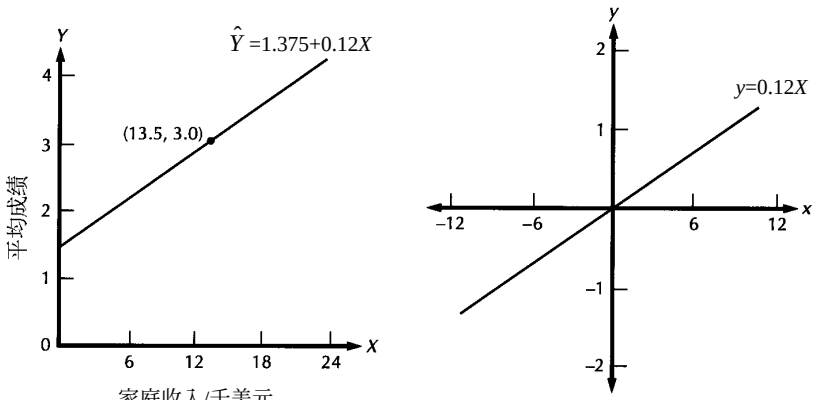
斜率的最小二乘估计式是：

$$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \tag{1-6}$$

图1-5显示了变量转换成离差形式的中心化过程。图 1-5a用原始观测值显示回归直线，而图1-5b用的是观测值的离差形式。注意：两种回归直线的斜率估计式是相同的，这一点从等式(1-6)看很明显，因为计算用的只是变量的离差形式。但是，图 1-5b中回归直线的截距等于 0，这是因为方程(1-4)以及$\overline{x}$和$\overline{Y}$等于零的事实。因此，如果我们选择用数据的离差形式计算，就是将回归直线的原点移到样本均值点，但是未改变直线的斜率。还应注意到图 1-5b中的直线通过原点，这等价于图1-5a中的直线通过均值点($\overline{X}$，$\overline{Y}$)这一事实。

⊖ 我们建议读者按照附录实际进行推导，以便更加熟悉我们所用的记号以及求和运算。

⊜ 例如，$\bar{x} = \frac{\Sigma x_i}{N} = \frac{\Sigma(X_i - \overline{X})}{N} = \frac{\Sigma X_i}{N} - \overline{X} = \overline{X} - \overline{X} = 0$



a) 原始回归直线

b) 变换后的回归直线

$x=X-13.5$ $y=Y-3.0$

图1-5 数据离差形式的使用

例1.1 平均成绩

在前面提到的学生平均成绩的例子中，最小二乘估计法使我们得到截距为 1.375、斜率为0.12的回归直线：$\hat{Y}=1.375+0.12X$ ⊖。计算的详细情况见表1-2。(回归直线l_2和原数据点如图1-2。)对于任意家庭收入X，回归直线能使我们预测学生的平均成绩Y。例如，一个家庭收入为12 000美元的学生的预测平均成绩为$\hat{Y}=1.375+0.12\times(12)=2.815$。虽然平均成绩的估计值不一定每一次都很精确，但它会提供一个很好的近似。例如，我们可能注意到，在原样本中(见表1-1)父母亲的收入同为12 000美元的两个学生，其平均成绩分别是3.0和2.5，而预测的平均成绩碰巧介于两个真实的数据之间。

表1-2 平均成绩例题(计算过程)

$\bar{X} = 13.5$ $\bar{Y} = 3.0$

(1) $x_i = X_i - \bar{X}$	(2) $y_i = Y_i - \bar{Y}$	(3) $x_i y_i$	(4) x_i^2
7.5	1.0	7.5	56.25
1.5	0.0	0.0	2.25
1.5	0.5	0.75	2.25
−4.5	−1.0	4.5	20.25
−1.5	0.0	0.0	2.25
4.5	0.5	2.25	20.25
−7.5	−0.5	3.75	56.25
−1.5	−0.5	0.75	2.25
$\Sigma x_i = 0$	$\Sigma y_i = 0$	$\Sigma x_i y_i = 19.50$	$\Sigma x_i^2 = 162.00$

$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} = 0.120$ $a = \bar{Y} - b\bar{X} = 1.375$ $\hat{Y} = 1.375 + 0.12$

回归直线的斜率告诉我们，家庭收入1 000美元的变化将导致平均成绩0.12的变化。斜率为正值和高平均成绩的学生来自于高收入家庭的假设相一致。截距为 1.375告诉我们如果家庭收入为0，平均成绩的最好预测值是1.375。因为在我们的样本中没有任何家庭的收入接近于0，我们不能过于信任这个结果。

⊖ 在整个第1部分中，我们用带“帽子”的非独立变量表示拟合值。为了简便起见，我们会在其他部分放松这一规定。

我们将在第3章中更加详细地研究双变量的线性回归模型，但是应该在这里做一点评价。在模型$Y=a+bX$中，斜率b是dY/dX(Y的变化率和X的变化率的比)的估计值，这使我们能够自然地解释回归斜率，然而截距的解释依赖于是否有足够多X值接近于0的观测，以便使其结果具有统计意义。如果是这种情况，我们就能把截距解释为$X=0$时Y的估计值。然而如果没有充足的X值接近于0的观测，截距只是最小二乘直线的高度。

例1.2 诉讼案件的急速增长

在美国法院立案的诉讼案件数量随时间增长有多快？增长的速度有多稳定？最近一项关于民事诉讼案件增长趋势的研究提供了一些有用的信息[⊖]。用1977年第2季度~1988年第3季度的季度时间序列数据，以每一个季度的诉讼案件数量为Y，时间变量为T，估计回归方程。对于1977年第2季度T为1，然后每一个季度增加1。估计回归方程为

$$\hat{Y} = 13.00 + 51.03T$$

回归斜率系数告诉我们，确实存在着诉讼案件急速增长的问题，每一个季度增加的案件数略多于51件。当然，一个回归方程的根本还不是计算诉讼案件增长速度。在第3章我们将会看到回归方法的一个重要优点：它能使我们判断增长速度估计值的统计显著性。

在整个被研究期间，诉讼案件的增长不是不变的，研究发现案件数量对商业周期很敏感，失业率U越高时，民事案件的数量越大，估计的回归方程为

$$\hat{Y} = -144.38 + 168.91U$$

这个方程告诉我们，失业率每1%的增长将会增加170个案件。

例1.3 公用事业公司股票价格

在一项比较大的关于公司财务的研究中，假设公共事业公司价格收益比受其负债产权率的影响，这是有道理的，因为人们一般认为较高的负债产权率能够引起收益规律上的较大变化，这将引起股票价格下跌风险的增加，从而造成价格收益比的降低。模型可以表示为：

$$Y=a+bX$$

其中 Y——公司的价格收益比(普通股每股价格除以每股收益)

X——负债产权率

我们认为b的值应该是负的，但是对它的大小没有任何估计。变量Y和X的观测是公共事业公司的一个截面数据(在某一给定时刻)。线性回归的结果是

$$\hat{Y} = 10.2 - 4.07X$$

系数-4.07证明了我们的假设，然而要想知道我们对这个假设的把握有多大，需要用到第2章中的一些统计检验。

⊖ 参见P. Siegelman and J. J. Donohue III, "The Selection of Employment Discrimination Disputes for Litigation: Using Business Cycle Effects to Test the Priest-Klein Hypothesis," *Journal of Legal Studies*, vol. 24, pp. 427-462, June 1995. 这里引用的结果只是该文章全面研究结果的一个简化。

附录1.1 求和算子的运用

因为计量经济学中的许多问题都涉及到数量的求和，所以复习一下（或熟悉）求和符号是很有益的。我们在这本书中采用大写希腊字母 $\sum$ 代表变量各个观测值的和。例如令 X 代表变量“家庭收入”，然后运用下标记号，

$$X_1, X_2, \ldots, X_N$$

表示N个家庭收入的观测值，则家庭收入的总和 $(X_1+X_2+\cdots+X_N)$可以被表示为

$$\sum_{i=1}^{N} X_i = X_1 + X_2 + \cdots + X_N \tag{A1-1}$$

下面是一些求和算子的运算法则：

法则1：一个变量的倍数的和等于那个变量的和的倍数。

$$\sum_{i=1}^{N} kX_i = k\sum_{i=1}^{N} X_i \tag{A1-2}$$

法则2：两个变量观测值和的总和等于它们分别总和的和。

$$\sum_{i=1}^{N} (X_i + Y_i) = \sum_{i=1}^{N} X_i + \sum_{i=1}^{N} Y_i \tag{A1-3}$$

法则3：N个常数观测值的和等于该常数的N倍。

$$\sum_{i=1}^{N} k = kN \tag{A1-4}$$

运用这三个法则我们可以得到一些有用的有关随机变量均值、方差以及协方差的结果。因为这些概念在第2章中有更全面的讨论，我们在这里仅讨论一些代数（而不是统计的）性质。首先，我们定义X变量的N个观测值的均值或平均值为：

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{A1-5}$$

运用这个定义，我们可以证明法则4。

法则4：变量X观测值关于其均值的离差和为0。

$$\sum_{i=1}^{N} (X_i - \bar{X}) = 0 \tag{A1-6}$$

（关于法则4的证明参见脚注[⊖]。）在本书的正文中，我们将会经常用到变量或观测值的离差形式。用小写字母表示离差形式，即 $x_i=X_i-\bar{X}$，则法则4变为

$$\sum_{i=1}^{N} x_i = 0 \tag{A1-7}$$

现在我们定义X的方差为

$$\text{Var}\,(X) = \frac{1}{N}\sum_{i=1}^{N} (X_i - \bar{X})^2 \tag{A1-8}$$

X与Y的协方差为

$$\text{Cov}\,(X, Y) = \frac{1}{N}\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) \tag{A1-9}$$

⊖ 例如，$\bar{x} = \frac{\Sigma x_i}{N} = \frac{\Sigma(X_i - \bar{X})}{N} = \frac{\Sigma X_i}{N} - \bar{X} = \bar{X} - \bar{X} = 0$

运用这些定义和前面的结果，我们可以证明最后两个求和法则。

法则5：X与Y的协方差等于X和Y乘积的均值与X和Y均值乘积的差。

$$\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-\bar{X}\,\bar{Y} \tag{A1-10}$$

证明 $$\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-\frac{1}{N}\sum_{i=1}^{N}\bar{X}Y_i-\frac{1}{N}\sum_{i=1}^{N}X_i\bar{Y}+\frac{1}{N}\sum_{i=1}^{N}\bar{X}\,\bar{Y}$$

并利用法则1，我们得到

$$\text{Cov}\,(X,Y)=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-\frac{1}{N}\bar{X}\sum_{i=1}^{N}Y_i-\frac{1}{N}\bar{Y}\sum_{i=1}^{N}X_i+\frac{1}{N}\sum_{i=1}^{N}\bar{X}\,\bar{Y}$$

再利用X和Y均值的定义，我们有：

$$\text{Cov}\,(X,Y)=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-\bar{X}\,\bar{Y}-\bar{Y}\,\bar{X}+\frac{1}{N}\sum_{i=1}^{N}\bar{X}\,\bar{Y}$$

$$=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-2\bar{X}\,\bar{Y}+\bar{X}\,\bar{Y}$$

因为由法则3，有 $$\sum_{i=1}^{N}\bar{X}\,\bar{Y}=N\bar{X}\,\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}X_iY_i-\bar{X}\,\bar{Y}$$

因为法则5把X和Y作为两个变量，所以法则6很容易由法则5推出。

法则6：X的方差等于X观测值平方的均值与X均值平方的差。

$$\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2=\frac{1}{N}\sum_{i=1}^{N}X_i^2-\bar{X}^2 \tag{A1-11}$$

注意，当X和Y的均值碰巧都为0时(或用离差形式表示时)，协方差和方差的定义成为(我们在这里省略了求和上下限)：

$$\text{Cov}\,(x,y)=\frac{1}{N}\sum x_iy_i \quad \text{以及} \quad \text{Var}\,(x)=\frac{1}{N}\sum x_i^2$$

有些情况下有必要把求和运算运用于两个随机变量，称为双重求和。具体来说，令 X_{ij}为一个随机变量，其中i=1,2,⋯,N,j=1,2,⋯,N。当然，共有N^2个值。现在我们定义N^2个值的双重求和为

$$\sum_{i=1}^{N}\sum_{j=1}^{N}X_{ij}=\sum_{i=1}^{N}(X_{i1}+X_{i2}+\cdots+X_{iN})$$

$$=(X_{11}+X_{12}+\cdots+X_{1N})+(X_{21}+X_{22}+\cdots+X_{2N})$$

$$+\cdots+(X_{N1}+X_{N2}+\cdots+X_{NN})$$

下面是双重求和的一些有用的法则。

法则7

$$\sum_{i=1}^{N}\sum_{j=1}^{N}X_iY_j=\left(\sum_{i=1}^{N}X_i\right)\left(\sum_{j=1}^{N}Y_j\right) \tag{A1-12}$$

注意法则7中的双重求和与单重求和$\sum_{i=1}^{N}X_iY_i$是不同的，后者只包含N项(不是N^2项)。

法则8：

$$\sum_{j=1}^{N}(X_{ij}+Y_{ij})=\sum_{i=1}^{N}\sum_{j=1}^{N}X_{ij}+\sum_{i=1}^{N}\sum_{j=1}^{N}Y_{ij} \tag{A1-13}$$

附录1.2 最小二乘参数估计的推导

正如正文中提到的那样，我们的目标是使 $\sum(Y_i-\hat{Y}_i)^2$最小，其中$\hat{Y}_i=a+bX_i$是相应于观测值X_i的Y_i的拟合值。

我们通过下面的做法达到求极小的目的：对上述表达式求关于 a，b偏导数，得到[⊖]：

$$\frac{\partial}{\partial a}\Sigma(Y_i-a-bX_i)^2=-2\Sigma(Y_i-a-bX_i) \tag{A1-14}$$

$$\frac{\partial}{\partial b}\Sigma(Y_i-a-bX_i)^2=-2\Sigma X_i(Y_i-a-bX_i) \tag{A1-15}$$

令这些偏导方程等于零，且在方程两边同时除以 -2得到：

$$\Sigma(Y_i-a-bX_i)=0 \tag{A1-16}$$

$$\Sigma X_i(Y_i-a-bX_i)=0 \tag{A1-17}$$

最后，整理式(A1-16)和(A1-17)得到方程组(称为正规方程组)：

$$\Sigma Y_i=aN+b\Sigma X_i \tag{A1-18}$$

$$\Sigma X_iY_i=a\Sigma X_i+b\Sigma X_i^2 \tag{A1-19}$$

用$\sum X_i$乘以式(A1-18)，用N乘以式(A1-19)，我们就能同时解出a和b：

$$\Sigma X_i\Sigma Y_i=aN\Sigma X_i+b(\Sigma X_i)^2 \tag{A1-20}$$

$$N\Sigma X_iY_i=aN\Sigma X_i+bN\Sigma X_i^2 \tag{A1-21}$$

从(A1-21)减去(A1-20)得到

$$N\Sigma X_iY_i-\Sigma X_i\Sigma Y_i=b[N\Sigma X_i^2-(\Sigma X_i)^2] \tag{A1-22}$$

由此得到

$$b=\frac{N\Sigma X_iY_i-\Sigma X_i\Sigma Y_i}{N\Sigma X_i^2-(\Sigma X_i)^2} \tag{A1-23}$$

对于已知的b，由式(A1-18)计算a:

$$a=\frac{\Sigma Y_i}{N}-b\frac{\Sigma X_i}{N} \tag{A1-24}$$

练习

1.1 假设你在一个虚构的国家中掌管中央货币机构，你有以下货币数量和国民收入的历史数据(单位均为百万美元)

年份	货币数量	国民收入	年份	货币数量	国民收入
1987	2.0	5.0	1992	4.0	7.7
1988	2.5	5.5	1993	4.2	8.4
1989	3.2	6.0	1994	4.6	9.0
1990	3.6	7.0	1995	4.8	9.7
1991	3.3	7.2	1996	5.0	10.0

⊖ 求和号中没有标出求和标号，假设求和标号包括所有观测值，即1，2，…，N。

(a) 绘制散点图，然后以国民收入 Y对货币数量 X做回归估计，并在散点图中绘出回归直线。

(b) 如何解释回归直线的斜率和截距？

(c) 如果你能独立控制货币供给并希望 1997年的国民收入达到 12.0，你的货币供给应控制在什么水平？请说明。

1.2　对本章学生平均分数例题用收入对平均成绩做回归分析，并将结果与平均成绩对收入的回归分析结果进行比较，两个结果为什么不同？

1.3　(a) 假设已经得到关系式 $Y=a+bX$的最小二乘估计，现在决定把 X变量的单位扩大10倍，这样做对回归斜率和截距会有什么样的影响？

(b) 将(a)的结果加以推广：用下面的方式改变 X和 Y的单位，评价这样做对回归方程的影响。你能得出什么结论？

$$Y^* = c_1 + c_2 Y \qquad X^* = d_1 + d_2 X$$

1.4　当所有自变量的值全都相同时，最小二乘的斜率和截距估计将会怎样？请直观地解释其原因。

1.5　证明估计回归直线通过 X 均值点 $(\overline{X}, \overline{Y})$。

提示　证明 $\overline{X}$ 和 $\overline{Y}$ 满足方程 $Y=a+bX$，其中 a，b由公式(1-2)和(1-3)定义。

1.6　如何解释例1-2中 Y对 U的回归截距 -144.38？说明为什么该截距值没有实际意义？

1.7　为检验斜率和截距的最小二乘估计值对异常值的敏感程度，做如下计算：

(1) 假设例1-1中第一个观测值是(21.0，1.0)而不是(21.0，4.0)，重新估计斜率和截距。

(2) 从样本中去掉第一个观测值，重新估计斜率和截距。

(a) 将(1)和(2)中的斜率和截距估计与例 1-1的结果进行比较，并进行说明。包含这两条回归直线的图会很有帮助。为什么最小二乘估计对单个数据值这么敏感？

(b) 画出(1)中的最小二乘直线图，能否推断第一个数据点是异常点？请讨论。

第2章

统计基础知识复习

即使从最应用性的角度，经济计量学的学习也要求对统计学有较好的了解。我们假设大多数读者学过统计学，但我们知道这些知识需要更新。在继续学习计量经济学之前，我们将复习统计学的观点，这些知识将在以后的各阶段中发挥它们的作用。为帮助读者把注意力放在重要的观点而不是细节上，我们把大部分推导放在附录 2-1中。

2.1 随机变量

随机变量是变量，它可以取不同的值，并且取每一个值的概率小于等于 1。我们可以通过研究随机变量生成各个取值的过程来描述一个随机变量，这个过程称做概率分布。概率分布列出所有可能出现的结果及每个结果发生的概率。我们可以将随机变量定义为一个函数，这个函数为每一个试验结果赋予一个实数值。例如，假设抛硬币出现正面的取值为 1，反面的取值为0(如果硬币是均匀的，出现正面的概率将为 1/2)。此例中我们可以把抛硬币的取值看作一个随机变量；生成这个随机变量的过程是二项概率分布。

弄清离散型随机变量和连续型随机变量之间的区别是很有用的。一个连续型随机变量可以取实数轴上的任何值，而一个离散型随机变量只能取若干特定的实数值。图 2-1表示的是离散型随机变量和连续型随机变量的概率函数。由图 2-1中离散型随机变量的分布，我们看到取值10和20发生的概率都是0.25，而取值40发生的概率为0.50。在图2-1的连续型分布中，随机变量的取值位于某两个值之间的概率是由这两个值之间连续密度函数之下的面积决定的。在此例中，随机变量取值位于10和20之间的概率约等于0.3，即图中的阴影部分。

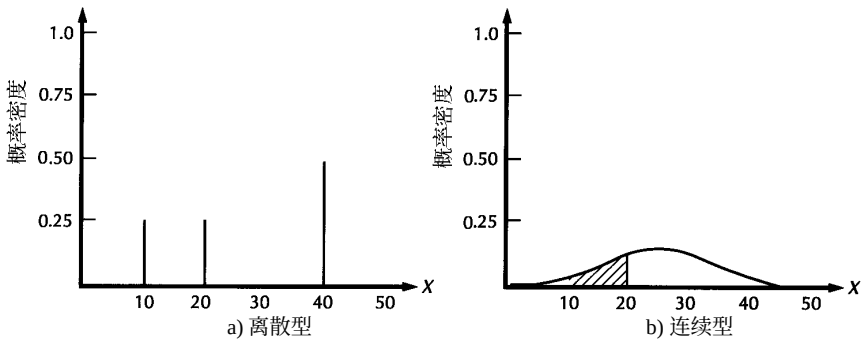


图2-1 概率密度

2.1.1 期望值

人们经常用均值和方差来描述概率分布，它们都是由期望算子 E定义的。因为我们将从离散型随机变量开始讨论，因此设 X_1、X_2…X_N代表随机变量X的N个可能结果，则X的均值或期望值是所有可能结果的一个加权平均值，其中权重为各结果发生的概率。具体说来， X的均值(记为μ_X)定义为：

$$\mu_X = E(X) = p_1X_1 + p_2X_2 + \cdots + p_NX_N = \sum_{i=1}^{N} p_iX_i \tag{2-1}$$

其中p_i为X_i发生的概率，$\sum p_i$=1，且E( )为期望算子。

期望值应与样本均值区分开来，后者表示样本的平均值，而样本是一组对某一概率分布进行观测得到的观测值(观测值的选取一般是随机的)。X的一组观测值的平均值记为$\overline{X}$。

随机变量的方差是随机变量在其均值周围分散或离散程度的一个度量，记为 σ_X^2。(在离散型随机变量情况下)，方差的定义为

$$\text{Var}\ (X) = \sigma_X^2 = \sum_{i=1}^{N} p_i[X_i - E(X)]^2 \tag{2-2}$$

因此方差是X的取值与其期望值之差平方的加权平均，其中权重为相应取值发生的概率。式(2-2)中的方差本身也是一个期望，这是因为

$$\sigma_X^2 = E[X - E(X)]^2 \tag{2-3}$$

方差的(正)平方根称为标准差。

期望算子有很多有用的性质，特别在讨论随机变量的期望和方差时更为有用。我们建议读者仔细阅读附录2-1中的细节。以下是有关期望算子的三个主要的结论：

结论1 $E(aX + b) = aE(X) + b$ 其中X是随机变量，a、b是常数

结论2 $E[(aX)^2] = a^2E(X^2)$

结论3 $\text{Var}\ (aX + b) = a^2\ \text{Var}\ (X)$

2.1.2 随机变量的联合分布

研究X和另一个随机变量Y之间的联合分布是很有用处的。在离散情形下，联合分布可以用一个概率分布表描述，这个分布表列出X和Y所有可能结果出现的概率。例如，如果Y是这样的一个随机变量，即当户主受过大学教育时，取值为 1；否则取值为0；而X是前面描述过的家庭收入变量，那么X和Y的联合分布如下：

结果	概率	结果	概率
X=\$5000, Y=1	0	X=\$10 000, Y=0	1/8
X=\$5000, Y=0	1/4	X=\$15 000, Y=1	1/3
X=\$10 000, Y=1	1/8	X=\$15 000, Y=0	1/6

注意：所有概率都是非负的，而且它们的和为 1。

与单个随机变量的情况一样，期望算子对于描述联合分布的重要性质也是很有用的。我们将X和Y的协方差定义为X、Y与各自均值离差乘积的期望；

$$\text{Cov}\ (X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}(X_i - E(X))(Y_j - E(Y)) \tag{2-4}$$

其中p_{ij}表示X与Y发生的联合概率。

协方差是X与Y之间线性相关关系的一个度量。如果两个变量总是同时大于或小于各自的均值，则协方差为正，如图 2-2b所示。如果Y小于其均值时X大于其均值，或者Y大于其均值时X小于其均值，则协方差为负，如图 2-2a所示。协方差的值依赖于X和Y的度量单位。因此我们经常用到相关系数

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{2-5}$$

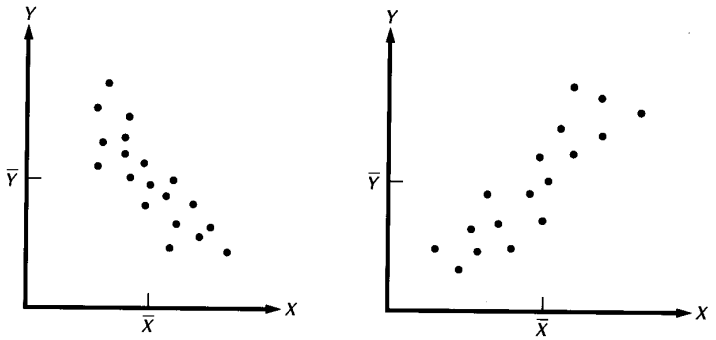
其中σ_x和σ_y分别代表X和Y的标准差。

与协方差不同，相关系数经过了标准化，没有量纲。可以证明相关系数的值总在 -1到+1之间。一个正的相关系数表示变量的运动是同方向的，而负相关系数表示变量的运动是反方向的。

在处理联合概率分布时一些关于期望的性质是很有用的。我们把这些结果陈述如下，其证明请见附录2-1。

结论4 如果X和Y是随机变量，$E(X+Y)=E(X)+E(Y)$

结论5 $\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X,Y)$



a) 协方差为负　　b) 协方差为正

图2-2 协方差

例2.1 协方差和相关系数

某小公司5个雇员的收入(单位：1000美元)和受教育年数的联合概率分布如下：

收入(X)	5	10	15	20	25
受教育年数(Y)	10	8	10	15	12

各变量的均值和方差为：

$$E(X) = (5 + 10 + 15 + 20 + 25)/5 = 15$$
$$\text{Var}(X) = [(-10)^2 + (-5)^2 + 0^2 + 5^2 + 10^2]/5 = 50$$
$$E(Y) = (10 + 8 + 10 + 15 + 12)/5 = 11$$
$$\text{Var}(Y) = [(-1)^2 + (-3)^2 + (-1)^2 + 4^2 + 1^2]/5 = 5.6$$

所以X和Y之间的协方差为：

$$\text{Cov}(X, Y) = [(-10)(-1) + (-5)(-3) + (0)(-1) + (5)(4) + (10)(1)]/5 = 11$$

最后，X和之间的相关系数为

$$\rho(X, Y) = 11/[(50)(5.6)]^{1/2} = 0.66$$

2.1.3 独立与相关

在某些情况下，Y结果发生的概率与X结果发生的概率无关。在这种情况下，我们称X和Y

是独立的随机变量。举个例子，如果抛硬币时正反两面发生的概率均为 1/2，假设前5次结果都是正面，第6次发生反面的概率还是1/2，它与前面发生的结果无关。

当两个变量独立时，关于期望算子的计算就变得简单了。有关结论归纳在结论 6和7中，关于它们的证明请见附录2-1。

结论6 如果X和Y独立，$E(XY)=E(X)\,E(Y)$。

结论7 如果X和Y独立，$\mathrm{Cov}(X, Y)=0$。

结论7说明，如果两个随机变量是独立的，它们的协方差为 0。这个结论很直观，因为X和Y之间的独立意味着一个变量的结果与另一个变量的结果没有关系。这样的话，X与其均值之间的离差和Y与其均值之间的离差也没有关系。然而，必须注意的是，这个结果不可逆，这一点很重要。相关系数为0的两个变量仍可能是不独立的。关键在于方差和相关系数度量的是线性相关性；相关系数为0的变量间可能会具有非线性相关关系。

例如，设X和Y的概率分布如下：

X	-2	-1	0	1	2
Y	4	1	0	1	4

设所有的结果以相等的概率(1/5)发生。在这个例子中，$E(X)=0$，$E(Y)=2$，并且

$$\mathrm{Cov}\,(X, Y) = 1/5 \sum_{i=1}^{5} X_i(Y_i - 2) = 0$$

但是，随机变量X和Y显然不独立。实际上，所有的5个联合分布的可能结果都符合$Y=X^2$的关系，这是恰恰是X与Y间的非线性关系。

2.2 估计

2.2.1 均值、方差和协方差的估计

只有在获得所有可能的结果，即总体时，我们才能够确定均值、方差和协方差。但是通常我们只有关于总体的一个样本，因此我们想要通过样本对总体的特征进行推断。我们将本章中讲述如何取得N个样本数据，对总体特征进行估计，最后得到关于样本估计与相应的总体参数之间关系的结论。由于无法获得均值、方差以及两个随机变量间协方差的真值，我们只能利用样本信息来寻找尽可能好的估计。

我们的目标是要确定一个法则，它能对每个可能的样本给出样本估计值。为了区别具体的估计值和普遍的估计法则，我们把后者称为估计量。对学生来说，混淆“估计值”和“估计量”是很常见的现象，但是如果我们记住估计量是一个法则，而估计值只是一个数，这种混淆就不会发生了。

寻找适用于任何样本的最优估计是一个复杂问题，我们将在 2.3节更详尽地进行讨论。在这里我们只假设最低的要求是参数的估计量所给出的估计值很接近于被估计的参数 (比如均值或者方差)。更具体地说，我们希望估计量是无偏的，即估计量的期望值等于被估计参数。作为一个例子，让我们来重新考虑均值为μ_x的随机变量X的样本均值。估计量$\overline{X}$的定义为：

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{2-6}$$

注意到下面一点十分重要，即$\overline{X}$是一个随机变量，虽然它所对应的总体参数是常数，它的取值将随着样本的不同而不同。由于样本估计值随样本的不同而不同，我们就可以描述出它的概率分布。重复抽取新的样本并每次分别计算样本均值和方差，我们可以得到这个抽样分布。均值

的抽样分布可以度量样本均值落在区间中概率 (回忆前面概率分布中的讨论)。

给定$\overline{X}$的抽样分布，我们很自然要问，估计量$\overline{X}$的期望值是否等于总体均值，换句话说，$\overline{X}$是无偏估计量吗？为了说明$\overline{X}$是无偏的，我们证明$E(\overline{X})=\mu_X$：

$$E(\overline{X}) = E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}E\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N} E(X_i)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu_X = \frac{1}{N}N\mu_X = \mu_X$$

随机变量方差估计量的合理选择是：

$$\frac{1}{N}\sum_{i=1}^{N} (X_i - \overline{X})^2$$

问题是这个估计是有偏的。附录 2-1的结论9给出了一个随机变量方差的无偏估计 (均值未知)：

$$\widehat{\text{Var}}\ (X) = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X})^2 \tag{2-7}$$

为什么我们要除以N-1(而不是N)来获得样本方差的无偏估计呢？真正的答案请见附录 2-1结论9的证明，但是也可以用自由度的概念直观地解释。已知我们的样本包含 N个数据，但在计算样本方差时，第一步就需要计算样本均值，对 N个数据点来说就有了一个约束条件，即 N个观测值之和等于N倍的$\overline{X}$。这样就剩下N-1个无约束的观测值估计样本方差。

最后，我们考虑如何求得两个随机变量之间协方差的一个无偏估计。因为协方差的定义是

$$\text{Cov}\ (X, Y) = E[(X - E(X))(Y - E(Y))]$$

因此协方差可用X，Y与各自均值的离差乘积的平均值来计算，即：

$$\frac{1}{N}\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})$$

但是，与样本方差估计量的情况一样，这个估计是有偏的。为了得到协方差的无偏估计量，我们用自由度除上面的和。在计算 X，Y与均值的离差乘积之和时，共有 N个X和Y的联合观测值，即有N项独立的信息；其中一项信息用来计算 X和Y的样本均值：约束条件为 X和Y的N个观测值之和分别等于N倍的X和Y的均值。因此自由度为N-1，则样本协方差的无偏估计为

$$\widehat{\text{Cov}}\ (X, Y) = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y}) \tag{2-8}$$

样本协方差与协方差真值的区别是在符号 Cov上加了个帽子(^)。

最后，我们可以定义两个随机变量间的样本相关系数，它对应于我们前面定义过的总体相关系数。样本相关系数是

$$r_{XY} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2 \sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \tag{2-9}$$

为了将r_{XY}与其他有关相关性更复杂的度量相区别，我们称它为 X和Y之间的简单相关系数。与总体相关系数相同，样本相关系数的值域为- 1到+1，因此它的平方在 0到1之间。我们可能会注意到X和Y的简单相关系数与样本协方差的直接关系如下：

$$r_{XY} = \frac{\widehat{\text{Cov}}\ (X, Y)}{\sqrt{\widehat{\text{Var}}\ (X)\ \widehat{\text{Var}}\ (Y)}}$$

计量经济学研究变量间的相关关系。由于协方差告诉我们两个变量间是否相关及相关到什

么程度，因此它是应用计量经济学的基础之一。协方差为正意味着当 X大于其均值时， Y也大于均值；X小于其均值， Y也如此。同样(见图2-2a)，协方差为负的一组点的最优拟合直线具有负的斜率。

如果我们把第1章中介绍的斜率的最小二乘估计与样本协方差联系起来，就能更清楚地理解这一点。将样本协方差用离差的形式重新写出，令 $x_i=X_i-\overline{X}$, $y_i=Y_i-\overline{Y}$，则

$$\widehat{\mathrm{Cov}}\ (X,\ Y)\ =\ \frac{1}{N\ -\ 1}\Sigma x_i y_i \tag{2-10}$$

注意，为方便起见我们没有标明求和指标 $i=1,2,\cdots,N$。关于X的方差的估计为：

$$\widehat{\mathrm{Var}}\ (X)\ =\ \frac{1}{N\ -\ 1}\Sigma x_i^2$$

现在考虑样本协方差除以样本方差得到的表达式：

$$\frac{\widehat{\mathrm{Cov}}\ (X,\ Y)}{\widehat{\mathrm{Var}}\ (X)}\ =\ \frac{[1/(N\ -\ 1)]\Sigma x_i y_i}{[1/(N\ -\ 1)]\Sigma x_i^2}\ =\ \frac{\Sigma x_i y_i}{\Sigma x_i^2} \tag{2-11}$$

这个比值等于(1-6)式中斜率的估计值。对于任意一组样本，斜率的最小二乘估计可以由样本协方差与样本方差的比值来计算，其中样本协方差决定直线的方向，而方差是一个正数，用来对用来计算的数据进行标准化。

考虑计算学生平均成绩的例子。斜率的最小二乘估计的计算可用来计算样本均值、样本方差和样本协方差。在数据的离差形式下， $\bar{x}$、$\bar{y}$=0。X与Y的协方差为：

$$\frac{1}{N\ -\ 1}\sum x_i y_i\ =\ \frac{19.50}{7}\ =\ 2.79$$

而X的方差为：

$$\frac{1}{N\ -\ 1}\sum x_i^2\ =\ \frac{162.00}{7}\ =\ 23.14$$

协方差为正，说明斜率为正，样本协方差和样本方差的比值为 2.79/23.14，得出斜率的估计值为0.12。

2.2.2 中心极限定理

当样本容量增大时，均值的抽样分布会发生什么变化？随着样本容量增大，我们会直观地认为均值的估计值总的来说会离总体均值越来越近。实际上如果样本容量非常大或等于总体，样本均值的估计值应当等于总体均值。

这种直观感觉适用于具有有限均值的概率分布，不局限于正态分布。它可以被规范地总结为中心极限定理。

中心极限定理：如果一个随机变量X具有均值μ和方差σ^2，则随着N增大，$\overline{X}$的抽样分布越来越接近于均值为μ方差为σ^2/N的正态分布。

中心极限定理为第2.4节中对正态分布的研究提供了重要根据。我们将看到，对于充分大的样本容量，正态分布的假设将使得我们能够大大地简化统计检验。在研究统计假设之前，我们先简单地讨论一下统计估计量的一些有用的性质。

2.3 估计量的有用性质

我们已经讨论过统计估计量的一个有用的性质是无偏。由于寻找估计量是计量经济学这门学科的核心，我们在这里先来考虑其他一些有用的性质。为了使我们的讨论针对回归模型分析，我们将考虑对任意的参数 β，如直线斜率，我们选择的估计量应当具有什么性质？有四个性质