第二章　经典单方程计量经济学模型：一元线性回归模型

一、内容提要

本章介绍了回归分析的基本思想与基本方法。首先，本章从总体回归模型与总体回归函数、样本回归模型与样本回归函数这两组概念开始，建立了回归分析的基本思想。总体回归函数是对总体变量间关系的定量表述，由总体回归模型在若干基本假设下得到，但它只是建立在理论之上，在现实中只能先从总体中抽取一个样本，获得样本回归函数，并用它对总体回归函数做出统计推断。

本章学习的一个重点是如何获取线性的样本回归函数，主要涉及到普通最小二乘法（OLS）的学习与掌握。同时，也介绍了极大似然估计法（ML）以及矩估计法（MM）。

本章的另一个学习的重点是对样本回归函数能否代表总体回归函数进行统计推断，即进行所谓的统计检验。统计检验包括两个方面，一是先检验样本回归函数与样本点的“拟合优度”，第二是检验样本回归函数与总体回归函数的“接近”程度。后者又包括两个层次：第一，检验解释变量对被解释变量是否存在着显著的线性影响关系，通过变量的 t 检验完成；第二，检验回归函数与总体回归函数的“接近”程度，通过参数估计值的“区间检验”完成。

本章还有三方面的内容不容忽视。其一，若干基本假设。样本回归函数参数的估计、对参数估计量的统计性质的分析以及所进行的统计推断都是建立在这些基本假设之上的。其二，参数估计量统计性质的分析，包括小样本性质与大样本性质，尤其是无偏性、有效性与一致性构成了对样本估计量优劣的最主要的衡量准则。Goss-markov 定理表明 OLS 估计量是最佳线性无偏估计量。其三，运用样本回归函数进行预测，包括被解释变量条件均值与个值的预测、预测置信区间的计算及其变化特征等。

二、典型例题分析

例 1、令 Y 表示一名妇女生育孩子的数目，X 表示该妇女接受过教育的年数。生育率对教育年数的简单回归模型为

$$Y = \beta_0 + \beta_1 X + \mu$$

（1）随机扰动项μ包含什么样的因素？它们可能与教育水平相关吗？

（2）上述简单回归分析能够揭示教育对生育率在其他条件不变下的影响吗？请解释。

解答：

（1）收入、年龄、家庭状况、政府的相关政策等也是影响生育率的重要的因素，在上述简单回归模型中，它们被包含在了随机扰动项之中。有些因素可能与增长率水平相关，如收入水平与教育水平往往呈正相关、年龄大小与教育水平呈负相关等。

（2）当归结在随机扰动项中的重要影响因素与模型中的教育水平X相关时，上述回归模型不能够揭示教育对生育率在其他条件不变下的影响，因为这时出现解释变量与随机扰动项相关的情形，违背了基本假设。

例 2. 已知回归模型$E = \alpha + \beta N + \mu$，式中$E$为某类公司一名新员工的起始薪金（元），$N$为所受教育水平（年）。随机扰动项$\mu$的分布未知，其他所有假设都满足。

（1）从直观及经济角度解释α和β。

（2）OLS 估计量$\hat{\alpha}$和$\hat{\beta}$满足线性性、无偏性及有效性吗？简单陈述理由。

（3）对参数的假设检验还能进行吗？简单陈述理由。

解答：

（1）$\alpha + \beta N$为接受过N年教育的员工的总体平均起始薪金。当N为零时，平均薪金为α，因此α表示没有接受过教育员工的平均起始薪金。β是每单位N变化所引起的E的变化，即表示每多接受一年学校教育所对应的薪金增加值。

（2）OLS 估计量$\hat{\alpha}$和仍$\hat{\beta}$满足线性性、无偏性及有效性，因为这些性质的成立无需随机扰动项μ的正态分布假设。

（3）如果μ的分布未知，则所有的假设检验都是无效的。因为 t 检验与 F 检验是建立在μ的正态分布假设之上的。

例 3、在例 2 中，如果被解释变量新员工起始薪金的计量单位由元改为 100 元，估计的截距项与斜率项有无变化？如果解释变量所受教育水平的度量单位由年改为月，估计的截距项与斜率项有无变化？

解答：

首先考察被解释变量度量单位变化的情形。以E^*表示以百元为度量单位的薪金，则

$$E = E^* \times 100 = \alpha + \beta N + \mu$$

由此有如下新模型

$$E^* = (\alpha / 100) + (\beta / 100) N + (\mu / 100)$$

或 $$E^* = \alpha^* + \beta^* N + \mu^*$$

这里$\alpha^* = \alpha/100$，$\beta^* = \beta/100$。所以新的回归系数将为原始模型回归系数的 1/100。

再考虑解释变量度量单位变化的情形。设N *为用月份表示的新员工受教育的时间长度，则N *=12N，于是

$$E = \alpha + \beta N + \mu = \alpha + \beta(N^*/12) + \mu$$

或 $$E = \alpha + (\beta/12)N^* + \mu$$

可见，估计的截距项不变，而斜率项将为原回归系数的 1/12。

例 4、对没有截距项的一元回归模型

$$Y_i = \beta_1 X_i + \mu_i$$

称之为**过原点回归**（regrission through the origin）。试证明

（1）如果通过相应的样本回归模型可得到通常的正规方程组

$$\sum e_i = 0$$
$$\sum e_i X_i = 0$$

则可以得到β_1的两个不同的估计值：$\tilde{\beta}_1 = \bar{Y}/\bar{X}$，$\hat{\beta}_1 = \left(\sum X_i Y_i\right)/\left(\sum X_i^2\right)$。

（2）在基本假设$E(\mu_i) = 0$下，$\tilde{\beta}_1$与$\hat{\beta}_1$均为无偏估计量。

（3）拟合线$\hat{Y} = \hat{\beta}_1 X$ 通常不会经过均值点$(\bar{X}, \bar{Y})$，但拟合线$\tilde{Y} = \tilde{\beta}_1 X$ 则相反。

（4）只有$\hat{\beta}_1$是β_1的 OLS 估计量。

解答：

（1）由第一个正规方程 $\sum e_t = 0$ 得

$$\sum (Y_t - \tilde{\beta}_1 X_t) = 0$$

或 $$\sum Y_t = \tilde{\beta}_1 \sum X_t$$

求解得 $$\tilde{\beta}_1 = \bar{Y}/\bar{X}$$

由第 2 个正规方程$\sum X_t (Y_t - \hat{\beta}_1 X_t) = 0$得

$$\sum X_t Y_t = \hat{\beta}_1 \sum X_t^2$$

求解得 $$\hat{\beta}_1 = (\sum X_t Y_t)/(\sum X_t^2)$$

（2）对于 $\tilde{\beta}_1 = \overline{Y} / \overline{X}$，求期望

$$E(\tilde{\beta}_1) = E(\overline{Y} \big/ \overline{X}) = \frac{1}{\overline{X}} E[\frac{1}{n}(\beta_1 X_t + \mu_t)]$$
$$= \frac{1}{\overline{X}}[E\{\frac{\beta_1 X_t}{n}) + E(\mu_t)]$$
$$= \frac{\overline{X}}{\overline{X}} \beta_1 = \beta_1$$

这里用到了 X_t 的非随机性。

对于 $\hat{\beta}_1 = (\sum X_t Y_t)/(\sum X_t^2)$，求期望

$$E(\hat{\beta}_1) = E(\sum X_t Y_t / \sum X_t^2)$$
$$= (\frac{1}{\sum X_t^2}) \sum E(X_t Y_t) = (\frac{1}{\sum X_t^2}) \sum E[X_t(\beta_1 X_t + \mu_t)]$$
$$= (\frac{1}{\sum X_t^2}) \beta_1 \sum (X_t^2) + (\frac{1}{\sum X_t^2}) \sum X_t E(\mu_t) = \beta_1$$

（3）要想拟合值 $\hat{Y} = \hat{\beta}_1 X$ 通过点 $(\overline{X}, \overline{Y})$，$\hat{\beta}_1 \overline{X}$ 必须等于 $\overline{Y}$。但 $\hat{\beta}_1 \overline{X} = \dfrac{\sum X_t Y_t}{\sum X_t^2} \overline{X}$，通常不等于 $\overline{Y}$。这就意味着点 $(\overline{X}, \overline{Y})$ 不太可能位于直线 $\hat{Y} = \hat{\beta}_1 X$ 上。

相反地，由于 $\tilde{\beta}_1 \overline{X} = \overline{Y}$，所以直线 $\hat{Y} = \tilde{\beta}_1 X$ 经过点 $(\overline{X}, \overline{Y})$。

（4）OLS 方法要求残差平方和最小

$$\text{Min} \qquad RSS = \sum e_t^2 = \sum (Y_t - \hat{\beta}_1 X_t)^2$$

关于 $\hat{\beta}_1$ 求偏导得

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 2 \sum (Y_t - \hat{\beta}_1 X_t)(-X_t) = 0$$

即

$$\sum X_t (Y_t - \hat{\beta}_1 X_t) = 0$$

$$\hat{\beta}_1 = \left(\sum X_i Y_i\right) \Big/ \left(\sum X_i^2\right)$$

可见 $\hat{\beta}_1$ 是 OLS 估计量。

例 5．假设模型为 $Y_t = \alpha + \beta X_t + \mu_t$。给定 n 个观察值 (X_1, Y_1)，(X_2, Y_2)，…，(X_n, Y_n)，按如下步骤建立 β 的一个估计量：在散点图上把第 1 个点和第 2 个点连接起来

并计算该直线的斜率；同理继续，最终将第 1 个点和最后一个点连接起来并计算该条线的斜率；最后对这些斜率取平均值，称之为 $\hat{\beta}$，即 β 的估计值。

（1）画出散点图，给出 $\hat{\beta}$ 的几何表示并推出代数表达式。

（2）计算 $\hat{\beta}$ 的期望值并对所做假设进行陈述。这个估计值是有偏的还是无偏的？解释理由。

（3）证明为什么该估计值不如我们以前用 OLS 方法所获得的估计值，并做具体解释。

解答：

（1）散点图如图 2-1 所示。

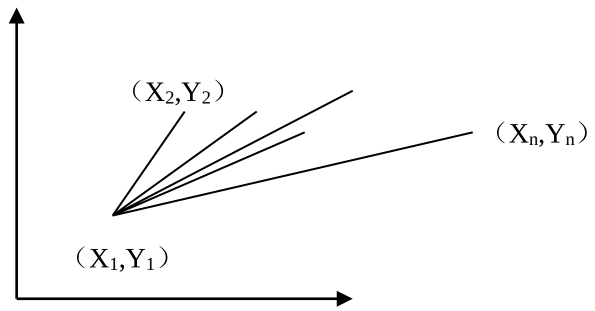

图 2-1

首先计算每条直线的斜率并求平均斜率。连接 (X_1, Y_1) 和 (X_t, Y_t) 的直线斜率为 $(Y_t - Y_1)/(X_t - X_1)$。由于共有 n －1 条这样的直线，因此

$$\hat{\beta} = \frac{1}{n-1}\sum_{t=2}^{n}[\frac{Y_t - Y_1}{X_t - X_1}]$$

（2）因为 X 非随机且 $E(\mu_t) = 0$，因此

$$E[\frac{Y_t - Y_1}{X_t - X_1}] = E[\frac{(\alpha + \beta X_t + \mu_t) - (\alpha + \beta X_1 + \mu_1)}{X_t - X_1}] = \beta + E[\frac{\mu_t - \mu_1}{X_t - X_1}] = \beta$$

这意味着求和中的每一项都有期望值 β，所以平均值也会有同样的期望值，则表明是无偏的。

（3）根据高斯－马尔可夫定理，只有 β 的 OLS 估计量是最佳线性无偏估计量，因此，这里得到的 $\hat{\beta}$ 的有效性不如 β 的 OLS 估计量，所以较差。

例 6. 对于人均存款与人均收入之间的关系式 $S_t = \alpha + \beta Y_t + \mu_t$ 使用美国 36 年的年度数据得如下估计模型，括号内为标准差：

$$\hat{S}_t = 384.105 + 0.067Y_t$$
$$(151.105) \qquad (0.011)$$

$R^2 = 0.538$

（1）β 的经济解释是什么？

（2）α 和 β 的符号是什么？为什么？实际的符号与你的直觉一致吗？如果有冲突的话，你可以给出可能的原因吗？

（3）对于拟合优度你有什么看法吗？

（4）检验是否每一个回归系数都与零显著不同（在 1%水平下）。同时对零假设和备择假设、检验统计值、其分布和自由度以及拒绝零假设的标准进行陈述。你的结论是什么？

解答：

（1）β 为收入的边际储蓄倾向，表示人均收入每增加 1 美元时人均储蓄的预期平均变化量。

（2）由于收入为零时，家庭仍会有支出，可预期零收入时的平均储蓄为负，因此 α 符号应为负。储蓄是收入的一部分，且会随着收入的增加而增加，因此预期 β 的符号为正。实际的回归式中，β 的符号为正，与预期的一致。但截距项为正，与预期不符。这可能是由于模型的错误设定造成的。如家庭的人口数可能影响家庭的储蓄形为，省略该变量将对截距项的估计产生影响；另一种可能就是线性设定可能不正确。

（3）拟合优度刻画解释变量对被解释变量变化的解释能力。模型中 53.8%的拟合优度，表明收入的变化可以解释储蓄中 53.8 %的变动。

（4）检验单个参数采用 t 检验，零假设为参数为零，备择假设为参数不为零。双变量情形下，在零假设下 t 分布的自由度为 n-2=36-2=34。由 t 分布表知，双侧 1%下的临界值位于 2.750 与 2.704 之间。斜率项计算的 t 值为 0.067/0.011=6.09；截距项计算的 t 值为 384.105/151.105=2.54。可见斜率项计算的 t 值大于临界值，截距项小于临界值，因此拒绝斜率项为零的假设，但不拒绝截距项为零的假设。

附录：一些理论结果的证明

1、令 $\hat{\beta}_{YX}$ 和 $\hat{\beta}_{XY}$ 分别为 Y 对 X 回归和 X 对 Y 回归中的斜率，证明

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

其中 r 为 X 与 Y 之相的线性相关系数。

证明：容易知道，在上述两回归中斜率项分别为

$$\hat{\beta}_{YX}=\frac{\sum x_i y_i}{\sum x_i^2}\text{，}\ \hat{\beta}_{XY}=\frac{\sum x_i y_i}{\sum y_i^2}$$

于是

$$\hat{\beta}_{YX}\hat{\beta}_{XY}=\frac{\sum x_i y_i}{\sum x_i^2}\frac{\sum x_i y_i}{\sum y_i^2}=\frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2\sum y_i^2}=r^2$$

2、记样本回归模型为$Y_i=\hat{\beta}_0+\hat{\beta}_1 X_i+e_i$，试证明：

1）估计的Y的均值等于实测的Y的均值：$\bar{\hat{Y}}=\bar{Y}$

2）残差和为零，从而残差的均值为零：$\sum e_i=0$，$\bar{e}=0$

3）残差项与X不相关：$\sum e_i X_i=0$

4）残差项与估计的Y不相关：$\sum e_i\hat{Y}_i=0$；

证明：1）由于

$$\hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1 X_i=(\bar{Y}-\hat{\beta}_1\bar{X})+\hat{\beta}_1 X_i=\bar{Y}+\hat{\beta}_1(X_i-\bar{X})$$

故 $$\bar{\hat{Y}}=\bar{Y}+\hat{\beta}_1\tfrac{1}{n}\sum(X_i-\bar{X})=\bar{Y}$$

这里用到了$\sum x_i=\sum(X_i-\bar{X})=0$

2）由一元回归中正规方程组中的第一个方程

$$\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i)=0$$

知： $$\sum e_i=0\text{，}$$

$$\bar{e}=\tfrac{1}{n}\sum e_i=0$$

3）由一元回归中正规方程组中的第二个方程

$$\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i)X_i=0$$

知： $$\sum e_i X_i=0$$

4）由 2）及 3）易知

$$\sum e_i\hat{Y}_i=\sum e_i(\hat{\beta}_0+\hat{\beta}_1 X_i)=\hat{\beta}_0\sum e_i+\hat{\beta}_1\sum e_i X_i=0$$

3、对一元线性回归模型$Y_i=\beta_0+\beta_1 X_i+\mu_i$，试证明普通最小二乘估计量$\hat{\beta}_1$在所有线性无偏估计量中具有最小方差性。

证：设 $\hat{\beta}_1^*$ 是其他方法得到的关于 β_1 的线性无偏估计量：

$$\hat{\beta}_1^* = \sum c_i Y_i$$

其中，$c_i = k_i + d_i$，d_i 为不全为零的常数，于是

$$E(\hat{\beta}_1^*) = E(\sum c_i Y_i) = \sum c_i E(Y_i) = \sum c_i(\beta_0 + \beta_1 X_i) = \beta_0 \sum c_i + \beta_1 \sum c_i X_i$$

由 $\hat{\beta}_1^*$ 的无偏性，即 $E(\hat{\beta}_1^*) = \beta_1$ 可知：

$$\beta_0 \sum c_i + \beta_1 \sum c_i X_i = \beta_1$$

已知 $\sum c_i = 0$ ， 从而 $\sum c_i X_i = 1$

$\hat{\beta}_1^*$ 的方差 $\operatorname{var}(\hat{\beta}_1^*) = \operatorname{var}(\sum c_i Y_i) = \sum c_i^2 \operatorname{var}(Y_i) = \sum c_i^2 \operatorname{var}(\mu_i) = \sum c_i^2 \sigma^2$

$$= \sum (k_i + d_i)^2 \sigma^2 = \sum k_i^2 \sigma^2 + \sum d_i^2 \sigma^2 + 2\sigma^2 \sum k_i d_i$$

由于 $\sum k_i d_i = \sum k_i (c_i - k_i) = \sum k_i c_i - \sum k_i^2$

$$= \sum \frac{x_i}{\sum x_i^2} c_i - \sum k_i^2 = \frac{\sum X_i c_i - \overline{X} \sum c_i}{\sum x_i^2} - \sum k_i^2 = \frac{1}{\sum x_i^2} - \frac{1}{\sum x_i^2} = 0$$

故 $\operatorname{var}(\hat{\beta}_1^*) = \sum k_i^2 \sigma^2 + \sum d_i^2 \sigma^2 = \frac{1}{\sum x_i^2} \sigma^2 + \sigma^2 \sum d_i^2 = \operatorname{var}(\hat{\beta}_1) + \sigma^2 \sum d_i^2$

因为 $\sum d_i^2 \geq 0$

所以 $\operatorname{var}(\hat{\beta}_1^*) \geq \operatorname{var}(\hat{\beta}_1)$

当 $d_i = 0$，（$i = 1,2\cdots,n$）等号成立，此时，$c_i = k_i$，$\hat{\beta}_1^*$ 就是 OLS 估计量 $\hat{\beta}_1$。

4、 试证明一元线性回归模型随机扰动项 μ 的方差 σ^2 的无偏估计量为 $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$。

证：给定一组样本 $\{X_i, Y_i\}$，容易写出模型 $Y_i = \beta_0 + \beta_1 X_i + \mu_i$ 的离差形式为：

$$y_i = \beta_1 x_i + (\mu_i - \overline{\mu})$$

根据样本回归函数的离差形式：

$$\hat{y}_i = \hat{\beta}_1 x_i$$

易知

$$
\begin{aligned}
\Sigma e_i^2 &= \Sigma(y_i-\hat{y}_i)^2 \\
&= \Sigma((\beta_1-\hat{\beta}_1)x_i+(\mu_i-\overline{\mu}))^2 \\
&= \Sigma((\beta_1-\hat{\beta}_1)^2x_i^2+2(\beta_1-\hat{\beta}_1)x_i(\mu_i-\overline{\mu})+(\mu_i-\overline{\mu})^2) \\
&= \Sigma\beta_1-\hat{\beta}_1)^2x_i^2+\Sigma(\mu_i-\overline{\mu})^2-2\Sigma(\Sigma k_i\mu_i)x_i(\mu_i-\overline{\mu}) \\
&= \Sigma(\beta_1-\hat{\beta}_1)^2x_i^2+\Sigma(\mu_i-\overline{\mu})^2-2\Sigma x_i\mu_i\Sigma k_i\mu_i+2\overline{\mu}\Sigma x_i\Sigma k_i\mu_i \\
&= \Sigma(\beta_1-\hat{\beta}_1)^2x_i^2+\Sigma(\mu_i-\overline{\mu})^2-2\Sigma x_i\mu_i\frac{\Sigma x_i\mu_i}{\Sigma x_i^2}
\end{aligned}
$$

因为

$$
E\sum(\beta_1-\hat{\beta}_1)^2x_i^2=\sum x_i^2\operatorname{var}(\hat{\beta}_1)=\frac{\sum x_i^2\sigma^2}{\sum x_i^2}=\sigma^2
$$

$$
E\sum(\mu_i-\overline{\mu})^2=E(\sum\mu_i^2-2\overline{\mu}\sum\mu_i+n\overline{\mu})=E(\sum\mu_i^2-n\overline{\mu}^2)=(n-1)\sigma^2
$$

$$
E\left[\frac{\left(\sum x_i\mu_i\right)^2}{\sum x_i^2}\right]=E\left[\frac{\sum x_i^2\mu_i^2+2\sum\limits_{i\neq j}(x_{i_i}x_j)(\mu\mu_j)}{\sum x_i^2}\right]=\sigma^2
$$

所以

$$
E(\sum e_i^2)=\sigma^2+(n-1)\sigma^2-2\sigma^2=(n-2)\sigma^2
$$

从而

$$
E(\frac{\Sigma e_i^2}{n-2})=\sigma^2
$$

5、 对一元线性回归模型$Y_i=\beta_0+\beta_1X_i+\mu_i$，试证明$Cov(\hat{\beta}_0,\hat{\beta}_1)=-\dfrac{\sigma^2\overline{X}}{\sum x_i^2}$。

证：

$$
\begin{aligned}
Cov(\hat{\beta}_0,\hat{\beta}_1) &= E(\hat{\beta}_0-\beta_0)(\hat{\beta}_1-\beta_1)=E(\hat{\beta}_0-E(\hat{\beta}_0))(\beta_1-E(\hat{\beta}_1)) \\
&= E(\overline{Y}-\hat{\beta}_1\overline{X}-(\overline{Y}-\overline{X}E(\hat{\beta}_1)))(\beta_1-E(\hat{\beta}_1)) \\
&= -\overline{X}E(\hat{\beta}_1-E(\hat{\beta}_1))(\beta_1-E(\hat{\beta}_1)) \\
&= -\overline{X}E(\hat{\beta}_1-E(\hat{\beta}_1))^2=-\overline{X}\operatorname{var}(\hat{\beta}_1) \\
&= -\frac{\sigma^2\overline{X}}{\sum x_i^2}
\end{aligned}
$$

四、补充练习题

2-1．解释下列概念：

1)总体回归函数

2)样本回归函数

3)随机的总体回归函数

4)线性回归模型

5)随机误差项 μ_i

6)残差项 e_i

7)条件期望

8)回归系数或回归参数

9)回归系数的估计量

10)最小二乘（平方）法

11)最大似然法

12)估计量的标准差

13)总离差平方和

14)回归平方和

15)残差平方和

16)协方差

17)拟合优度检验

18)t 检验

2-2．判断正误并说明理由：

1) 随机误差项 μ_i 和残差项 e_i 是一回事

2) 总体回归函数给出了对应于每一个自变量的因变量的值

3) 线性回归模型意味着变量是线性的

4) 在线性回归模型中，解释变量是原因，被解释变量是结果

5) 随机变量的条件均值与非条件均值是一回事

2-3．回答下列问题：

1) 总体方差与参数估计方差的区别与联系。

2) 随机误差项 μ_i 和残差项 e_i 的区别与联系。

3) 根据最小二乘原理，所估计的模型已经使得拟合误差达到最小，为什么还要讨论模型的拟合优度问题？

4) 为什么用决定系数 R^2 评价拟合优度，而不用残差平方和作为评价标准？

5) 回归分析与相关分析的区别与联系。

6) 最小二乘法和最大似然法的基本原理各是什么？说明它们有何区别？

7) 为什么要进行解释变量的显著性检验？

8) 是否任何两个变量之间的关系，都可以用两变量线性回归模型进行分析？

2-4．表 2-2 列出若干对自变量与因变量。对每一对变量，你认为它们之间的关系如何？是正的、负的、还是无法确定？并说明理由。

表 2-2

因变量	自变量
（1） GNP	利率
（2）个人储蓄	利率
（3）小麦产出	降雨量
（4）美国国防开支	前苏联国防开支
（5）棒球明星本垒打的次数	其年薪
（6）总统声誉	任职时间
（7）学生计量经济学成绩	其统计学成绩
（8）日本汽车的进口量	美国人均国民收入

2-5．参数估计量的无偏性和有效性的含义是什么？从参数估计量的无偏性和有效性证明过程说明，为什么说满足基本假设的计量经济学模型的普通最小二乘参数估计量才具有无偏性和有效性？

2-6．试证明过原点回归模型 $Y_i = \beta_1 X_i + \mu_i$ 中斜率项有

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum X_i^2}$$

2-7. 为什么在一元线性方程中，最小二乘估计量与极大似然估计量的表达式是一致的？证明 σ^2 的 ML 估计量 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ 是有偏的。

2-8．现代投资分析的特征线涉及如下回归方程：$r_t = \beta_0 + \beta_1 r_{mt} + u_t$；其中：$r$ 表示股票或债券的收益率；r_m 表示有价证券的收益率（用市场指数表示，如标准普尔 500 指数）；t 表示时间。在投资分析中，β_1 被称为债券的安全系数 β，是用来度量市场的风险程度的，即市场的发展对公司的财产有何影响。依据 1956~1976 年间 240 个月的数据，Fogler 和 Ganpathy 得到 IBM 股票收益率的回归方程如下：

$$\hat{r}_t = 0.7264 + 1.0598 r_{mt}$$

(0.3001) (0.0728)

$$R^2 = 0.4710$$

要求：（1）解释回归参数的意义；

（2）如何解释 R^2？

（3）安全系数 $\beta>1$ 的证券称为不稳定证券，建立适当的零假设及备选假设，检验 IBM 是否是易变股票（$\alpha=5\%$）。

2-9. 已知模型 $Y_i=\alpha+\beta X_i+\mu_i$，证明：估计量 α 可以表示为：$\hat{\alpha}=\sum_{i=1}^{n}(\frac{1}{n}-\bar{X}W_i)Y_i$，这里 $W_i=\frac{x_i}{\sum x_i^2}$。

2-10. 一个消费分析者论证了消费函数 $C_i=\alpha+\beta Y_i$ 是无用的，因为散点图上的点（C_i，Y_i）不在直线 $C_i=\alpha+\beta Y_i$ 上。他还注意到，有时 Y_i 上升但 C_i 下降。因此他下结论：C_i 不是 Y_i 的函数。请你评价他的论据（这里 C_i 是消费，Y_i 是收入）。

2-11. 证明：仅当 $R^2=1$ 时，Y 对 X 的线性回归的斜率估计量等于 X 对 Y 的线性回归的斜率估计量的倒数。

2-12. 证明：相关系数的另一个表达式是：$r=\hat{\beta}_1\frac{S_x}{S_y}$ 其中 $\hat{\beta}_1$ 为一元线性回归模型一次项系数的估计值，S_x、S_y 分别为 X 与 Y 的样本标准差。

2-13. 对于经济计量模型：$Y_i=\beta_0+\beta_1X_i+\mu_i$，其 OLS 估计参数 β_1 的特性在下列情况下会受到什么影响：

（1）观测值数目 n 增加；

（2）X_i 各观测值差额增加；

（3）X_i 各观测值近似相等；

（4）$\mathrm{E}(\mu^2)=0$。

2-14. 假定有如下的回归结果：$\hat{Y}_t=2.6911-0.4795X_t$，其中，$Y$ 表示美国的咖啡的消费量（每天每人消费的杯数），X 表示咖啡的零售价格（美元/杯），t 表示时间。问

（1）这是一个时间序列回归还是横截面序列回归？做出回归线；

（2）如何解释截距的意义，它有经济含义吗？如何解释斜率？

（3）能否求出真实的总体回归函数？

（4）根据需求的价格弹性定义：弹性＝斜率×（X/Y），依据上述回归结果，你能求出

对咖啡需求的价格弹性吗？如果不能，计算此弹性还需要其他什么信息？

2-15．假设某人通过一容量为 19 的样本估计了如下消费函数 $C_i = \alpha + \beta Y_i + \mu_i$，并获得下列结果：

$$\hat{C}_i = 15 + 0.81 Y_i$$

（3.1）　(18.7)

$R^2 = 0.98$

要求：（1）利用 t 值检验假设：$\beta = 0$（取显著水平为 5%）;

（2）确定参数估计量的标准差；

（3）构造 β 的 95%的置信区间，这个区间包括 0 吗？

2-16．表 2-3 给出了某社区每月家庭的收入 X 与消费支出 Y 的调查数据。

表 2-3

每月收入 X（元）	每月消费支出 Y（元）
800	550，600，650，700，750
1000	650，700，740，800，850，880
1200	790，840，900，940，980
1400	800，930，950，1030，1080，1130，1150
1600	1020，1070，1100，1160，1180，1250
1800	1100，1150，1200，1300，1350，1400
2000	1200，1360，1400，1440，1450
2200	1350，1370，1400，1520，1570，1600，1620
2400	1370，1450，1550，1650，1750，1890
2600	1500，1520，1750，1780，1800，1850，1910

要求：（1）对每一收入水平，计算平均的消费支出，$E(Y \mid X_i)$，即条件期望值；

（2）以收入为横轴、消费支出为纵轴作散点图；在散点图中，做出（1）中的条件均值点；你认为 X 与 Y 之间、X 与 Y 的均值之间的关系如何？

（3）写出其总体回归函数。

（4）如果对每一个 X 值，随机抽取一个 Y 值，结果如下：

Y	700	650	900	950	1100	1150	1200	1400	1550	1500
X	800	1000	1200	1400	1600	1800	2000	2200	2400	2600

求样本回归函数。

（5）在同一个图中，做出总体回归线与样本回归线，它们相同吗？

2-17．表 2-4 给出了某国 1990~1996 年间的 CPI 指数与 S&P500 指数。要求：

（1）以 CPI 指数为横轴、S&P 指数为纵轴做图；

（2）你认为 CPI 指数与 S&P 指数之间关系如何？

（3）考虑下面的回归模型：$S\&P_t = \beta_1 + \beta_2 CPI_t + \mu_t$，根据表中的数据运用 OLS 估计上述方程，并解释所得结果。

表 2-4

年份	CPI	S&P500 指	年份	CPI	S&P500 指数
1990	130.7	334.59	1994	148.2	460.33
1991	136.2	376.18	1995	152.4	541.64
1992	140.3	415.74	1996	159.6	670.83
1993	144.5	451.41			

2-18．表 2-5 给出了美国 30 所知名学校的 MBA 学生 1994 年基本年薪（ASP）、GPA 分数（从 1~4 共四个等级）、GMAT 分数以及每年学费（X）的数据。要求：

（1）用双变量回归模型分析 GPA 是否对 ASP 有影响？

（2）用合适的回归模型分析 GMAT 分数是否与 ASP 有关？

（3）每年的学费与 ASP 有关吗？你是如何知道的？如果两变量之间正相关，是否意味着进到最高费用的商业学校是有利的；

（4）你同意高学费的商业学校意味着高质量的 MBA 成绩吗？为什么？

表 2-5

学校	ASP (美元)	GPA	GMAT	X (美元)
Harvard	102630	3.4	650	23894
Stanford	100800	3.3	665	21189
Columbian	100480	3.3	640	21400
Dartmouth	95410	3.4	660	21225
Wharton	89930	3.4	650	21050
Northwestern	84640	3.3	640	20634
Chicago	83210	3.3	650	21656
MIT	80500	3.5	650	21690
Virginia	74280	3.2	643	17839
UCLA	74010	3.5	640	14496
Berkeley	71970	3.2	647	14361
Cornell	71970	3.2	630	20400
NUY	70660	3.2	630	20276
Duke	70490	3.3	623	21910
Carnegie Mellon	59890	3.2	635	20600
North Carolina	69880	3.2	621	10132

Michigan	67820	3.2	630	20960
Texas	61890	3.3	625	8580
Indiana	58520	3.2	615	14036
Purdue	54720	3.2	581	9556
Case Western	57200	3.1	591	17600
Georgetown	69830	3.2	619	19584
Michigan State	41820	3.2	590	16057
Penn State	49120	3.2	580	11400
Southern Methodist	60910	3.1	600	18034
Tulane	44080	3.1	600	19550
Illinois	47130	3.2	616	12628
Lowa	41620	3.2	590	9361
Minnesota	48250	3.2	600	12618
Washington	44140	3.3	617	11436

2-19．表 2-6 给出了 1988 年 9 个工业国的名义利率（Y）与通货膨胀率（X）的数据。要求：（1）以利率为纵轴、通货膨胀率为横轴做图；

（2）用 OSL 进行回归分析；

（3）如果实际利率不变，则名义利率与通货膨胀率的关系如何？

表 2-6

国家	Y（%）	X（%）	国家	Y（%）	X（%）
澳大利亚	11.9	7.7	墨西哥	66.3	51.0
加拿大	9.4	4.0	瑞典	2.2	2.0
法国	7.5	3.1	英国	10.3	6.8
德国	4.0	1.6	美国	7.6	4.4
意大利	11.3	4.8			

补充练习参考答案

2-1．答：

⑴总体回归函数是指在给定 X_i 下 Y 分布的总体均值与 X_i 所形成的函数关系（或者说：将总体被解释变量的条件期望表示为解释变量的某种函数)。

⑵样本回归函数指从总体中抽出的关于 Y 、 X 的若干组值形成的样本而建立的回归函数。

⑶随机的总体回归函数指含有随机误差项的总体回归函数(是相对与条件期望形式而言

的）。

⑷教材中所讲的线性回归模型既指对变量是线性的，也指对参数β为线性的。即解释变量与参数β只以它们的 1 次方出现。

⑸随机误差项也称误差项，是一个随机变量，针对总体回归函数而言。

⑹残差项是一随机变量，针对样本回归函数而言。

⑺条件期望又称条件均值，指X取特定X_i值时的Y的期望值。

⑻回归系数（或回归参数）指回归模型中β_0、β_1等未知但却是固定的参数。

⑼回归系数的估计量指用$\hat{\beta}_0$、$\hat{\beta}_1$等表示的用已知样本提供的信息所估计出来总体未知参数的结果。

⑽最小二乘法又称最小平方法，指使估计的剩余平方和最小的原则确定样本回归函数的方法。

⑾最大似然法又称最大或然法，指用产生该样本概率最大的原则去确定样本回归函数的方法。

⑿估计量的标准差是度量一个变量变化大小的测度量。

⒀总离差平方和用 TSS 表示，用以度量被解释变量的总变动。

⒁回归平方和用 ESS 表示，用以度量由解释变量变化引起的被解释变量的变化部分。

⒂残差平方和用 RSS 表示，用以度量实际值与拟合值之间的差异，是由除解释变量以外的其他因素引起的被解释变量变化的部分。

⒃协方差用$Cov(X,Y)$表示，是用来度量 X、Y 二个变量关联程度的统计量。

⒄拟和优度检验指检验模型对样本观测值的拟合程度，用R^2表示，该值越接近 1，模型对样本观测值拟合得越好。

⒅t 检验是针对每个解释变量进行的显著性检验，即构造一个 t 统计量，如果该统计量的值落在置信区间外，就拒绝原假设。

2-2．答：

（1）错误；随机误差项是针对总体回归模型而言的，它是模型中其他没有包含的因素的综合体；而残差项是针对样本回归模型而言的，它是实际观测值与样本回归线上值的离差。两者的含义不同，后者只能说成是对前者的一个估计。

（2）错误；总体回归函数给出了对应于第一个自变量的被解释变量的均值。

（3）（在不考虑参数非线性的情况下是）正确的。

（4）正确；这是建立回归模型的前提。

（5）错误；只有在解释变量与随机误差项不相关时，随机误差项与条件均值与非条件均值才是一回事。在基本假设成立的情况下，两者是一回事。

2-3．答：

（1）总体方差又称随机误差项的方差，用$Var(\mu_i|X_i)=\sigma^2$表示。它是参数估计量方差的有机组成部分。如在一元线性回归模型$Y_i=\beta_0+\beta_1X_i+\mu_i$中，

$$Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_i^2}，\ Var(\hat{\beta}_0)=\frac{\sum X_i^2}{n\sum x_i^2}\sigma^2，$$

（2）随机误差项μ_i是指总体观测值与回归方程理论值之间的偏差，而残差项e_i是指样本观测值与回归方程理论值之间的偏差，二者是有区别的；但是，由于总体观察值无法得到，从而造成总体回归函数事实上是未知的，因此，一般的做法是通过样本观测获得的信息去估计总体回归函数，这样，残差e_i就是随机误差项μ_i的一个样本估计量。

（3）普通最小二乘法所保证的最好拟合是同一个问题内部的比较，即使用给出的样本数据满足残差的平方和最小；拟合优度检验结果所表示的优劣可以对不同的问题进行比较，即可以辨别不同的样本回归结果谁好谁坏。

（4）判定系数$R^2=\dfrac{ESS}{TSS}=1-\dfrac{RSS}{TSS}$，含义为由解释变量引起的被解释变量的变化占被解释变量总变化的比重，用来判定回归直线拟合的优劣。该值越大说明拟合得越好；而残差平方和与样本容量关系密切，当样本容量比较小时，残差平方和的值也比较小；尤其是不同样本得到的残差平方和是不能做比较的。此外，作为检验统计量的一般应是相对量而不能用绝对量，因而不能使用残差平方和判断模型的拟合优度。

（5）回归分析是讨论被解释变量与一个或多个解释变量之间具体依存关系的分析方法；相关分析是讨论变量之间线性相关程度的分析方法；二者的区别在于：研究的目的不同，相关分析着重探讨变量间的关联程度，而回归分析却要进一步探寻变量间具体依赖关系，即希望根据解释变量的固定值去估计和预测被解释变量的平均值；对变量的处理不同，相关分析对称地处理相互联系的变量，而回归分析必须明确解释变量与被解释变量。二者的联系在于：回归分析建立在相关分析基础之上，当相互有关联的变量进一步有因果关系时，可进一步进

行回归分析。相关分析中线性相关系数的平方等于回归分析中的拟合优度。

（6）最小二乘法和最大似然法都是常用于对线性回归模型参数进行估计的方法。最小二乘法的基本原理是：用使估计的剩余平方和最小的原则确定样本回归函数；最大似然法的基本原理是：用产生该样本概率最大的原则去确定样本回归函数。它们的区别在于：最小二乘法的估计量具有线性、无偏性与有效性，随机误差项方差估计量也是无偏的；而最大似然法的估计量仅具有线性、无偏性、有效性，其随机误差项方差的估计量是有偏的。

（7）对解释变量进行显著性检验的目的，是为了决定该变量是否应作为解释变量被保留在模型中，如果该变量对被解释变量的影响并不显著，就应该将其剔除，并寻找其他可能的变量建立模型。

（8）不是。当变量间存在非线性关系时，可建立非线性回归模型。

2-4．答：

（1）无法确定；

（2）正的因果关系；

（3）因果关系但不能确定正负；

（4）正的因果关系；

（5）正的相关关系；

（6）无法确定；

（7）正的因果关系；

（8）正的相关关系。

2-5．答：参数估计量的无偏性是指：参数估计量 $\hat{\beta}$ 的均值等于模型参数值，即 $E(\hat{\beta}_0)=\beta_0$，$E(\hat{\beta}_1)=\beta_1$。参数估计量的有效性是指：在所有线性、无偏估计量中，该参数估计量的方差最小。从参数估计量的无偏性和有效性的证明过程中看出，得出无偏性、有效性是利用了随机误差项具有零均值和同方差及随机误差项与解释变量之间不相关的基本假设，所以说只有满足基本假设的 OLS 参数估计量才具有无偏性和有效性。

2-6．证明：

易知，模型 $Y_i=\beta_1 X_i+\mu_i$ 的参数的 OLS 估计量为：

$$\hat{\beta}_1=\frac{\sum X_i Y_i}{\sum X_i^2}=\frac{\sum X_i(\beta_1 X_i+\mu_i)}{\sum X_i^2}=\beta_1+\frac{\sum X_i \mu_i}{\sum X_i^2}$$

因此，

$$Var(\hat{\beta}_1) = Var(\beta + \frac{\sum X_i \mu_i}{\sum X_i^2})$$

$$= \sum \left(\frac{X_i}{\sum X_i^2} \right)^2 Var(\mu_i) = \frac{\sigma^2}{\sum X_i^2}$$

2-7．答：

在一元线性方程中，最小二乘估计量与最大似然估计量的表达式之所以会一致，原因在于两种方法都存在对$\sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$式求极小的步骤。

证明最大似然估计量　$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ 是有偏的过程如下：

对 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ 式两边取数学期望，并考虑 $E(\frac{\sum e_i^2}{n-2}) = \sigma^2$ 得

$$E(\hat{\sigma}^2) = \frac{1}{n} E(\sum_{i=1}^{n} e_i^2)$$
$$= (\frac{n-2}{n})\sigma^2$$
$$= \sigma^2 - \frac{2}{n}\sigma^2$$

结果表明，在小样本下，$\hat{\sigma}^2$ 偏小，使得估计值比真实值低，是有偏的；在样本无限增大时，$\frac{2}{n} \to 0$，从而使 $\lim E(\hat{\sigma}^2) = \sigma^2$，即 $\hat{\sigma}^2$ 是渐近无偏的。

2-8．答：

⑴回归方程的截距 0.7264 表明，当 r_m 为 0 时的股票或债券收益率，它本身没有经济意义；回归方程的斜率 1.0598 表明当有价证券的收益率每上升(或下降)1 个点将使股票或债券收益率上升(或下降)1.0598 个点；

⑵ R^2 为判定系数，是度量回归方程拟合优度的指标，它表明该回归方程中 47.1%的股票或债券收益率的变化是由 r_m 的变化引起的。当然 $R^2 = 0.4710$ 也表明回归方程对数据的拟合效果不是很好。

⑶建立零假设 $H_0: \beta_1 = 1$，备择假设 $H_1: \beta_1 > 1$，$\alpha = 0.05$，$n = 240$，

查表得临界值 $t_{0.05}(238) = 1.645$，由于

$$t=\frac{\hat{\beta}_1-1}{S_{\hat{\beta}_1}}=\frac{1.0598-1}{0.0728}=0.8214$$

故 $0.8214<1.645$，接受零假设 $H_0:\beta_1=1$，拒绝备择假设 $H_1:\beta_1>1$。说明此期间 IBM 不是易变证券。

2-9．证明：

$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}=\sum_{i=1}^{n}\frac{Y_i}{n}-\bar{X}\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$$

由于

$$\sum_{i=1}^{n}x_iy_i=\sum_{i=1}^{n}x_i(Y_i-\bar{Y})=\sum_{i=1}^{n}x_iY_i-\bar{Y}\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_iY_i$$

于是

$$\hat{\alpha}=\sum_{i=1}^{n}\frac{Y_i}{n}-\frac{\bar{X}}{\sum_{i=1}^{n}x_i^2}\sum_{i=1}^{n}x_iY_i=\sum_{i=1}^{n}(\frac{1}{n}-\bar{X}\frac{x_i}{\sum_{i=1}^{n}x_i^2})Y_i=\sum(\frac{1}{n}-\bar{X}W_i)Y_i$$

2-10．答：

他的论据是错误的。原因是他忽略了随机误差项 μ_i，这个随机误差项可取正值和负值，但是 $\mathrm{E}(\mu_i)=0$。另外，将 C_i 与 Y_i 的关系表达为 $C_i=\alpha+\beta Y_i$ 是不准确的，它应是一个平均关系。

2-11．证明：

设：$\hat{Y}_i=\hat{\alpha}_0+\hat{\alpha}_1X_i$，

$\hat{X}_i=\hat{\beta}_0+\hat{\beta}_1Y_i$，

则 $\hat{\alpha}_1=\frac{\sum x_iy_i}{\sum x_i^2}$，$\hat{\beta}_1=\frac{\sum x_iy_i}{\sum y_i^2}$

于是 $\hat{\alpha}_1\hat{\beta}_1=\frac{(\sum x_iy_i)^2}{\sum x_i^2\sum y_i^2}=R^2=1$

从而 $\hat{\alpha}_1=\frac{1}{\hat{\beta}_1}$

2-12．证明：

由于　$\hat{\beta}=\dfrac{\sum xy}{\sum x^2}$

又　$S_x=\sqrt{\dfrac{\sum x^2}{n-1}}$，　$S_y=\sqrt{\dfrac{\sum y^2}{n-1}}$

所以　$\hat{\beta}\dfrac{S_x}{S_y}=\dfrac{\sum xy}{\sum x^2}\cdot\dfrac{\sqrt{\dfrac{\sum x^2}{n-1}}}{\sqrt{\dfrac{\sum y^2}{n-1}}}=\dfrac{\sum xy}{\sqrt{x^2y^2}}=r$

2-13．解：由 $\hat{\beta}_1=\dfrac{\sum xy}{\sum x^2}=\beta_1+\dfrac{\sum x\mu}{\sum x^2}=\beta_1+\dfrac{\sum x\mu/n}{\sum x^2/n}$ 知：

⑴随着观测值数目的增加，根据大样本性质，参数 $\hat{\beta}_1$ 更接近真实值；

⑵由于 X_i 各观测值差额增加，意味着 $\sum x^2$ 增加，将使得 $\hat{\beta}_1$ 更接近真实值；

⑶如果 X_i 各观测值近似相等，意味着 $\sum x^2$ 趋于零，会使得 $\hat{\beta}_1$ 变得很不稳定，甚至无法计算；

⑷ $E(\mu^2)=0$，并违背随机误差项同方差的性质，所以既不会对 $\hat{\beta}_1$ 的无偏性产生影响，也不会对有效性产生影响。

2-14．解：

⑴这是一个时间序列回归（图略）。

⑵截距 2.6911 表示咖啡零售价在 t 时刻为每磅 0 美元时，美国平均消费量为每天每人 2.6911 杯，这个数字没有明显的经济意义；斜率-0.4795 表示咖啡零售价与消费量负相关，表明咖啡价格每上升 1 美元，则平均每天每人消费量减少 0.4795 杯，即约半杯；

⑶不能；原因在于要了解全美国所有人的咖啡消费情况几乎是不可能的；

⑷不能；在同一条需求曲线上不同点的价格弹性不同，若要求出，须给出具体的 X 值及与之对应的 Y 值。

2-15．解：

⑴由于参数估计量 β 的 t 的绝对值为 18.7，明显大于 2，故拒绝零假设 $H_0:\beta=0$，从而 β 在统计上是显著的；

⑵参数 α 的估计量的标准差为 15/3.1=4.84，参数 β 的估计量的标准差为 0.81/18.7=0.043；

⑶由⑵的结果，β 的 95%的置信区间为：

$$\begin{aligned}&(0.81 - t_{0.025}(19-2)\times 0.043, \quad 0.81 + t_{0.025}(19-2)\times 0.043)\\&=(0.81-0.091, \quad 0.81+0.091)\\&=(0.719, 0.901)\end{aligned}$$

显然这个区间不包括 0。

2-16．解：

(1)每月家庭收入水平下，消费支出的条件均值计算见表 2-7。

表 2-7

	每月家庭收入X（元）									
	800	1000	1200	1400	1600	1800	2000	2200	2400	2600
每月家庭消费支出Y（元）	550	650	790	800	1020	1100	1200	1350	1370	1500
	600	700	840	930	1070	1150	1360	1370	1450	1520
	650	740	900	950	1100	1200	1400	1400	1550	1750
	700	800	940	1030	1160	1300	1440	1520	1650	1780
	750	850	980	1080	1180	1350	1450	1570	1750	1800
		880		1130	1250	1400		1600	1890	1850
				1150				1620		1910
条件概率	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
条件均值	650	770	890	1010	1130	1250	1370	1490	1610	1730

⑵以收入为横轴、消费支出为纵轴作散点图如图 2-3 所示。其中条件均值点显示为“°”的点，这些点近似形成一条直线（虚线表示）。显然，X 与 Y 之间是正向相关关系；X 与 Y 的均值间的关系近似线性。

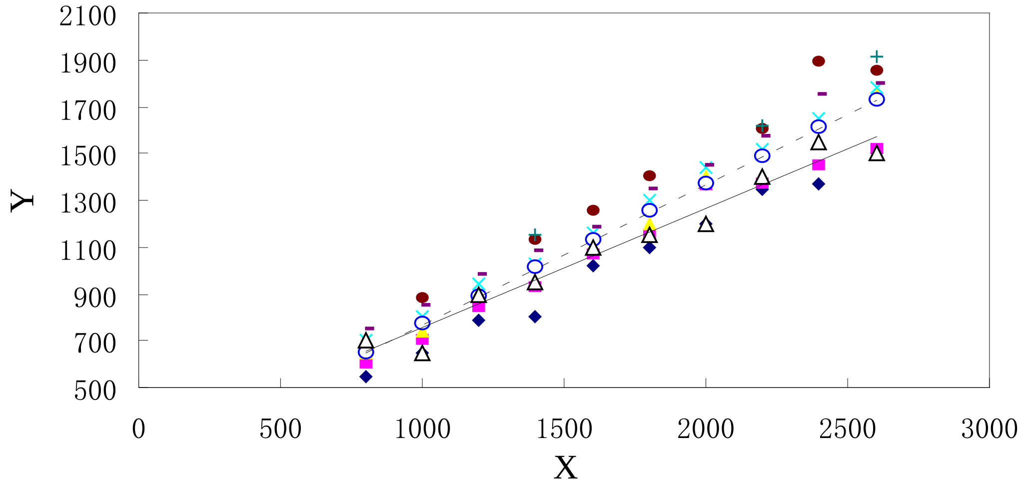

图 2-3

（3）由于X与Y的均值间的关系呈现线性关系，可以估计总体回归函数如下。任取两点（800，650），（2600，1730），写出总体回归方程：

$$\frac{E(Y \mid X)-650}{1730-650}=\frac{X-800}{2600-800}$$

整理得

$$E(Y \mid X)=170+0.6X$$

（4）在 Eviews 软件下，容易得到样本回归函数为

$$\hat{Y}_i = 244.6+0.5091X_i$$

（3.81） (14.24)

R^2=0.9620

(5)图中同时画出了总体回归线（虚线）与样本回归线（实线），可以看出两者并不相同。原因是存在随机误差。

2-17．解：

⑴利用所给数据作图 2-4。

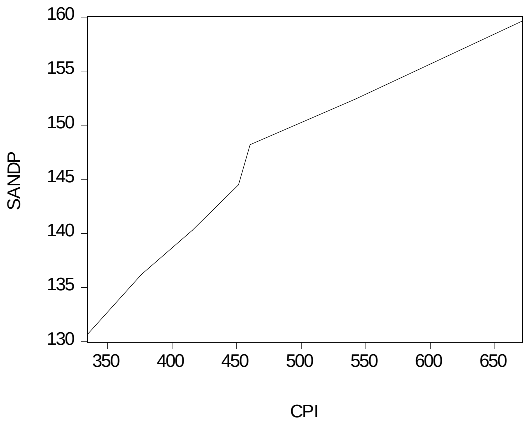

图2-4

⑵从图 2-4 可见，CPI 指数与 S&P 指数正相关，且呈近似的线性关系。

⑶使用 Eviews 软件回归结果如下：

Dependent Variable: S&P				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1137.826	177.9488	-6.394122	0.0014
CPI	11.08361	1.228555	9.021662	0.0003
R-squared	0.942123	Mean dependent var		464.3886
Adjusted R-squared	0.930548	S.D. dependent var		112.3728
S.E. of regression	29.61448	Akaike info criterion		9.849360
Sum squared resid	4385.086	Schwarz criterion		9.833906
Log likelihood	-32.47276	F-statistic		81.39039
Durbin-Watson stat	1.187041	Prob(F-statistic)		0.000279

回归结果显示，CPI 指数与 S&P 指数正相关，斜率表示当 CPI 变化 1 个点，会使 S&P 变化 11.08 个点；截距表示当 CPI 为 0 时，S&P 为-1137.8264，此数据没有明确的经济意义。

2-18．解：

⑴使用 Eviews 软件，ASP 对 GPA 的回归结果如下：

Dependent Variable: ASP				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-270566.7	82444.69	-3.281797	0.0028
GPA	104041.0	25302.40	4.111900	0.0003
R-squared	0.376499	Mean dependent var		68260.00
Adjusted R-squared	0.354231	S.D. dependent var		18187.78
S.E. of regression	14615.65	Akaike info criterion		22.08191
Sum squared resid	5.98E+09	Schwarz criterion		22.17533
Log likelihood	-329.2287	F-statistic		16.90772
Durbin-Watson stat	1.027123	Prob(F-statistic)		0.000311

从回归结果看，GPA 的系数是统计显著的，对 ASP 有正的影响。

⑵使用 Eviews 软件，ASP 对 GMAT 的回归结果如下：

Dependent Variable: ASP				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-332306.8	47572.09	-6.985332	0.0000
GMAT	641.6598	76.15036	8.426222	0.0000
R-squared	0.717175	Mean dependent var		68260.00
Adjusted R-squared	0.707074	S.D. dependent var		18187.78
S.E. of regression	9843.701	Akaike info criterion		21.29139
Sum squared resid	2.71E+09	Schwarz criterion		21.38480
Log likelihood	-317.3709	F-statistic		71.00122
Durbin-Watson stat	1.128809	Prob(F-statistic)		0.000000

从回归结果看，GMAT 与 ASP 显著正相关。

⑶使用 Eviews 软件，ASP 对学费（X）的回归结果如下：

Dependent Variable: ASP				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23126.32	9780.863	2.364446	0.0252
X	2.633483	0.551601	4.774252	0.0001
R-squared	0.448748	Mean dependent var		68260.00
Adjusted R-squared	0.429061	S.D. dependent var		18187.78
S.E. of regression	13742.78	Akaike info criterion		21.95876
Sum squared resid	5.29E+09	Schwarz criterion		22.05217
Log likelihood	-327.3813	F-statistic		22.79348
Durbin-Watson stat	1.142178	Prob(F-statistic)		0.000051

从计算结果看，每年的学费与 ASP 显著正相关。学费高，ASP 就高；但学费仅解释了 ASP 变化的一部分（不到 50%），明显还有其他因素影响着 ASP。

⑷使用 Eviews 软件回归结果如下：

Dependent Variable: GPA				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.147579	0.072559	43.37936	0.0000
X	6.17E-06	4.09E-06	1.507952	0.1428

R-squared	0.075112	Mean dependent var	3.253333
Adjusted R-squared	0.042080	S.D. dependent var	0.104166
S.E. of regression	0.101951	Akaike info criterion	-1.664311
Sum squared resid	0.291032	Schwarz criterion	-1.570897
Log likelihood	26.96466	F-statistic	2.273920
Durbin-Watson stat	1.702758	Prob(F-statistic)	0.142768

从回归结果看，尽管高学费的商业学校与高质量的 MBA 成绩略有正相关性，但学费对 GAP 的影响是不显著的。而且也无法得出学费是影响 GAP 的主要原因的结论。

2-19．解：

⑴以利率（Y）为纵轴、通胀率（X）为横轴作图 2-5。

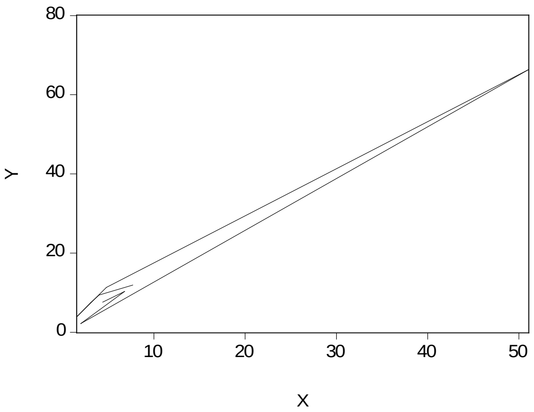

图 2-5

(2)Eviews 软件中的 OLS 回归如下：

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.636174	0.691303	3.813340	0.0066
X	1.250286	0.039326	31.79274	0.0000

R-squared	0.993122	Mean dependent var	14.50000
Adjusted R-squared	0.992140	S.D. dependent var	19.69150
S.E. of regression	1.745810	Akaike info criterion	4.145445
Sum squared resid	21.33498	Schwarz criterion	4.189272
Log likelihood	-16.65450	F-statistic	1010.778
Durbin-Watson stat	1.819376	Prob(F-statistic)	0.000000

(3)上述回归结果表明，如果实际利率不变，名义利率与通胀率呈正向关系；斜率 1.2503 表明通胀率上升 1 个点，名义利率上升 1.25 个点。