

(亿元)。

$$(2) C_t = 180 + 1.2Y_t$$

其中， C 、 Y 分别是城镇居民消费支出和可支配收入。

$$(3) \ln Y_t = 1.15 + 1.62 \ln K_t - 0.28 \ln L_t$$

其中， Y 、 K 、 L 分别是工业总产值、工业生产资金和职工人数。

1-19. 下列假想的计量经济模型是否合理，为什么？

$$(1) \text{GDP} = \alpha + \sum \beta_i \text{GDP}_i + \varepsilon$$

其中， $\text{GDP}_i (i = 1, 2, 3)$ 是第 i 产业的国内生产总值。

$$(2) S_1 = \alpha + \beta S_2 + \varepsilon$$

其中， S_1 、 S_2 分别为农村居民和城镇居民年末储蓄存款余额。

$$(3) Y_t = \alpha + \beta_1 I_t + \beta_2 L_t + \varepsilon$$

其中， Y 、 I 、 L 分别为建筑业产值、建筑业固定资产投资和职工人数。

$$(4) Y_t = \alpha + \beta P_t + \varepsilon$$

其中， Y 、 P 分别为居民耐用消费品支出和耐用消费品物价指数。

$$(5) \text{财政收入} = f(\text{财政支出}) + \varepsilon$$

$$(6) \text{煤炭产量} = f(L, K, X_1, X_2) + \varepsilon$$

其中， L 、 K 分别为煤炭工业职工人数和固定资产原值， X_1 、 X_2 分别为发电量和钢铁产量。

1-20. 模型参数对模型有什么意义？

习题参考答案

第一章 绪论

1-1. 什么是计量经济学？

答：计量经济学是经济学的一个分支学科，是以揭示经济活动中客观存在的数量关系为内容的分支学科，是由经济学、统计学和数学三者结合而成的交叉学科。

1-2. 答：计量经济学自 20 年代末、30 年代初形成以来，无论在技术方法还是在应用方面发展都十分迅速，尤其是经过 50 年代的发展阶段和 60 年代的扩张阶段，使其在经济学科占据重要的地位，主要表现在：①在西方大多数大学和学院中，计量经济学的讲授已成为经济学课程表中有权威的一部分；②从 1969~2003 年诺贝尔经济学奖的 XX 位获奖者中有 XX 位是与研究和应用计量经济学有关；著名经济学家、诺贝尔经济学奖获得者萨缪尔森甚至说：“第二次世界大战后的经济学是计量经济学的时代”。③计量经济学方法与其他经济数学方法结合应用得到发展；④计量经济学方法从主要用于经济预测转向经济理论假设和政策假设的检验；⑤计量经济学模型的应用从传统的领域转向新的领域，如货币、工资、就业、福利、国际贸易等；⑥计量经济学模型的规模不再是水平高低的衡量标准，人们更喜欢建立一些简单的模型，从总量上、趋势上说明经济现象。

1-3. 答：计量经济学方法揭示经济活动中各个因素之间的定量关系，用随机性的数学方程加以描述；一般经济数学方法揭示经济活动中各个因素之间的理论关系，用确定性的数学方程加以描述。

1-4. 答：

1-5. 答：从计量经济学的定义看，它是量化的经济学；其次，从计量经济学在西方国家经济学科中居于最重要的地位看，也是如此，尤其是从诺贝尔经济学奖设立之日起，已有多人因直接或间接对计量经济学的创立和发展作出贡献而获得诺贝尔经济学奖；计量经济学与数理统计学有严格的区别，它仅限于经济领域；从建立与应用计量经济学模型的全过程看，不论是理论模型的设定还是样本数据的收集，都必须以对经济理论、对所研究的经济现象有透彻的认识为基础。综上所述，计量经济学确实是一门经济学科。

1-6. 计量经济学的研究的对象和内容是什么？计量经济学模型研究的经济关系有哪两个基本特征？

答：计量经济学的研究对象是经济现象，是研究经济现象中的具体数量规律（或者说，计量经济学是利用数学方法，根据统计测定的经济数据，对反映经济现象本质的经济数量关系进行研究）。计量经济学的内容大致包括两个方面：一是方法论，即计量经济学方法或理论计量经济学；二是应用，即应用计量经济学；无论是理论计量经济学还是应用计量经济学，都包括理论、方法和数据三种要素。

计量经济学模型研究的经济关系有两个基本特征：一是随机关系；二是因果关系。

1-7. 答：

1-8. 答：计量经济学方法，就是定量分析经济现象中各因素之间的因果关系。所以，第一步，要根据经济理论分析所研究的经济现象，找出经济现象之间的因果关系及相互间的联系，把问题作为被解释变量，把影响问题的主要因素作为解释变量，把非主要因素归入随机项；第二步，要按照它们之间的行为关系选择适当的数学形式描述这些变量之间的关系，一般是用一组数学上彼此独立、互不矛盾、完整有解的方程组表示。在建立理论模型的时，要求理论模型在参数估计、模型检验的过程中不断得到修正，以便得到一个较好的、能够解释过去的、反映客观经济规律的数学模型。此外，还可以通过散电图或模拟的方法，选择一个拟合效果较好的数学模型。

1-9. 答：计量经济学模型主要有以下几个方面的用途：①结构分析，即研究一个或几个经济变量发生变化及结构参数的变动对其他变量以至整个经济系统产生何种的影响；其原理是弹性分析、乘数分析与比较静力分析。②经济预测，即用其进行中短期经济的因果预测；其原理是模拟历史，从已经发生的经济活动中找出变化规律；③政策评价，即利用计量经济模型定量分析政策变量变化对经济系统运行的影响，是对不同政策执行情况的“模拟仿真”。④检验与发展经济理论，即利用计量经济模型和实际统计资料实证分析某个理论假说的正确与否；其原理是如果按照某种经济理论建立的计量经济模型可以很好地拟合实际观察数据，则意味着该理论是符合客观事实的，否则则表明该理论不能说明客观事实。

1-10. 答：时间序列数据的例子如：改革开放以来 25 年中的 GDP、居民人均消费支出、人均可支配收入、零售物价指数、固定资产投资等；横截面数据的例子如：2003 年各省的 GDP、该年各工业部门的销售额、该年不同收入的城镇居民消费支出、该年不同城镇居民的可支配收入、该年各省的固定资产投资等。这两类数据都是反映经济规律的经济现象的数量信息，不同点：时间序列数据是含义、口径相同的同一指标按时间先后排列的统计数据列；而横截面数据是一批发生在同一时间截面上不同统计单元的相同统计指标组成的数据列。

1-11. 答：如果模型系统只包含一个方程，即只研究单一的经济活动过程，揭示其因素之间

的单向因果关系，则称该模型为单方程模型；如果模型系统涉及到多个经济关系而需要构造一个方程组，则称该模型为联立方程模型。二者之间有着密切联系，如：单方程模型是联立方程模型的组成元素，而联立方程模型又是由若干个单方程模型有机组合而成。二者又有区别，如：单方程模型都是随机方程，而联立方程模型中既有随机方程也有恒等方程。

1-12. 答：模型的检验主要包括：经济意义检验、统计检验、计量经济学检验、模型的预测检验。在经济意义检验中，需要检验模型是否符合经济意义，检验求得的参数估计值的符号与大小是否与根据人们的经验和经济理论所拟订的期望值相符合；在统计检验中，需要检验模型参数估计值的可靠性，即检验模型的统计学性质；在计量经济学检验中，需要检验模型的计量经济学性质，包括随机扰动项的序列相关检验、异方差性检验、解释变量的多重共线性检验等；模型的预测检验主要检验模型参数估计量的稳定性以及对样本容量变化时的灵敏度，以确定所建立的模型是否可以用于样本观测值以外的范围。

1-13. 答：常用的样本数据包括：时间序列数据、横截面数据、虚变量数据和面板数据。

1-14. 答：由于客观经济现象的复杂性，以至于人们目前仍难以完全地透彻地了解它的全貌。对于某一种经济现象而言，往往受到很多因素的影响，而人们在认识这种经济现象的时候，只能从影响它的很多因素中选择一种或若干种来说明。这样就会有許多因素未被选上，这些未被选上的因素必然也会影响所研究的经济现象。因此，由被选因素构成的数学模型与由全部因素构成的数学模型去描述同一经济现象，必然会有出入。为使模型更加确切地说明客观经济现象，所以有必要引入随机误差项。随机误差项形成的原因：①在解释变量中被忽略的因素；②变量观测值的观测误差；③模型的关系误差或设定误差；④其他随机因素的影响。

1-15. 答：

1-16. 答：经济数据是通过对经济变量进行观测和统计得到的，它们反映经济活动相关方面的水平和情况。从计量经济学的角度看，经济数据是计量经济分析的材料，或者说发现经济规律的信息载体，对经济规律的实证研究起十分关键的作用。为此，要求经济数据须具备完整性、准确性、可比性和一致性。

1-17.

1-18.

1-19.

1-20.

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以及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。Gauss-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，以及预测置信区间的计算及其变化特征。

二、典型例题分析

例 1、令 $kids$ 表示一名妇女生育孩子的数目， $educ$ 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

$$kids = \beta_0 + \beta_1 educ + \mu$$

- (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 educ 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，基本假设 4 不满足。

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 μ 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_t 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 (regression 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

$$\begin{aligned} \sum e_i &= 0 \\ \sum e_i X_i &= 0 \end{aligned}$$

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

(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 $\sum e_i = 0$ 得

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

或
$$\sum Y_i = \tilde{\beta}_1 \sum X_i$$

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

由第 2 个下规方程 $\sum X_t(Y_t - \hat{\beta}_1 X_t) = 0$ 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 $\tilde{\beta}_1 = \bar{Y} / \bar{X}$, 求期望

$$\begin{aligned} E(\tilde{\beta}_1) &= E(\bar{Y} / \bar{X}) = \frac{1}{\bar{X}} E[\frac{1}{n}(\beta_1 X_t + \mu_t)] \\ &= \frac{1}{\bar{X}} [E\{\frac{\beta_1 X_t}{n}\} + E(\mu_t)] \\ &= \frac{\bar{X}}{\bar{X}} \beta_1 = \beta_1 \end{aligned}$$

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

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

(3) 要想拟合值 $\hat{Y} = \hat{\beta}_1 X$ 通过点 $(\bar{X}, \bar{Y})$, $\hat{\beta}_1 \bar{X}$ 必须等于 $\bar{Y}$ 。但 $\hat{\beta}_1 \bar{X} = \frac{\sum X_t Y_t}{\sum X_t^2} \bar{X}$,

通常不等于 $\bar{Y}$ 。这就意味着点 $(\bar{X}, \bar{Y})$ 不太可能位于直线 $\hat{Y} = \hat{\beta}_1 X$ 上。

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

(4) OLS 方法要求残差平方和最小

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

关于 $\hat{\beta}_1$ 求偏导得

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

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

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

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) , $\dots$, (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。

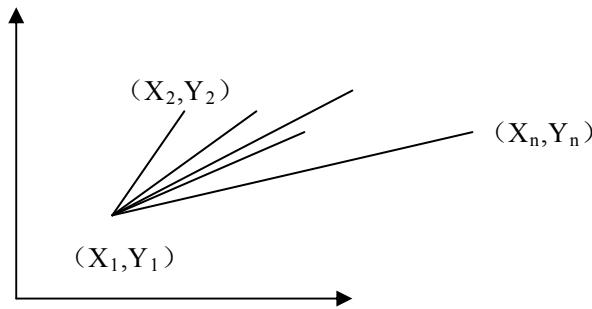
(1) 画出散点图, 给出 $\hat{\beta}$ 的几何表示并推出代数表达式。

(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) 散点图如下图所示。



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

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

(2) 因为 X 非随机且 $E(\mu_t) = 0$, 因此

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

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 估计量是最付佳线性无偏估计量, 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：

$$\hat{S}_t = 384.105 + 0.067Y_t$$

$$(151.105) \quad (0.011)$$

$$R^2 = 0.538 \quad \hat{\sigma} = 199.023$$

(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。但截距项为负，与预期不符。这可能与由于模型的错误设定形造成的。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。

三、习题

(一) 基本知识类题型

2-1. 解释下列概念:

- | | |
|-----------------------------------|-------------|
| 1) 总体回归函数 | 11) 最大似然法 |
| 2) 样本回归函数 | 12) 估计量的标准差 |
| 3) 随机的总体回归函数 | 13) 总离差平方和 |
| 4) 线性回归模型 | 14) 回归平方和 |
| 5) 随机误差项 (u_i) 和残差项 (e_i) | 15) 残差平方和 |
| 6) 条件期望 | 16) 协方差 |
| 7) 非条件期望 | 17) 拟合优度检验 |
| 8) 回归系数或回归参数 | 18) t 检验 |
| 9) 回归系数的估计量 | 19) F 检验 |
| 10) 最小平方法 | |

2-2. 判断正误并说明理由:

- 1) 随机误差项 u_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) 线性回归模型有哪些基本假设? 违背基本假设的计量经济学模型是否就不可估计?
- 2) 总体方差与参数估计误差的区别与联系。
- 3) 随机误差项 u_i 和残差项 e_i 的区别与联系。
- 4) 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?
- 5) 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?
- 6) R^2 检验与 F 检验的区别与联系。
- 7) 回归分析与相关分析的区别与联系。

- 8) 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？
- 9) 为什么要进行解释变量的显著性检验？
- 10) 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？

2-2. 下列方程哪些是正确的？哪些是错误的？为什么？

- (1) $y_t = \alpha + \beta x_t \quad t = 1, 2, \dots, n$
- (2) $y_t = \alpha + \beta x_t + \mu_t \quad t = 1, 2, \dots, n$
- (3) $y_t = \hat{\alpha} + \hat{\beta} x_t + \mu_t \quad t = 1, 2, \dots, n$
- (4) $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t + \mu_t \quad t = 1, 2, \dots, n$
- (5) $y_t = \hat{\alpha} + \hat{\beta} x_t \quad t = 1, 2, \dots, n$
- (6) $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t \quad t = 1, 2, \dots, n$
- (7) $y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\mu}_t \quad t = 1, 2, \dots, n$
- (8) $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\mu}_t \quad t = 1, 2, \dots, n$

其中带“^”者表示“估计值”。

2-3. 下表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。

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

(二) 基本证明与问答类题型

2-4. 对于一元线性回归模型，试证明：

- (1) $E(y_i) = \alpha + \beta x_i$
- (2) $D(y_i) = \sigma^2$

$$(3) \text{Cov}(y_i, y_j) = 0 \quad i \neq j$$

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. 对于过原点回归模型 $Y_i = \beta_1 X_i + u_i$ ，试证明

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

2-7. 试证明：

$$(1) \sum e_i = 0, \text{ 从而: } \bar{e} = 0$$

$$(2) \sum e_i x_i = 0$$

$$(3) \sum e_i \hat{Y}_i = 0; \text{ 即残差 } e_i \text{ 与 } Y_i \text{ 的估计值之积的和为零。}$$

2-8. 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估计量为 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i^2$ ，并且是有偏的。

2-9. 熟悉 t 统计量的计算方法和查表判断。

2-10. 证明： $R^2 = (r_{yx})^2$ ；其中 R^2 是一元线性回归模型的判定系数， r_{yx} 是 y 与 x 的相关系数。

2-11. 试根据置信区间的概念解释 t 检验的概率意义，即证明：对于显著性水平 α ，当 $|t_i| > t_{\alpha/2}$

时， b_i 的 $100(1-\alpha)\%$ 的置信区间不包含 0。

2-12. 线性回归模型

$$y_t = \alpha + \beta x_t + \mu_t \quad t = 1, 2, \dots, n$$

的 0 均值假设是否可以表示为 $\frac{1}{n} \sum_{t=1}^n \mu_t = 0$ ？为什么？

2-13. 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 股票的回归方程；市场指数是在芝加哥大学建立的市场有价证券指数：

$$\hat{r}_t = 0.7264 + 1.0598r_{mt} \quad r^2 = 0.4710$$

$$(0.3001) \quad (0.0728)$$

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，并用t检验进行检验（ $\alpha = 5\%$ ）。

2-14. 已知模型 $Y_i = \alpha + \beta x_i + u_i$ ，证明：估计量 α 可以表示为： $\hat{\alpha} = \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}W_i\right)y_i$ 这

$$\text{里 } W_i = \frac{\bar{x}_i}{\sum x_i^2}$$

2-15. 已知两个量X和Y的一组观察值 (x_i, y_i) ， $i=1, 2, \dots, n$ 。

证明：Y的真实值和拟合值有共同的均值。

2-16. 一个消费分析者论证了消费函数 $C_i = a + bY_i$ 是无用的，因为散点图上的点 (C_i, Y_i)

不在直线 $C_i = a + bY_i$ 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）。

2-17. 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。

2-18. 证明：相关系数的另一个表达式是： $r = \hat{\beta} \frac{S_x}{S_y}$ 其中 $\hat{\beta}$ 为一元线性回归模型一次项

系数的估计值， S_x 、 S_y 分别为样本标准差。

2-19. 对于经济计量模型： $Y_i = b_0 + b_1 X_i + u_i$ ，其OLS估计参数 b_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） $E(u^2) = 0$ 。

2-20. 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。

要求：

- （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 $\times$ (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？

(三) 基本计算类题型

2-21. 下面数据是对 X 和 Y 的观察值得到的。

$$\sum Y_i = 1110; \quad \sum X_i = 1680; \quad \sum X_i Y_i = 204200$$

$$\sum X_i^2 = 315400; \quad \sum Y_i^2 = 133300$$

假定满足所有的古典线性回归模型的假设，要求：(1) b_1 和 b_2 ？(2) b_1 和 b_2 的标准差？(3) r^2 ？(4) 对 B_1 、 B_2 分别建立 95% 的置信区间？利用置信区间法，你可以接受零假设： $B_2=0$ 吗？

2-22. 假设王先生估计消费函数（用模型 $C_i = a + bY_i + u_i$ 表示），并获得下列结果：

$$\hat{C}_i = 15 + 0.81Y_i, \quad n=19$$

(3.1) (18.7) $R^2=0.98$ 这里括号里的数字表示相应参数的 T 比率值。

要求：(1) 利用 T 比率值检验假设： $b=0$ （取显著水平为 5%）；(2) 确定参数估计量的标准方差；(3) 构造 b 的 95% 的置信区间，这个区间包括 0 吗？

2-23. 下表给出了每周家庭的消费支出 Y（美元）与每周的家庭的收入 X（美元）的数据。

每周收入 (X)	每周消费支出 (Y)
80	55, 60, 65, 70, 75
100	65, 70, 74, 80, 85, 88
120	79, 84, 90, 94, 98
140	80, 93, 95, 103, 108, 113, 115
160	102, 107, 110, 116, 118, 125
180	110, 115, 120, 130, 135, 140
200	120, 136, 140, 144, 145
220	135, 137, 140, 152, 157, 160, 162
240	137, 145, 155, 165, 175, 189
260	150, 152, 175, 178, 180, 185, 191

要求：

- (1) 对每一收入水平，计算平均的消费支出， $E(Y | X_i)$ ，即条件期望值；
- (2) 以收入为横轴、消费支出为纵轴作散点图；
- (3) 在散点图中，做出 (1) 中的条件均值点；
- (4) 你认为 X 与 Y 之间、X 与 Y 的均值之间的关系如何？
- (5) 写出其总体回归函数及样本回归函数；总体回归函数是线性的还是非线性的？

2-24. 根据上题中给出的数据，对每一个 X 值，随机抽取一个 Y 值，结果如下：

Y	70	65	90	95	110	115	120	140	155	150
X	80	100	120	140	160	180	200	220	240	260

要求：

- (1) 以 Y 为纵轴、X 为横轴作图，并说明 Y 与 X 之间是怎样的关系？
- (2) 求样本回归函数，并按要求写出计算步骤；
- (3) 在同一个图中，做出样本回归函数及从上题中得到的总体回归函数；比较二者相同吗？为什么？

2-25. 下表给出了 1990~1996 年间的 CPI 指数与 S&P500 指数。

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

资料来源：总统经济报告，1997，CPI 指数见表 B-60，第 380 页；S&P 指数见表 B-93，第 406 页。

- 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： $(S \& P)_t = B_1 + B_2CPI_t + u_t$ ，根据表中的数据运用 OLS

估计上述方程，并解释你的结果；你的结果有经济意义吗？

2-26. 下表给出了美国 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 分数以及每年学费的数据。

学校	ASP/美元	GPA	GMAT	学费/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

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？

2-27. 从某工业部门抽取 10 个生产单位进行调查，得到下表所列的数据：

单位序号	年产量（万吨）y	工作人员数（千人）x
1	210.8	7.062
2	210.1	7.031
3	211.5	7.018
4	208.9	6.991
5	207.4	6.974
6	205.3	7.953
7	198.8	6.927
8	192.1	6.302
9	183.2	6.021
10	176.8	5.310

要求：假定年产量与工作人员数之间存在线性关系，试用经典回归估计该工业部门的生产函

数及边际劳动生产率。

2-28. 下表给出了 1988 年 9 个工业国的名义利率（Y）与通货膨胀率（X）的数据：

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

资料来源：原始数据来自国际货币基金组织出版的《国际金融统计》

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，写出求解步骤；
- （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？
- （四）自我综合练习类题型

2-29. 综合练习：自己选择研究对象，收集样本数据（利用我国公开发表的统计资料），应用计量经济学软件（建议使用 Eviews3.1）完成建立计量经济学模型的全过程，并写出详细的研究报告。（通过练习，能够熟练应用计量经济学软件 Eviews3.1 中的最小二乘法）

四、习题参考答案

2-1. 答：

- (1) 总体回归函数是指在给定 X_i 下的 Y 的分布的总体均值与 X_i 有函数关系。
- (2) 样本回归函数指对应于某个给定的 X 的 Y 值的一个样本而建立的回归函数。
- (3) 随机的总体回归函数指含有随机误差项的总体回归函数，形如：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

(4) 线性回归模型指对参数 β 为线性的回归，即 β 只以它的 1 次方出现，对 X 可以是或不是线性的。

- (5) 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。

(6)残差项是一随机变量，针对样本回归函数而言。

(7)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。

(9)回归系数（或回归参数）指 β_1 、 β_2 等未知但却是固定的参数。

(10)回归系数的估计量指用 $\hat{\beta}_1$ 、 $\hat{\beta}_2$ 等表示的用已知样本所提供的信息去估计出来的量。

(13)估计量的标准差指度量一个变量变化大小的标准。

(14)总离差平方和用 TSS 表示，用以度量被解释变量的总变动。

(15)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。

(16)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。

(17)协方差用 $\text{Cov}(X, Y)$ 表示，是用来度量 X 、 Y 二个变量同时变化的统计量。

2-2. 答：错；错；错；错；错。（理由见本章其他习题答案）

2-3. 答：

(1)线性回归模型的基本假设（实际是针对普通最小二乘法的基本假设）是：解释变量是确定性变量，而且解释变量之间互不相关；随机误差项具有 0 均值和同方差；随机误差项在不同样本点之间是独立的，不存在序列相关；随机误差项与解释变量之间不相关；随机误差项服从 0 均值、同方差的正态分布。违背基本假设的计量经济学模型还是可以估计的，只是不能使用普通最小二乘法进行估计。

(5)判定系数 $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ ，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。

(10)不是。

2-8. 证明：

由于 $\hat{\beta}_1 = \frac{\sum X_t Y_t}{\sum X_t^2}$ ，因此

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var}\left(\frac{\sum X_t Y_t}{\sum X_t^2}\right) = \text{Var}\left(\sum \frac{X_t}{X_t^2} Y_t\right) = \sum \left(\frac{X_t}{X_t^2}\right)^2 \text{Var}(\beta_1 X_t + \mu_t) \\ &= \sum \frac{X_t^2}{(\sum X_t^2)^2} \text{Var}(\mu_t) = \sigma_\mu^2 \frac{\sum X_t^2}{(\sum X_t^2)^2} = \frac{\sigma_\mu^2}{\sum X_t^2} \end{aligned}$$

2-9. 证明：

(1)根据定义得知,

$$\begin{aligned}\sum e_i &= \sum (Y_i - \hat{Y}_i) = \sum (Y_i - \beta_1 - \beta_2 X_i) = \sum Y_i - n\beta_1 - \beta_2 \sum X_i \\ &= n\bar{Y} - n\beta_1 - n\beta_2 \bar{X} = n(\bar{Y} - \beta_1 - \beta_2 \bar{X})\end{aligned}$$

$$\therefore \bar{Y} = \beta_1 + \beta_2 \bar{X}$$

$$\therefore \sum e_i = 0$$

$$\text{从而使得: } \bar{e} = \frac{\sum e_i}{n} = 0$$

证毕。

(2)

$$\begin{aligned}\therefore \sum e_i X_i &= \sum (Y_i - \hat{Y}_i)(X_i - \bar{X}) = \sum (Y_i X_i - \bar{X} Y_i - X_i \hat{Y}_i + \bar{X} \hat{Y}_i) \\ &= \sum [Y_i X_i - \bar{X} Y_i - (Y_i - e_i) X_i + \bar{X} (Y_i - e_i)] \\ &= \sum (Y_i X_i - \bar{X} Y_i - Y_i X_i + e_i X_i + \bar{X} Y_i - e_i \bar{X}) \\ &= \sum (e_i X_i - e_i \bar{X}) \\ &= \sum e_i \bar{X} (n-1) \\ &= 0 \\ \therefore \sum e_i X_i &= 0\end{aligned}$$

证毕。

(3)

$$\begin{aligned}\sum e_i \hat{Y}_i &= \sum e_i (\beta_1 + \beta_2 X_i) = \beta_1 \sum e_i + \beta_2 \sum e_i X_i \\ &= \beta_1 \sum e_i + n\bar{X} \beta_2 \sum e_i \\ &= 0\end{aligned}$$

证毕。

2-14. 答: 线性回归模型: $y_t = \alpha + \beta x_t + \mu_t$ 中的 0 均值假设 $E(u^2) = 0$ 不可以表示为:

$$\frac{1}{n} \sum_{t=1}^n \mu_t = 0, \text{ 因为前者表示取完所有的可能的样本组合后的平均状态, 而后者只是一个样本的}$$

平均值。

2-16. 证明:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \sum_{i=1}^n \frac{y_i}{n} - \bar{x} \frac{\sum_{i=1}^n \dot{x}_i \dot{y}_i}{\sum_{i=1}^n \dot{x}_i^2}$$

$$\because \sum_{i=1}^n \dot{x}_i \dot{y}_i = \sum_{i=1}^n \dot{x}_i (y_i - \bar{y}) = \sum_{i=1}^n \dot{x}_i y_i - \bar{y} \sum_{i=1}^n \dot{x}_i = \sum_{i=1}^n \dot{x}_i y_i$$

$$\hat{\alpha} = \sum_{i=1}^n \frac{y_i}{n} - \frac{\bar{x}}{\sum_{i=1}^n \dot{x}_i^2} \sum_{i=1}^n \dot{x}_i y_i = \sum_{i=1}^n \left(\frac{1}{n} - \frac{\bar{x}}{\sum_{i=1}^n \dot{x}_i^2} \right) y_i$$

证毕。

2-17. 证明:

$$\because \hat{\alpha} \text{ 和 } \hat{\beta} \text{ 满足正规方程 } \sum_{i=1}^n [y_i - (\hat{\alpha} + \hat{\beta}x_i)] = 0$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$$

$$\therefore \sum_{i=1}^n (y_i - \hat{y}_i) = 0 \text{ 即表明 } Y \text{ 的真实值与拟合值有共同的均值。}$$

证毕。

2-18. 答: 他的论据是错误的。原因是他忽略了随机误差项 u_i , 这个随机误差项可取正值和负值, 但是 $E(u_i) = 0$, 将 C_i 与 Y_i 的关系表达为 $C_i = \alpha + \beta Y_i$ 是不准确的, 而是一个平均关系。

2-19. 证明:

$$\text{设: } \hat{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 x_i,$$

$$\hat{x}_i = \hat{\beta}_0 + \hat{\beta}_1 y_i,$$

$$\text{由于: } R^2 = \frac{(\sum \dot{x}_i \dot{y}_i)^2}{\sum \dot{x}_i^2 \sum \dot{y}_i^2} = 1 \Rightarrow \sum \dot{x}_i^2 = \frac{(\sum \dot{x}_i \dot{y}_i)^2}{\sum \dot{y}_i^2}$$

$$\text{线性回归的斜率估计量: } \hat{\alpha}_1 = \frac{\sum \dot{x}_i \dot{y}_i}{\sum \dot{x}_i^2} = \frac{1}{\sum (\dot{x}_i \dot{y}_i) / \sum \dot{y}_i^2} = \frac{1}{\hat{\beta}_1}$$

证毕。

2-20. 证明:

$$\therefore \hat{\beta} = \frac{\sum \dot{x}\dot{y}}{\sum \dot{x}^2} \quad \text{又} \therefore S_x = \sqrt{\frac{\sum \dot{x}^2}{n-1}}, \quad S_y = \sqrt{\frac{\sum \dot{y}^2}{n-1}}$$

$$\therefore \hat{\beta} \frac{S_x}{S_y} = \frac{\sum \dot{x}\dot{y}}{\sum \dot{x}^2} \cdot \frac{\sqrt{\frac{\sum \dot{x}^2}{n-1}}}{\sqrt{\frac{\sum \dot{y}^2}{n-1}}} = \frac{\sum \dot{x}\dot{y}}{\sqrt{\sum \dot{x}^2 \sum \dot{y}^2}} = r$$

证毕。

2-22. 解:

(1)这是一个横截面序列回归。(图略)

(2)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, 这个数字没有经济意义; 斜率 -0.4795 表示咖啡零售价与消费量负相关, 在 t 时刻, 价格上升 1 美元/磅, 则平均每天每人消费量减少 0.4795 杯;

(3)不能;

(4)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。

2-23. 解:

$$(1) \therefore \bar{X} = \frac{\sum X_i}{n} = 168, \quad \bar{Y} = \frac{\sum Y_i}{n} = 111$$

$$\begin{aligned} \therefore \sum (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum (X_i Y_i - \bar{Y} X_i - Y_i \bar{X} + \bar{X} \bar{Y}) \\ &= 204200 - 1680 \times 111 - 168 \times 1110 + 10 \times 168 \times 111 \\ &= 17720 \end{aligned}$$

$$\begin{aligned} \text{又} \therefore \sum (X_i - \bar{X})^2 &= \sum (X_i^2 - 2X_i \bar{X} + \bar{X}^2) \\ &= \sum X_i^2 - 2 \times 10 \bar{X}^2 + 10 \bar{X}^2 \\ &= 315400 - 10 \times 168 \times 168 \\ &= 33160 \end{aligned}$$

$$\therefore \beta_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{17720}{33160} = 0.5344$$

$$\beta_1 = \bar{Y} - \beta_2 \bar{X} = 111 - 0.5344 \times 168 = 21.22$$

$$(2) \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{10-2} = \frac{\sum (Y_i^2 - 2Y_i \hat{Y}_i + \hat{Y}_i^2)}{8}$$

$$\therefore \hat{Y}_i = 21.22 + 0.5344X_i$$

$$\begin{aligned} \therefore \sum (Y_i^2 - 2Y_i\hat{Y}_i + \hat{Y}_i^2) &= \sum (Y_i^2 - 2 \times 21.22Y_i - 2 \times 0.5344X_iY_i + \beta_1^2 + \beta_2^2X_i^2 + 2\beta_1\beta_2X_i) \\ &= 133300 - 2 \times 21.22 \times 1110 - 2 \times 0.5344 \times 204200 + 10 \times 21.22 \times 21.22 \\ &\quad + 0.5344 \times 0.5344 \times 315400 + 2 \times 21.22 \times 0.5344 \times 1680 \\ &= 620.81 \end{aligned}$$

$$\therefore \hat{\sigma}^2 = \frac{620.81}{8} = 77.60$$

$$\therefore \text{Var}(\beta_1) = \frac{\sum X_i^2 \sigma^2}{n \sum (X_i - \bar{X})^2} = \frac{77.60 \times 315400}{10 \times 33160} = 73.81, \quad \text{se}(\beta_1) = \sqrt{73.81} = 8.5913$$

$$\text{Var}(\beta_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{77.60}{33160} = 0.0023, \quad \text{se}(\beta_2) = \sqrt{0.0023} = 0.0484$$

$$^{(3)} r^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2},$$

$$\therefore \sum e_i^2 = 620.81,$$

$$\text{又} \therefore \sum (Y_i - \bar{Y})^2 = 133300 - 123210 = 10090$$

$$\therefore r^2 = 1 - \frac{620.81}{10090} = 0.9385$$

$$(4) \therefore p(|t| \leq 2.306) = 95\%, \quad \text{自由度为 } 8$$

$$\therefore -2.306 \leq \frac{21.22 - \beta_1}{8.5913} \leq 2.306, \quad \text{解得: } 1.4085 \leq \beta_1 \leq 41.0315 \text{ 为 } \beta_1 \text{ 的 } 95\% \text{ 的置信区间。}$$

$$\text{同理, } \therefore -2.306 \leq \frac{0.5344 - \beta_2}{0.0484} \leq 2.306, \quad \text{解得: } 0.4227 \leq \beta_2 \leq 0.646 \text{ 为 } \beta_2 \text{ 的 } 95\% \text{ 的置信区间。}$$

信区间。

由于 $\beta_2 = 0$ 不在 β_2 的置信区间内，故拒绝零假设： $\beta_2 = 0$ 。

2-24. 解：

(1) 由于参数估计量 β 的 T 比率值的绝对值为 18.7 且明显大于 2，故拒绝零假设

$H_0: \beta = 0$ ，从而 β 在统计上是显著的；

(2) 参数 α 的估计量的标准方差为 $15/3.1=4.84$ ，参数 β 的估计量的标准方差为 $0.81/18.7=0.043$ ；

(3) 由(2)的结果， β 的 95% 的置信区间为：

$(0.81 - t_{0.975}(n-2)0.043, 0.81 + t_{0.975}(n-2)0.043) = (0.81 - 0.091, 0.81 + 0.091)$, 显然这个区间不包括 0。

2-25. 解:

$$(1) E(Y|X_i = 80) = 65$$

$$E(Y|X_i = 100) = 77$$

$$E(Y|X_i = 120) = 89$$

$$E(Y|X_i = 140) = 101$$

$$E(Y|X_i = 160) = 113$$

$$E(Y|X_i = 180) = 125$$

$$E(Y|X_i = 200) = 137$$

$$E(Y|X_i = 220) = 149$$

$$E(Y|X_i = 240) = 161$$

$$E(Y|X_i = 260) = 173$$

第三章、经典单方程计量经济学模型：多元线性回归模型

一、内容提要

本章将一元回归模型拓展到了多元回归模型，其基本的建模思想与建模方法与一元的情形相同。主要内容仍然包括模型的基本假定、模型的估计、模型的检验以及模型在预测方面的应用等方面。只不过为了多元建模的需要，在基本假设方面以及检验方面有所扩充。

本章仍重点介绍了多元线性回归模型的基本假设、估计方法以及检验程序。与一元回归分析相比，多元回归分析的基本假设中引入了多个解释变量间不存在（完全）多重共线性这一假设；在检验部分，一方面引入了修正的可决系数，另一方面引入了对多个解释变量是否对被解释变量有显著线性影响关系的联合性 F 检验，并讨论了 F 检验与拟合优度检验的内在联系。

本章的另一个重点是将线性回归模型拓展到非线性回归模型，主要学习非线性模型如何转化为线性回归模型的常见类型与方法。这里需要注意各回归参数的具体经济含义。

本章第三个学习重点是关于模型的约束性检验问题，包括参数的线性约束与非线性约束检验。参数的线性约束检验包括对参数线性约束的检验、对模型增加或减少解释变量的检验以及参数的稳定性检验三方面的内容，其中参数稳定性检验又包括邹氏参数稳定性检验与邹氏预测检验两种类型的检验。检验都是以 F 检验为主要检验工具，以受约束模型与无约束模型是否有显著差异为检验基点。参数的非线性约束检验主要包括最大似然比检验、沃尔德检验与拉格朗日乘数检验。它们仍以估计无约束模型与受约束模型为基础，但以最大似然原理进行估计，且都适用于大样本情形，都以约束条件个数为自由度的 χ^2 分布为检验统计量的分布特征。非线性约束检验中的拉格朗日乘数检验在后面的章节中多次使用。

二、典型例题分析

例 1. 某地区通过一个样本容量为 722 的调查数据得到劳动力受教育的一个回归方程为

$$edu = 10.36 - 0.094sibs + 0.131medu + 0.210fedu$$

$$R^2=0.214$$

式中，edu 为劳动力受教育年数，sibs 为该劳动力家庭中兄弟姐妹的个数，medu 与 fedu 分别为母亲与父亲受到教育的年数。问

(1) sibs 是否具有预期的影响? 为什么? 若 medu 与 fedu 保持不变, 为了使预测的受教育水平减少一年, 需要 sibs 增加多少?

(2) 请对 medu 的系数给予适当的解释。

(3) 如果两个劳动力都没有兄弟姐妹, 但其中一个的父母受教育的年数为 12 年, 另一个的父母受教育的年数为 16 年, 则两人受教育的年数预期相差多少?

解答:

(1) 预期 sibs 对劳动者受教育的年数有影响。因此在收入及支出预算约束一定的条件下, 子女越多的家庭, 每个孩子接受教育的时间会越短。

根据多元回归模型偏回归系数的含义, sibs 前的参数估计值-0.094 表明, 在其他条件不变的情况下, 每增加 1 个兄弟姐妹, 受教育年数会减少 0.094 年, 因此, 要减少 1 年受教育的时间, 兄弟姐妹需增加 $1/0.094=10.6$ 个。

(2) medu 的系数表示当兄弟姐妹数与父亲受教育的年数保持不变时, 母亲每增加 1 年受教育的机会, 其子女作为劳动者就会预期增加 0.131 年的教育机会。

(3) 首先计算两人受教育的年数分别为

$$10.36+0.131 \times 12+0.210 \times 12=14.452$$

$$10.36+0.131 \times 16+0.210 \times 16=15.816$$

因此, 两人的受教育年限的差别为 $15.816-14.452=1.364$

例 2. 以企业研发支出 (R&D) 占销售额的比重为被解释变量 (Y), 以企业销售额 (X1) 与利润占销售额的比重 (X2) 为解释变量, 一个有 32 容量的样本企业的估计结果如下:

$$Y = 0.472 + 0.32 \log(X_1) + 0.05X_2$$

$$(1.37) \quad (0.22) \quad (0.046)$$

$$R^2 = 0.099$$

其中括号中为系数估计值的标准差。

(1) 解释 $\log(X_1)$ 的系数。如果 X1 增加 10%, 估计 Y 会变化多少个百分点? 这在经济上是一个很大的影响吗?

(2) 针对 R&D 强度随销售额的增加而提高这一备择假设, 检验它不随 X1 而变化的假设。分别在 5% 和 10% 的显著性水平上进行这个检验。

(3) 利润占销售额的比重 X2 对 R&D 强度 Y 是否在统计上有显著的影响?

解答:

(1) $\log(x_1)$ 的系数表明在其他条件不变时, $\log(x_1)$ 变化 1 个单位, Y 变化的单位数, 即 $\Delta Y = 0.32 \Delta \log(X_1) \approx 0.32 (\Delta X_1 / X_1) = 0.32 \times 100\%$, 换言之, 当企业销售 X1 增长 100% 时, 企业研发支出占销售额的比重 Y 会增加 0.32 个百分点。由此, 如果 X1 增加 10%, Y 会增加 0.032 个百分点。这在经济上不是一个较大的影响。

(2) 针对备择假设 $H_1: \beta_1 > 0$, 检验原假设 $H_0: \beta_1 = 0$ 。易知计算的 t 统计量的值为 $t = 0.32/0.22 = 1.468$ 。在 5% 的显著性水平下, 自由度为 $32-3=29$ 的 t 分布的临界值为 1.699 (单侧), 计算的 t 值小于该临界值, 所以不拒绝原假设。意味着 R&D 强度不随销售额的增加而变化。在 10% 的显著性水平下, t 分布的临界值为 1.311, 计算的 t 值小于该值, 拒绝原假设, 意味着 R&D 强度随销售额的增加而增加。

(3) 对 X2, 参数估计值的 t 统计值为 $0.05/0.046 = 1.087$, 它在 10% 的显著性水平下的临界值还小, 因此可以认为它对 Y 在统计上没有显著的影响。

例 3. 下表为有关经批准的私人住房单位及其决定因素的 4 个模型的估计量和相关统计值（括号内为 p-值）（如果某项为空，则意味着模型中没有此变量）。数据为美国 40 个城市的数据。模型如下：

$$\text{housing} = \beta_0 + \beta_1 \text{density} + \beta_2 \text{value} + \beta_3 \text{income} + \beta_4 \text{popchang} + \beta_5 \text{unemp} + \beta_6 \text{localtax} + \beta_7 \text{statetax} + \mu$$

式中 housing——实际颁发的建筑许可证数量，density——每平方英里的人口密度，value——自由房屋的均值（单位：百美元），income——平均家庭的收入（单位：千美元），popchang——1980~1992 年的人口增长百分比，unemp——失业率，localtax——人均交纳的地方税，statetax——人均缴纳的州税

变量	模型 A	模型 B	模型 C	模型 D
C	813 (0.74)	-392 (0.81)	-1279 (0.34)	-973 (0.44)
Density	0.075 (0.43)	0.062 (0.32)	0.042 (0.47)	
Value	-0.855 (0.13)	-0.873 (0.11)	-0.994 (0.06)	-0.778 (0.07)
Income	110.41 (0.14)	133.03 (0.04)	125.71 (0.05)	116.60 (0.06)
Popchang	26.77 (0.11)	29.19 (0.06)	29.41 (0.001)	24.86 (0.08)
Unemp	-76.55 (0.48)			
Localtax	-0.061 (0.95)			
Statetax	-1.006 (0.40)	-1.004 (0.37)		
RSS	4.763e+7	4.843e+7	4.962e+7	5.038e+7
R ²	0.349	0.338	0.322	0.312
$\hat{\sigma}^2$	1.488e+6	1.424e+6	1.418e+6	1.399e+6
AIC	1.776e+6	1.634e+6	1.593e+6	1.538e+6

- （1）检验模型 A 中的每一个回归系数在 10%水平下是否为零（括号中的值为双边备择 p-值）。根据检验结果，你认为应该把变量保留在模型中还是去掉？
- （2）在模型A中，在 10%水平下检验联合假设 $H_0: \beta_i = 0 (i=1, 5, 6, 7)$ 。说明被择假设，计算检验统计值，说明其在零假设条件下的分布，拒绝或接受零假设的标准。说明你的结论。
- （3）哪个模型是“最优的”？解释你的选择标准。
- （4）说明最优模型中有哪些系数的符号是“错误的”。说明你的预期符号并解释原因。确认其是否为正确符号。

解答：

（1）直接给出了 P-值，所以没有必要计算 t-统计值以及查 t 分布表。根据题意，如果 p-值<0.10, 则我们拒绝参数为零的原假设。

由于表中所有参数的 p-值都超过了 10%，所以没有系数是显著不为零的。但由此去掉所有解释变量，则会得到非常奇怪的结果。其实正如我们所知道的，多元回归中在省略变量时一定要谨慎，要有所选择。本例中，value、income、popchang 的 p-值仅比 0.1 稍大一点，在略掉 unemp、localtax、statetax 的模型 C 中，这些变量的系数都是显著的。

（2）针对联合假设 $H_0: \beta_i = 0 (i=1, 5, 6, 7)$ 的备择假设为 $H1: \beta_i \neq 0 (i=1, 5, 6, 7)$ 中至少有一个不为零。检验假设 H_0 ，实际上就是参数的约束性检验，非约束模型为模型A，

约束模型为模型D，检验统计值为

$$F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{RSS_U/(n - k_U - 1)} = \frac{(5.038e + 7 - 4.763e + 7)/(7 - 3)}{(4.763e + 7)/(40 - 8)} = 0.462$$

显然，在 H_0 假设下，上述统计量满足 F 分布，在 10% 的显著性水平下，自由度为 (4, 32) 的 F 分布的临界值位于 2.09 和 2.14 之间。显然，计算的 F 值小于临界值，我们不能拒绝 H_0 ，所以 $\beta_i (i=1, 5, 6, 7)$ 是联合不显著的。

(3) 模型 D 中的 3 个解释变量全部通过显著性检验。尽管 R^2 与残差平方和较大，但相对来说其 AIC 值最低，所以我们选择该模型为最优的模型。

(4) 随着收入的增加，我们预期住房需要会随之增加。所以可以预期 $\beta_3 > 0$ ，事实上其估计值确是大于零的。同样地，随着人口的增加，住房需求也会随之增加，所以我们预期 $\beta_4 > 0$ ，事实其估计值也是如此。随着房屋价格的上升，我们预期对住房的需求人数减少，即我们预期 β_5 估计值的符号为负，回归结果与直觉相符。出乎预料的是，地方税与州税为不显著的。由于税收的增加将使可支配收入降低，所以我们预期住房的需求将下降。虽然模型 A 是这种情况，但它们的影响却非常微弱。

4、在经典线性模型基本假定下，对含有三个自变量的多元回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$$

你想检验的虚拟假设是 $H_0: \beta_1 - 2\beta_2 = 1$ 。

(1) 用 $\hat{\beta}_1, \hat{\beta}_2$ 的方差及其协方差求出 $Var(\hat{\beta}_1 - 2\hat{\beta}_2)$ 。

(2) 写出检验 $H_0: \beta_1 - 2\beta_2 = 1$ 的 t 统计量。

(3) 如果定义 $\beta_1 - 2\beta_2 = \theta$ ，写出一个涉及 β_0, θ, β_2 和 β_3 的回归方程，以便能直接得

到 θ 估计值 $\hat{\theta}$ 及其标准误。

解答：

(1) 由数理统计学知识易知

$$Var(\hat{\beta}_1 - 2\hat{\beta}_2) = Var(\hat{\beta}_1) - 4Cov(\hat{\beta}_1, \hat{\beta}_2) + 4Var(\hat{\beta}_2)$$

(2) 由数理统计学知识易知

$$t = \frac{\hat{\beta}_1 - 2\hat{\beta}_2 - 1}{se(\hat{\beta}_1 - 2\hat{\beta}_2)}, \text{ 其中 } se(\hat{\beta}_1 - 2\hat{\beta}_2) \text{ 为 } (\hat{\beta}_1 - 2\hat{\beta}_2) \text{ 的标准差。}$$

(3) 由 $\beta_1 - 2\beta_2 = \theta$ 知 $\beta_1 = \theta + 2\beta_2$ ，代入原模型得

$$\begin{aligned} Y &= \beta_0 + (\theta + 2\beta_2)X_1 + \beta_2 X_2 + \beta_3 X_3 + \mu \\ &= \beta_0 + \theta X_1 + \beta_2 (2X_1 + X_2) + \beta_3 X_3 + \mu \end{aligned}$$

这就是所需的模型，其中 θ 估计值 $\hat{\theta}$ 及其标准误都能通过对该模型进行估计得到。

三、习题

(一) 基本知识类题型

3-1. 解释下列概念：

- | | |
|-----------|------------------|
| 1) 多元线性回归 | 6) 参数估计量的置信区间 |
| 2) 虚变量 | 7) 被解释变量预测值的置信区间 |
| 3) 正规方程组 | 8) 受约束回归 |
| 4) 无偏性 | 9) 无约束回归 |
| 5) 一致性 | 10) 参数稳定性检验 |

3-2. 观察下列方程并判断其变量是否呈线性？系数是否呈线性？或都是？或都不是？

- 1) $Y_i = \beta_0 + \beta_1 X_i^3 + \varepsilon_i$
- 2) $Y_i = \beta_0 + \beta_1 \log X_i + \varepsilon_i$
- 3) $\log Y_i = \beta_0 + \beta_1 \log X_i + \varepsilon_i$
- 4) $Y_i = \beta_0 + \beta_1 (\beta_2 X_i) + \varepsilon_i$
- 5) $Y_i = \frac{\beta_0}{\beta_1 X_i} + \varepsilon_i$
- 6) $Y_i = 1 + \beta_0 (1 - X_i^{\beta_1}) + \varepsilon_i$
- 7) $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} / 10 + \varepsilon_i$

3-3. 多元线性回归模型与一元线性回归模型有哪些区别？

3-4. 为什么说最小二乘估计量是最优的线性无偏估计量？多元线性回归最小二乘估计的正规方程组，能解出唯一的参数估计的条件是什么？

3-5. 多元线性回归模型的基本假设是什么？试说明在证明最小二乘估计量的无偏性和有效性的过程中，哪些基本假设起了作用？

3-6. 请说明区间估计的含义。

(二) 基本证明与问答类题型

3-7. 什么是正规方程组？分别用非矩阵形式和矩阵形式写出模型：

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$ ， $i = 1, 2, \dots, n$ 的正规方程组，及其推导过程。

3-8. 对于多元线性回归模型，证明：

$$(1) \sum e_i = 0$$

$$(2) \sum \hat{y}_i e_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}) e_i = 0$$

3-9. 为什么从计量经济学模型得到的预测值不是一个确定的值？预测值的置信区间和置信度的含义是什么？在相同的置信度下如何才能缩小置信区间？为什么？

3-10. 在多元线性回归分析中， t 检验与 F 检验有何不同？在一元线性回归分析中二者是否有等价的作用？

3-11. 设有模型： $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ ，试在下列条件下：

$$(1) \beta_1 + \beta_2 = 1$$

$$(2) \beta_1 = \beta_2$$

分别求出 β_1 和 β_2 的最小二乘估计量。

3-12. 多元线性计量经济学模型

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i \quad i = 1, 2, \cdots, n \quad (2.11.1)$$

的矩阵形式是什么？其中每个矩阵的含义是什么？熟练地写出用矩阵表示的该模型的普通最小二乘参数估计量，并证明在满足基本假设的情况下该普通最小二乘参数估计量是无偏和有效的估计量。

3-13. 有如下生产函数： $\ln X = 1.37 + 0.632 \ln K + 0.452 \ln L$

$$(0.257) \quad (0.219)$$

$$R^2 = 0.98 \quad \text{Cov}(b_K, b_L) = 0.055$$

其中括号内数值为参数标准差。请检验以下零假设：

(1) 产出量的资本弹性和劳动弹性是等同的；

(2) 存在不变规模收益，即 $\alpha + \beta = 1$ 。

3-14. 对模型 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$ 应用 OLS 法，得到回归方程如下：

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}$$

要求：证明残差 $\varepsilon_i = y_i - \hat{y}_i$ 与 $\hat{y}_i$ 不相关，即： $\sum \hat{y}_i \varepsilon_i = 0$ 。

3-15.

3-16. 考虑下列两个模型：

$$\text{I、 } y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

$$\text{II、 } (y_i - x_{2i}) = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + u'_i$$

要求：（1）证明： $\hat{\alpha}_2 = \hat{\beta}_2 - 1$, $\hat{\alpha}_1 = \hat{\beta}_1$, $\hat{\alpha}_3 = \hat{\beta}_3$

（2）证明：残差的最小二乘估计量相同，即： $\hat{u}_i = \hat{u}'_i$

（3）在何种情况下，模型 II 的拟合优度 R_2^2 会小于模型 I 拟合优度 R_1^2 。

3-17. 假设要求你建立一个计量经济模型来说明在学校跑道上慢跑一英里或一英里以上的人数，以便决定是否修建第二条跑道以满足所有的锻炼者。你通过整个学年收集数据，得到两个可能的解释性方程：

$$\text{方程 A: } \hat{Y} = 125.0 - 15.0X_1 - 1.0X_2 + 1.5X_3 \quad \bar{R}^2 = 0.75$$

$$\text{方程 B: } \hat{Y} = 123.0 - 14.0X_1 + 5.5X_2 - 3.7X_4 \quad \bar{R}^2 = 0.73$$

其中：Y——某天慢跑者的人数

X_1 ——该天降雨的英寸数

X_2 ——该天日照的小时数

X_3 ——该天的最高温度（按华氏温度）

X_4 ——第二天需交学期论文的班级数

请回答下列问题：（1）这两个方程你认为哪个更合理些，为什么？

（2）为什么用相同的数据去估计相同变量的系数得到不同的符号？

3-18. 对下列模型： $y_i = \alpha + \beta x_i + 2z_i + u_i$ （1）

$$y_i = \alpha + \beta x_i - \beta z_i + u_i \quad (2)$$

求出 β 的最小二乘估计值；并将结果与下面的三变量回归方程的最小二乘估计值作比较：

$$(3) \quad y_i = \alpha + \beta x_i - \gamma z_i + u_i \quad , \quad \text{你认为哪一个估计值更好?}$$

3-19. 假定以校园内食堂每天卖出的盒饭数量作为被解释变量，盒饭价格、气温、附近餐厅的盒饭价格、学校当日的学生数量（单位：千人）作为解释变量，进行回归分析；假设不管是否有假期，食堂都营业。不幸的是，食堂内的计算机被一次病毒侵犯，所有的存储丢失，无法恢复，你不能说出独立变量分别代表着哪一项！下面是回归结果（括号内为标准差）：

$$\hat{Y}_i = 10.6 + 28.4X_{1i} + 12.7X_{2i} + 0.61X_{3i} - 5.9X_{4i}$$

$$(2.6) \quad (6.3) \quad (0.61) \quad (5.9) \quad \bar{R}^2 = 0.63 \quad n = 35$$

要求：

(1) 试判定每项结果对应着哪一个变量？

(2) 对你的判定结论做出说明。

(三) 基本计算类题型

3-20. 试对二元线性回归模型： $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ ，（ $i = 1, 2, \dots, n$ ）作回归分

析，要求：(1) 求出未知参数 $\beta_0, \beta_1, \beta_2$ 的最小二乘估计量 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ ；

(2) 求出随机误差项 u 的方差 σ^2 的无偏估计量；

(3) 对样本回归方程作拟合优度检验；

(4) 对总体回归方程的显著性进行 F 检验；

(5) 对 β_1, β_2 的显著性进行 t 检验；

(6) 当 $X_0 = (1, X_{10}, X_{20})'$ 时，写出 $E(Y_0 | X_0)$ 和 Y_0 的置信度为 95% 的预测区间。

3-21. 下表给出三变量模型的回归结果：

方差来源	平方和 (SS)	自由度 (d.f.)	平方和的均值(MSS)
来自回归	65965	—	—
来自残差	—	—	—
总离差(TSS)	66042	14	

要求：(1) 样本容量是多少？

(2) 求 RSS？

(3) ESS 和 RSS 的自由度各是多少？

(4) 求 R^2 和 $\bar{R}^2$ ？

(5) 检验假设： X_2 和 X_3 对 Y 无影响。你用什么假设检验？为什么？

(6) 根据以上信息，你能否确定 X_2 和 X_3 各自对 Y 的贡献吗？

3-22. 下面给出依据 15 个观察值计算得到的数据：

$$\bar{Y} = 367.693, \quad \bar{X}_2 = 402.760, \quad \bar{X}_3 = 8.0, \quad \sum y_i^2 = 66042.269$$

$$\sum x_{2i}^2 = 84855.096, \quad \sum x_{3i}^2 = 280.0, \quad \sum y_i x_{2i} = 74778.346$$

$$\sum y_i x_{3i} = 4250.9, \quad \sum x_{2i} x_{3i} = 4796.0$$

其中小写字母代表了各值与其样本均值的离差。

要求：（1）估计三个多元回归系数；

（2）估计它们的标准差；并求出 R^2 与 $\bar{R}^2$ ？

（3）估计 B_2 、 B_3 95% 的置信区间；

（4）在 $\alpha = 5\%$ 下，检验估计的每个回归系数的统计显著性（双边检验）；

（5）检验在 $\alpha = 5\%$ 下所有的部分系数都为零，并给出方差分析表。

3-23. 考虑以下方程（括号内为估计标准差）：

$$\hat{W}_t = 8.562 + 0.364P_t + 0.004P_{t-1} - 2.560U_t$$

(0.080) (0.072) (0.658) $n = 19$ $R^2 = 0.873$

其中： W —— t 年的每位雇员的工资和薪水

P —— t 年的物价水平

U —— t 年的失业率

要求：（1）对个人收入估计的斜率系数进行假设检验；（尽量在做本题之前不参考结果）

（2）讨论 P_{t-1} 在理论上的正确性，对本模型的正确性进行讨论； P_{t-1} 是否应从方程中删除？为什么？

3-24. 下表是某种商品的需求量、价格和消费者收入十年的时间序列资料：

年份	1	2	3	4	5	6	7	8	9	10
需求量	59190	65450	62360	64700	67400	64440	68000	72400	75710	70680
价 格	23.56	24.44	32.07	32.46	31.15	34.14	35.30	38.70	39.63	46.68
收 入	76200	91200	106700	111600	119000	129200	143400	159600	180000	193000

要求：（1）已知商品需求量 Y 是其价格 X_1 和消费者收入 X_2 的函数，试求 Y 对 X_1 和 X_2 的最

小二乘回归方程： $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$

（2）求 Y 的总变差中未被 X_1 和 X_2 解释的部分，并对回归方程进行显著性检验；

（3）对回归参数 $\hat{\beta}_1$ ， $\hat{\beta}_2$ 进行显著性 t 检验。

3-25. 参考习题 2-28 给出的数据，要求：

（1）建立一个多元回归模型，解释 MBA 毕业生的平均初职工资，并且求出回归结果；

（2）如果模型中包括了 GPA 和 GMAT 分数这两个解释变量，先验地，你可能会遇到什么问题，为什么？

(3) 如果学费这一变量的系数为正、并且在统计上是显著的，是否表示进入最昂贵的商业学校是值得的。学费这个变量可用什么来代替？

3-26. 经研究发现，学生用于购买书籍及课外读物的支出与本人受教育年限和其家庭收入水平有关，对 18 名学生进行调查的统计资料如下表所示：

学生 序号	购买书籍及课外 读物支出 Y (元/	受教育年限 X_1 (年)	家庭月可支配收 入 X_2 (元/月)
1	450.5	4	171.2
2	507.7	4	174.2
3	613.9	5	204.3
4	563.4	4	218.7
5	501.5	4	219.4
6	781.5	7	240.4
7	541.8	4	273.5
8	611.1	5	294.8
9	1222.1	10	330.2
10	793.2	7	333.1
11	660.8	5	366.0
12	792.7	6	350.9
13	580.8	4	357.9
14	612.7	5	359.0
15	890.8	7	371.9
16	1121.0	9	435.3
17	1094.2	8	523.9
18	1253.0	10	604.1

要求：

(1) 试求出学生购买书籍及课外读物的支出 Y 与受教育年限 X_1 和家庭收入水平 X_2 的估计

的回归方程： $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$

(2) 对 β_1, β_2 的显著性进行 t 检验；计算 R^2 和 $\overline{R}^2$ ；

(3) 假设有一学生的受教育年限 $X_1 = 10$ 年，家庭收入水平 $X_2 = 480$ 元/月，试预测该学生全年购买书籍及课外读物的支出，并求出相应的预测区间 ($\alpha = 0.05$)。

3-27. 根据 100 对 (x_1, y) 的观察值计算出：

$$\sum \dot{x}_1^2 = 12 \quad \sum \dot{x}_1 \dot{y} = -9 \quad \sum \dot{y}^2 = 30$$

要求：

(1) 求出一元模型 $y = \beta_0 + \beta_1 x_1 + u$ 中的 β_1 的最小二乘估计量及其相应的标准差估计量;

(2) 后来发现 y 还受 x_2 的影响, 于是将一元模型改为二元模型 $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + v$,

收集 x_2 的相应观察值并计算出:

$$\sum \dot{x}_2^2 = 6 \quad \sum \dot{x}_2 \dot{y} = 8 \quad \sum \dot{x}_1 \dot{x}_2 = 2$$

求二元模型中的 α_1 , α_2 的最小二乘估计量及其相应的标准差估计量;

(3) 一元模型中的 $\hat{\beta}_1$ 与二元模型中的 $\hat{\alpha}_1$ 是否相等? 为什么?

3-28. 考虑以下预测的回归方程:

$$\hat{Y}_t = -120 + 0.10F_t + 5.33RS_t \quad \overline{R^2} = 0.50$$

其中: Y_t ——第 t 年的玉米产量 (蒲式耳/亩)

F_t ——第 t 年的施肥强度 (磅/亩)

RS_t ——第 t 年的降雨量 (英寸)

要求回答下列问题:

(1) 从 F 和 RS 对 Y 的影响方面, 说出本方程中系数 0.10 和 5.33 的含义;

(2) 常数项 -120 是否意味着玉米的负产量可能存在?

(3) 假定 β_F 的真实值为 0.40, 则估计值是否有偏? 为什么?

(4) 假定该方程并不满足所有的古典模型假设, 即并不是最佳线性无偏估计值, 则是否意味着 β_{RS} 的真实值绝对不等于 5.33? 为什么?

3-29. 已知线性回归模型 $\mathbf{Y} = \mathbf{XB} + \mathbf{U}$ 式中 $\mathbf{U} \sim (0, \sigma^2 \mathbf{I})$, $n = 13$ 且 $k = 3$ (n 为样本

容量, k 为参数的个数), 由二次型 $(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})$ 的最小化得到如下线性方程组:

$$\hat{\beta}_1 + 2\hat{\beta}_2 + \hat{\beta}_3 = 3$$

$$2\hat{\beta}_1 + 5\hat{\beta}_2 + \hat{\beta}_3 = 9$$

$$\hat{\beta}_1 + \hat{\beta}_2 + 6\hat{\beta}_3 = -8$$

要求: (1) 把问题写成矩阵向量的形式; 用求逆矩阵的方法求解之;

(2) 如果 $\mathbf{Y}'\mathbf{Y} = 53$, 求 $\hat{\sigma}^2$;

(3) 求出 $\hat{\beta}$ 的方差—协方差矩阵。

3-30. 已知数据如下表：

Y	X_1	X_2
1	1	10
3	2	9
8	3	5
15	4	1
28	5	-6

要求：(1) 先根据表中数据估计以下回归模型的方程（只估计参数不用估计标准差）：

$$y_i = \alpha_0 + \alpha_1 x_{1i} + u_{1i}$$

$$y_i = \lambda_0 + \lambda_2 x_{2i} + u_{2i}$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

(2) 回答下列问题： $\alpha_1 = \beta_1$ 吗？为什么？ $\lambda_2 = \beta_2$ 吗？为什么？

(四) 自我综合练习类题型

3-31. 自己选择研究对象(最好是一个实际经济问题)，收集样本数据，应用计量经济学软件（建议使用 Eviews3.1），完成建立多元线性计量经济模型的全过程，并写出详细研究报告。

四、习题参考答案

（一）基本知识类题型

3-1. 解释下列概念

（1）在现实经济活动中往往存在一个被解释变量受到多个解释变量的影响的现象，表现为在线性回归模型中有多个解释变量，这样的模型被称为多元线性回归模型，多元指多个解释变量。

（2）形如 $\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\hat{\mathbf{B}}$ 的关于参数估计值的线性代数方程组称为正规方程组。

3-2. 答：变量非线性、系数线性；变量、系数均线性；变量、系数均线性；变量线性、系数非线性；变量、系数均为非线性；变量、系数均为非线性；变量、系数均为线性。

3-3. 答：多元线性回归模型与一元线性回归模型的区别表现在如下几方面：一是解释变量的个数不同；二是模型的经典假设不同，多元线性回归模型比一元线性回归模型多了“解释变量之间不存在线性相关关系”的假定；三是多元线性回归模型的参数估计式的表达更复杂；

3-4. 在多元线性回归模型中，参数的最小二乘估计量具备线性、无偏性、最小方差性，同时多元线性回归模型满足经典假定，所以此时的最小二乘估计量是最优的线性无偏估计量，又称 BLUE 估计量。对于多元线性回归最小二乘估计的正规方程组，

3-5. 答：多元线性回归模型的基本假定有：零均值假定、随机项独立同方差假定、解释变量的非随机性假定、解释变量之间不存在线性相关关系假定、随机误差项 u_i 服从均值为 0 方差为 σ^2 的正态分布假定。在证明最小二乘估计量的无偏性中，利用了解释变量与随机误差项不相关的假定；在有效性的证明中，利用了随机项独立同方差假定。

3-6. 答：区间估计是指研究用未知参数的点估计值（从一组样本观测值算得的）作为近似值的精确程度和误差范围。

（二）基本证明与问答类题型

3-7. 答：含有待估关系估计量的方程组称为正规方程组。

正规方程组的非矩阵形式如下：

$$\begin{cases} \sum y_i - \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}) = 0 \\ \sum y_i x_{1i} - \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}) x_{1i} = 0 \\ \sum y_i x_{2i} - \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}) x_{2i} = 0 \\ \vdots \\ \sum y_i x_{ki} - \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}) x_{ki} = 0 \end{cases}$$

正规方程组的矩阵形式如下：

$$X'Y = X'XB$$

推导过程略。

3-16. 解：

(1) 证明：由参数估计公式可得下列参数估计值

$$\begin{aligned} \hat{\alpha}_2 &= \frac{\begin{vmatrix} \sum \dot{x}_{2i}(\dot{y}_i - \dot{x}_{2i}) & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{3i}(\dot{y}_i - \dot{x}_{2i}) & \sum \dot{x}_{3i}^2 \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} \\ &= \frac{\begin{vmatrix} \sum \dot{x}_{2i}\dot{y}_i & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{3i}\dot{y}_i & \sum \dot{x}_{3i}^2 \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} - \frac{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} \\ &= \hat{\beta}_2 - 1 \end{aligned}$$

$$\begin{aligned} \hat{\alpha}_3 &= \frac{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}(\dot{y}_i - \dot{x}_{2i}) \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}(\dot{y}_i - \dot{x}_{2i}) \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} \\ &= \frac{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{y}_i \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}\dot{y}_i \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} - \frac{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}^2 \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{2i}\dot{x}_{3i} \end{vmatrix}}{\begin{vmatrix} \sum \dot{x}_{2i}^2 & \sum \dot{x}_{2i}\dot{x}_{3i} \\ \sum \dot{x}_{2i}\dot{x}_{3i} & \sum \dot{x}_{3i}^2 \end{vmatrix}} \\ &= \hat{\beta}_3 \end{aligned}$$

$$\begin{aligned} \hat{\alpha}_1 &= \bar{y} - \bar{x} - \hat{\alpha}_2 \bar{x}_2 - \hat{\alpha}_3 \bar{x}_3 \\ &= \bar{y} - (1 + \hat{\alpha}_2) \bar{x}_2 - \hat{\alpha}_3 \bar{x}_3 \\ &= \bar{y} - \hat{\beta}_2 \bar{x}_2 - \hat{\beta}_3 \bar{x}_3 \\ &= \hat{\beta}_1 \end{aligned}$$

证毕。

(2)证明:

$$\begin{aligned}\hat{u}'_i &= y_i - x_{2i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i} \\ &= y_i - \hat{\alpha}_1 - (1 + \hat{\alpha}_2)x_{2i} - \hat{\alpha}_3 x_{3i} \\ &= y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i} \\ &= \hat{u}_i\end{aligned}$$

证毕。

(3)设: $z_i = y_i - x_{2i}$

I 式的拟合优度为:

$$R_1^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2}$$

II 式的拟合优度为:

$$R_2^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum \hat{u}'_i^2}{\sum (z_i - \bar{z})^2}$$

在(2)中已经证得 $\hat{u}_i = \hat{u}'_i$ 成立, 即二式分子相同, 若要模型 II 的拟合优度 R_2^2 小于模型

I 的拟合优度 R_1^2 , 必须满足: $\sum (z_i - \bar{z})^2 < \sum (y_i - \bar{y})^2$ 。

3-17. 答:

(1)方程 B 更合理些。原因是: 方程 B 中的参数估计值的符号与现实更接近些, 如与日照的小时数同向变化, 天长则慢跑的人会多些; 与第二天需交学期论文的班级数成反向变化, 这一点在学校的跑道模型中是一个合理的解释变量。

(2)解释变量的系数表明该变量的单位变化在方程中其他解释变量不变的条件下对被解释变量的影响, 在方程 A 和方程 B 中由于选择了不同的解释变量, 如方程 A 选择的是“该天的最高温度”而方程 B 选择的是“第二天需交学期论文的班级数”, 由此造成 X_2 与这两个变量之间的关系不同, 所以用相同的数据估计相同的变量得到不同的符号。

3-18. 答:

将模型(1)改写成 $(y_i - 2z_i) = \alpha + \beta x_i + u_i$, 则 β 的估计值为:

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - 2z_i)}{\sum (x_i - \bar{x})^2}$$

将模型(2)改写成 $y_i = \alpha + \beta(x_i - z_i) + u_i$ ，则 β 的估计值为：

$$\hat{\beta} = \frac{\sum (x_i - z_i - \bar{x} + \bar{z}) y_i}{\sum (x_i - z_i - \bar{x} + \bar{z})^2}$$

这两个模型都是三变量回归模型(3)在某种限制条件下的变形。如果限制条件正确，则前两个回归参数会更有效；如果限制条件不正确则前两个回归参数会有偏。

3-19. 答：

(1)答案并不唯一，猜测为： X_1 为学生数量， X_2 为附近餐厅的盒饭价格， X_3 为气温，

X_4 为校园内食堂的盒饭价格；

(2)理由是被解释变量应与学生数量成正比，并且应该影响显著；与本食堂盒饭价格成反比，这与需求理论相吻合；与附近餐厅的盒饭价格成正比，因为彼此是替代品；与气温的变化关系不是十分显著，因为大多数学生不会因为气温升高不吃饭。

(三) 基本计算类题型

3-22. 解：(1)

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \sum x_{2i} x_{3i} \sum x_{2i} x_{3i}} \\&= \frac{74778.346 \times 280 - 4250.9 \times 4796.0}{84855.096 \times 280 - 4796.0^2} \\&= \frac{550620}{757810} \\&= 0.7266 \\ \hat{\beta}_3 &= \frac{\sum y_i x_{3i} \sum x_{2i}^2 - \sum y_i x_{2i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \sum x_{2i} x_{3i} \sum x_{2i} x_{3i}} \\&= \frac{4250.9 \times 84855.096 - 74778.346 \times 4796.0}{84855.096 \times 280 - 4796.0^2} \\&= \frac{2073580}{757810} \\&= 2.7363 \\ \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \\&= 367.693 - 0.7266 \times 402.760 - 2.7363 \times 8.0 \\&= 53.1572\end{aligned}$$

(2)

$$\begin{aligned}\sigma^2 &= \frac{\sum e_i^2}{n-3} = \frac{\sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}}{15-3} \\ &= \frac{66042.269 - 0.7266 \times 74778.346 - 2.7363 \times 4250.9}{12} \\ &= 6.3821\end{aligned}$$

$$se(\beta_1) = \sqrt{Var(\beta_1)} = \sqrt{\frac{1}{15} \times A \times \sigma^2} = 12.768$$

$$\text{其中: } A = \frac{\bar{X}^2 \times \sum x_{3i}^2 + \bar{X}^2 \times \sum x_{2i}^2 - 2\bar{X}\bar{X} \times \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \sum x_{2i}x_{3i} \sum x_{2i}x_{3i}}$$

同理, 可得: $se(\beta_2) = 0.0486$, $se(\beta_3) = 0.8454$

$$\text{拟合优度为: } R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} = 0.9988$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 0.9986$$

(3) $d.f. = 12, \alpha = 5\%$, 查表得 $P(|t| \leq 2.179) = 0.95$

$$-2.179 \leq \frac{0.7266 - \beta_2}{0.0486} \leq 2.179, \text{ 得到 } 0.6207 \leq \beta_2 \leq 0.8325$$

$$-2.179 \leq \frac{2.7363 - \beta_3}{0.8454} \leq 2.179, \text{ 得到 } 0.8942 \leq \beta_3 \leq 4.5784$$

$\therefore \beta_2$ 95%的置信区间为: $0.6207 \leq \beta_2 \leq 0.8325$,

β_3 95%的置信区间为: $0.8942 \leq \beta_3 \leq 4.5784$

(4) $H_0: B_i = 0, (i = 1, 2, 3)$, $H_1: B_i \neq 0$

$\alpha = 5\%$ (双边), $d.f. = 15 - 3 = 12$ 查表得临界值为 $-2.179 \leq t \leq 2.179$

$$\text{则: } t_{\beta_1} = \frac{53.1572 - 0}{12.9768} = 4.0963 > 2.179, \therefore \text{拒绝零假设: } B_1 = 0$$

$$t_{\beta_2} = \frac{0.7266 - 0}{0.0486} = 14.9509 > 2.179, \therefore \text{拒绝零假设: } B_2 = 0$$

$$t_{\beta_3} = \frac{2.7363 - 0}{0.8454} = 3.2367 > 2.179, \therefore \text{拒绝零假设: } B_3 = 0$$

(5) 所有的部分系数为 0, 即: $H_0 = B_1 = B_2 = 0$, 等价于 $H_0: R^2 = 0$

方差来源	平方和	自由度	平方和的均值
来自回归	65963.018	2	32981.509
来自残差	79.2507	12	6.6042

$$F = \frac{32981.509}{6.6042} = 4994.0203, \alpha = 5\%, d.f. = 2, 12, F \text{ 临界值为 } 3.89$$

$\therefore F$ 值是显著的，所以拒绝零假设。

3-23. 解：

(1) 对给定在 5% 的显著水平下，可以进行 t 检验，得到的结果如下：

系数	$\hat{\beta}_{PI}$	$\hat{\beta}_{PI-1}$	$\hat{\beta}_U$
假设符号	+	+	
T 值			
5% 显著水平			

3-28. 解：

(1) 在降雨量不变时，每亩增加一磅肥料将使第 t 年的玉米产量增加 0.1 蒲式耳/亩；在每亩施肥量不变的情况下，每增加一英寸的降雨量将使第 t 年的玉米产量增加 5.33 蒲式耳/亩；

(2) 在种地的一年中不施肥、也不下雨的现象同时发生的可能性极小，所以玉米的负产量不可能存在；

(3) 如果 β_F 的真实值为 0.40，并不能说明 0.1 是有偏的估计，理由是 0.1 是本题估计的参数，而 0.40 是从总体得到的系数的均值。

(4) 不一定。即便该方程并不满足所有的古典模型假设、不是最佳线性无偏估计值，也有可能得出的估计系数等于 5.33。

3-29. 解：

(1) 该方程组的矩阵向量形式为：

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 6 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ -8 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 9 \\ -8 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}$$

$$(2) \hat{\sigma}^2 = \frac{(TSS - RSS)}{n - k} = \frac{53 - 3 \times 3 - 1 \times 9 - 2 \times 8}{13 - 3} = 1.9$$

(3) $\hat{\beta}$ 的方差—协方差矩阵为:

$$V - Cov(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 1.9 \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 6 \end{bmatrix}^{-1} = \begin{bmatrix} 6.525 & -2.475 & -0.675 \\ -2.475 & 1.125 & 0.225 \\ -0.675 & 0.225 & 0.619 \end{bmatrix}$$

第四章 经典单方程计量经济学模型：放宽基本假定的模型

一、内容提要

本章主要介绍计量经济模型的二级检验问题，即计量经济检验。主要讨论对回归模型的若干基本经典假定是否成立进行检验、当检验发现不成立时继续采用 OLS 估计模型所带来的不良后果以及如何修正等问题。具体包括异方差性问题、序列相关性问题、多重共线性问题以及随机解释变量这四大类问题。

异方差是模型随机扰动项的方差不同时产生的一类现象。在异方差存在的情况下，OLS 估计尽管是无偏、一致的，但通常的假设检验却不再可靠，这时仍采用通常的 t 检验和 F 检验，则有可能导致出现错误的结论。同样地，由于随机项异方差的存在而导致的参数估计值的标准差的偏误，也会使采用模型的预测变得无效。对模型的异方差性有若干种检测方法，如图示法、Park 与 Gleiser 检验法、Goldfeld-Quandt 检验法以及 White 检验法等。而当检测出模型确实存在异方差性时，通过采用加权最小二乘法进行修正的估计。

序列相关性也是模型随机扰动项出现序列相关时产生的一类现象。与异方差的情形相类似，在序列相关存在的情况下，OLS 估计量仍具无偏性与一致性，但通常的假设检验不再可靠，预测也变得无效。序列相关性的检测方法也有若干种，如图示法、回归检验法、Durbin-Watson 检验法以及 Lagrange 乘子检验法等。存在序列相关性时，修正的估计方法有广义最小二乘法（GLS）以及广义差分法。

多重共线性是多元回归模型可能存在的一类现象，分为完全共线与近似共线两类。模型的多个解释变量间出现完全共线性时，模型的参数无法估计。更多的情况则是近似共线性，这时，由于并不违背所有的基本假定，模型参数的估计仍是无偏、一致且有效的，但估计的参数标准差往往较大，从而使得 t-统计值减小，参数的显著性下降，导致某些本应存在于模型中的变量被排除，甚至出现参数正负号方面的一些混乱。显然，近似多重共线性使得模型偏回归系数的特征不再明显，从而很难对单个系数的经济含义进行解释。多重共线性的检验包括检验多重共线性是否存在以及估计多重共线性的范围两层递进的检验。而解决多重共线性的办法通常有逐步回归法、差分法以及使用额外信息、增大样本容量等方法。

当模型中的解释变量是随机解释变量时，需要区分三种类型：随机解释变量与随机扰动项独立，随机解释变量与随机扰动项同期无关、但异期相关，随机解释变量与随机扰动项

同期相关。第一种类型不会对 OLS 估计带来任何问题。第二种类型则往往导致模型估计的有偏性，但随着样本容量的增大，偏误会逐渐减小，因而具有一致性。所以，扩大样本容量是克服偏误的有效途径。第三种类型的 OLS 估计则既是有偏、也是非一致的，需要采用工具变量法来加以克服。

二、典型例题分析

1、下列哪种情况是异方差性造成的结果？

- (1) OLS 估计量是有偏的
- (2) 通常的 t 检验不再服从 t 分布。
- (3) OLS 估计量不再具有最佳线性无偏性。

解答：

第 (2) 与 (3) 种情况可能由于异方差性造成。异方差性并不会引起 OLS 估计量出现偏误。

2、已知模型

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

$$\text{Var}(u_t) = \sigma_t^2 = \sigma^2 Z_t^2$$

式中， Y 、 X_1 、 X_2 和 Z 的数据已知。假设给定权数 w_t ，加权最小二乘法就是求下式中的各 β ，以使的该式最小

$$RSS = \sum (w_t u_t)^2 = \sum (w_t Y_t - \beta_0 w_t - \beta_1 w_t X_{1t} - \beta_2 w_t X_{2t})^2$$

- (1) 求 RSS 对 β_1 、 β_2 和 β_0 的偏微分并写出正规方程。
- (2) 用 Z 去除原模型，写出所得新模型的正规方程组。
- (3) 把 $w_t = 1/Z_t$ 带入 (1) 中的正规方程，并证明它们和在 (2) 中推导的结果一样。

解答：

(1) 由 $RSS = \sum (w_t u_t)^2 = \sum (w_t Y_t - \beta_0 w_t - \beta_1 w_t X_{1t} - \beta_2 w_t X_{2t})^2$ 对各 β 求偏导得如下正规方程组：

$$\sum (w_t Y_t - \beta_0 w_t - \beta_1 w_t X_{1t} - \beta_2 w_t X_{2t}) w_t = 0$$

$$\sum (w_t Y_t - \beta_0 w_t - \beta_1 w_t X_{1t} - \beta_2 w_t X_{2t}) w_t X_{1t} = 0$$

$$\sum (w_t Y_t - \beta_0 w_t - \beta_1 w_t X_{1t} - \beta_2 w_t X_{2t}) w_t X_{2t} = 0$$

(2) 用 Z 去除原模型，得如下新模型

$$\frac{Y_t}{Z_t} = \frac{\beta_0}{Z_t} + \beta_1 \frac{X_{1t}}{Z_t} + \beta_2 \frac{X_{2t}}{Z_t} + \frac{u_t}{Z_t}$$

对应的正规方程组如下所示：

$$\sum \left(\frac{Y_t}{Z_t} - \frac{\beta_0}{Z_t} - \beta_1 \frac{X_{1t}}{Z_t} - \beta_2 \frac{X_{2t}}{Z_t} \right) \frac{1}{Z_t} = 0$$

$$\sum \left(\frac{Y_t}{Z_t} - \frac{\beta_0}{Z_t} - \beta_1 \frac{X_{1t}}{Z_t} - \beta_2 \frac{X_{2t}}{Z_t} \right) \frac{X_{1t}}{Z_t} = 0$$

$$\sum \left(\frac{Y_t}{Z_t} - \frac{\beta_0}{Z_t} - \beta_1 \frac{X_{1t}}{Z_t} - \beta_2 \frac{X_{2t}}{Z_t} \right) \frac{X_{2t}}{Z_t} = 0$$

(3) 如果用 $\frac{1}{Z_t}$ 代替 (1) 中的 w_t ，则容易看到与 (2) 中的正规方程组是一样的。

3、已知模型 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

式中， Y_i 为某公司在第 i 个地区的销售额； X_{1i} 为该地区的总收入； X_{2i} 为该公司在该地区投入的广告费用 ($i=0, 1, 2, \dots, 50$)。

(1) 由于不同地区人口规模 P_i 可能影响着该公司在该地区的销售，因此有理由怀疑随机误差项 u_i 是异方差的。假设 σ_i 依赖于总体 P_i 的容量，请逐步描述你如何对此进行检验。

需说明：1) 零假设和备择假设；2) 要进行的回归；3) 要计算的检验统计值及它的分布（包括自由度）；4) 接受或拒绝零假设的标准。

(2) 假设 $\sigma_i = \sigma P_i$ 。逐步描述如何求得 BLUE 并给出理论依据。

解答：

(1) 如果 σ_i 依赖于总体 P_i 的容量，则随机扰动项的方差 σ_i^2 依赖于 P_i^2 。因此，要进行回归的一种形式为 $\sigma_i^2 = \alpha_0 + \alpha_1 P_i^2 + \varepsilon_i$ 。于是，要检验的零假设 $H_0: \alpha_1 = 0$ ，备择

假设 $H_1: \alpha_1 \neq 0$ 。检验步骤如下：

第一步：使用 OLS 方法估计模型，并保存残差平方项 $\tilde{e}_i^2$ ；

第二步：做 $\tilde{e}_i^2$ 对常数项 C 和 P_i^2 的回归

第三步：考察估计的参数 α_1 的 t 统计量，它在零假设下服从自由度为 2 的 t 分布。

第四步：给定显著性水平 0.05（或其他），查相应的自由度为 2 的 t 分布的临界值，如果估计的参数 $\hat{\alpha}_1$ 的 t 统计值大于该临界值，则拒绝同方差的零假设。

(2) 假设 $\sigma_i = \sigma P_i$ 时, 模型除以 P_i 有:

$$\frac{Y_i}{P_i} = \beta_0 \frac{1}{P_i} + \beta_1 \frac{X_{1i}}{P_i} + \beta_2 \frac{X_{2i}}{P_i} + \frac{u_i}{P_i}$$

由于 $\text{Var}(u_i / P_i) = \sigma_i^2 / P_i^2 = \sigma^2$, 所以在该变换模型中可以使用 OLS 方法, 得出 BLUE 估计值。方法是对 Y_i / P_i 关于 $1/P_i$ 、 X_{1i} / P_i 、 X_{2i} / P_i 做回归, 不包括常数项。

4、以某地区 22 年的年度数据估计了如下工业就业回归方程

$$Y = -3.89 + 0.51 \ln X_1 - 0.25 \ln X_2 + 0.62 \ln X_3$$

$$(-0.56) \quad (2.3) \quad (-1.7) \quad (5.8)$$

$$\bar{R}^2 = 0.996 \quad DW = 1.147$$

式中, Y 为总就业量; X1 为总收入; X2 为平均月工资率; X3 为地方政府的总支出。

(1) 试证明: 一阶自相关的 DW 检验是无定论的。

(2) 逐步描述如何使用 LM 检验

解答:

(1) 由于样本容量 $n=22$, 解释变量个数为 $k=3$, 在 5% 显著性水平下, 相应的上下临界值为 $d_U = 1.664$ 、 $d_L = 1.503$ 。由于 $DW=1.147$ 位于这两个值之间, 所以 DW 检验是无定论的。

(2) 进行 LM 检验:

第一步, 做 Y 关于常数项、 $\ln X_1$ 、 $\ln X_2$ 和 $\ln X_3$ 的回归并保存残差 $\tilde{e}_t$;

第二步, 做 $\tilde{e}_t$ 关于常数项、 $\ln X_1$ 、 $\ln X_2$ 和 $\ln X_3$ 和 $\tilde{e}_{t-1}$ 的回归并计算 R^2 ;

第三步, 计算检验统计值 $(n-1)R^2 = 21 \times 0.996 = 20.916$;

第四步, 由于在不存在一阶序列相关的零假设下 $(n-1)R^2$ 呈自由度为 1 的 χ^2 分布。在 5% 的显著性水平下, 该分布的相应临界值为 3.841。由于 $20.916 > 3.841$, 因此拒绝零假设, 意味着原模型随机扰动项存在一阶序列相关。

5、某地区供水部门利用最近 15 年的用水年度数据得出如下估计模型:

$$water = -326.9 + 0.305 house + 0.363 pop - 0.005 pcy - 17.87 price - 1.123 rain$$

$$(-1.7) \quad (0.9) \quad (1.4) \quad (-0.6) \quad (-1.2) \quad (-0.8)$$

$$\bar{R}^2 = 0.93 \quad F=38.9$$

式中, water——用水总量 (百万立方米), house——住户总数 (千户), pop——总人口 (千人), pcy——人均收入 (元), price——价格 (元/100 立方米), rain——降雨量 (毫米)。

(1) 根据经济理论和直觉, 请计回归系数的符号是什么(不包括常量), 为什么? 观察符号与你的直觉相符吗?

(2) 在 10% 的显著性水平下, 请进行变量的 t-检验与方程的 F-检验。T 检验与 F 检验结果有相矛盾的现象吗?

(3) 你认为估计值是 (1) 有偏的; (2) 无效的或 (3) 不一致的吗? 详细阐述理由。

解答:

(1) 在其他变量不变的情况下, 一城市的人口越多或房屋数量越多, 则对用水的需求越高。所以可期望 house 和 pop 的符号为正; 收入较高的个人可能用水较多, 因此 pcy 的预期符号为正, 但它可能是不显著的。如果水价上涨, 则用户会节约用水, 所以可预期 price 的系数为负。显然如果降雨量较大, 则草地和其他花园或耕地的用水需求就会下降, 所以可以期望 rain 的系数符号为负。从估计的模型看, 除了 pcy 之外, 所有符号都与预期相符。

(2) t-统计量检验单个变量的显著性, F-统计值检验变量是否是联合显著的。

这里 t-检验的自由度为 $15-5-1=9$, 在 10% 的显著性水平下的临界值为 1.833。可见, 所有参数估计值的 t 值的绝对值都小于该值, 所以即使在 10% 的水平下这些变量也不是显著的。

这里, F-统计值的分子自由度为 5, 分母自由度为 9。10% 显著性水平下 F 分布的临界值为 2.61。可见计算的 F 值大于该临界值, 表明回归系数是联合显著的。

T 检验与 F 检验结果的矛盾可能是由于多重共线性造成的。house、pop、pcy 都是高度相关的, 这将使它们的 t-值降低且表现为不显著。price 和 rain 不显著另有原因。根据经验, 如果一个变量的值在样本期间没有很大的变化, 则它对被解释变量的影响就不能够很好地被度量。可以预期水价与年降雨量在各年中一般没有太大的变化, 所以它们的影响很难度量。

(3) 多重共线性往往表现的是解释变量间的样本观察现象, 在不存在完全共线性的情况下, 近似共线并不意味着基本假定的任何改变, 所以 OLS 估计量的无偏性、一致性和有效性仍然成立, 即仍是 BLUE 估计量。但共线性往往导致参数估计值的方差大于不存在多重共线性的情况。

6、一个对某地区大学生就业增长影响的简单模型可描述如下

$$gEMP_t = \beta_0 + \beta_1 gMIN_{1t} + \beta_2 gPOP + \beta_3 gGDP_{1t} + \beta_4 gGDP_t + \mu_t$$

式中, 为新就业的大学生人数, MIN1 为该地区最低限度工资, POP 为新毕业的大学生人数, GDP1 为该地区国内生产总值, GDP 为该国国内生产总值; g 表示年增长率。

(1) 如果该地区政府以多多少少不易观测的却对新毕业大学生就业有影响的因素作为基础来选择最低限度工资, 则 OLS 估计将会存在什么问题?

(2) 令 MIN 为该国的最低限度工资, 它与随机扰动项相关吗?

(3) 按照法律, 各地区最低限度工资不得低于国家最低工资, 那么 gMIN 能成为 gMIN1 的工具变量吗?

解答:

(1) 由于地方政府往往是根据过去的经验、当前的经济状况以及期望的经济发展前景来定制地区最低限度工资水平的, 而这些因素没有反映在上述模型中, 而是被归结到了模型

的随机扰动项中，因此 $gMIN1$ 与 μ 不仅异期相关，而且往往是同期相关的，这将引起 OLS 估计量的偏误，甚至当样本容量增大时也不具有一致性。

(2) 全国最低限度的制定主要根据全国整体的情况而定，因此 $gMIN$ 基本与上述模型的随机扰动项无关。

(3) 由于地方政府在制定本地区最低工资水平时往往考虑全国的最低工资水平的要求，因此 $gMIN1$ 与 $gMIN$ 具有较强的相关性。结合 (2) 知 $gMIN$ 可以作为 $gMIN1$ 的工具变量使用。

三、习题

(一) 基本知识类题型

4-1. 解释下列概念：

- | | |
|-------------|--------------|
| (1) 异方差性 | (6) 不完全多重共线性 |
| (2) 序列相关性 | (7) 随机解释变量 |
| (3) 多重共线性 | (8) 差分法 |
| (4) 偏回归系数 | (9) 广义最小二乘法 |
| (5) 完全多重共线性 | (10) D.W.检验 |

4-2. 判断下列各题对错，并简单说明理由：

- 1) 在存在异方差情况下，普通最小二乘法 (OLS) 估计量是有偏的和无效的；
- 2) 如果存在异方差，通常使用的 t 检验和 F 检验是无效的；
- 3) 在存在异方差情况下，常用的 OLS 法总是高估了估计量的标准差；
- 4) 如果从 OLS 回归中估计的残差呈现系统模式，则意味着数据中存在着异方差；
- 5) 当存在序列相关时，OLS 估计量是有偏的并且也是无效的；
- 6) 消除序列相关的一阶差分变换假定自相关系数 ρ 必须等于 1；
- 7) 两个模型，一个是一阶差分形式，一个是水平形式，这两个模型的 R^2 值是不可以直接比较的。
- 8) 回归模型中误差项 u_t 存在异方差时，OLS 估计不再是有效的；
- 9) 回归模型中误差项 u_t 存在序列相关时，OLS 估计不再是无偏的；

4-3. 简述异方差对下列各项有何影响：(1) OLS 估计量及其方差；(2) 置信区间；(3) 显

著性 t 检验和 F 检验的使用。

4-4. 在存在 AR (1) 自相关的情形下, 什么估计方法能够产生 BLUE 估计量? 简述这个方法的具体步骤。

(二) 基本证明与问答类题型

4-5. 在存在 AR (1) 的情形下, 估计自相关参数 ρ 有哪些不同的方法?

4-6. 在如下回归中, 你是否预期存在着异方差?

Y	X	样本
a) 公司利润	净财富	《财富》500 强
b) 公司利润的对数	净财富的对数	《财富》500 强
c) 道琼斯工业平均指数	时间	1960~1990 年 (年平均)
d) 婴儿死亡率	人均收入	100 个发达国家和发展中国家
e) 通货膨胀率	货币增长率	美国、加拿大和 15 个拉美国家

4-7. 已知消费模型: $y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_t$

其中: y_t ——消费支出

x_{1t} ——个人可支配收入

x_{2t} ——消费者的流动资产

$$E(u_t) = 0$$

$$Var(u_t) = \sigma^2 x_{1t}^2 \text{ (其中 } \sigma^2 \text{ 为常数)}$$

要求:

(1) 进行适当变换消除异方差, 并证明之;

(2) 写出消除异方差后, 模型的参数估计量的表达式。

4-8. 什么是异方差性? 举例说明经济现象中的异方差性。检验异方差性的方法思路是什么?

4-9. 什么是序列相关性? 举例说明经济现象中序列相关性的存在。检验序列相关性的方法思路是什么? 熟悉 D.W.统计量的计算方法和查表判断。

4-10. 什么是多重共线性? 产生多重共线性的经济背景是什么? 多重共线性的危害是什么? 为什么会造成这些危害? 检验多重共线性的方法思路是什么? 有哪些克服方法?

4-11. 随机解释变量的来源有哪些? 随机解释变量可以造成哪些结果?

4-12. 当模型中出现随机解释变量时, 最小二乘估计量具有什么特征?

4-13. 试比较说明普通最小二乘法与加权最小二乘法的区别与联系。

4-14. 估计量的渐近统计性质的含义是什么？什么是渐近无偏性？

4-15. 什么是估计的一致性？证明对于工具变量法的估计量 $\hat{\beta}$ 是 β 的一致估计。

4-16. 为什么回归残差序列可以作为检验线性回归模型误差项的各种问题的基础？

4-17. 对于线性回归模型： $Y_t = \beta_0 + \beta_1 X_t + u_t$ ，已知 u 为一阶自回归形式： $u_t = \rho u_{t-1} + \varepsilon_t$ ，

要求：证明 ρ 的估计值为： $\hat{\rho} \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$

4-18. 证明下面方程中的误差项 v_i 是同方差的。

$$\frac{Y_i}{X_i} = \beta_1 \left(\frac{1}{X_i} \right) + \beta_2 \left(\frac{u_i}{X_i} \right) = \beta_1 \left(\frac{1}{X_i} \right) + \beta_2 + v_i, \quad \text{其中: } v_i = u_i / X_i$$

(三) 基本计算类题型

4-19. 某上市公司的子公司的年销售额 Y_t 与其总公司年销售额 X_t 的观测数据如下表：

序号	X	Y	序号	X	Y
1	127.3	20.96	11	148.3	24.54
2	130.0	21.40	12	146.4	24.30
3	132.7	21.96	13	150.2	25.00
4	129.4	21.52	14	153.1	25.64
5	135.0	22.39	15	157.3	26.36
6	137.1	22.76	16	160.7	26.98
7	141.2	23.48	17	164.2	27.52
8	142.8	23.66	18	165.6	27.78
9	145.5	24.10	19	168.7	28.24
19	145.3	24.01	20	171.7	28.78

要求：

(1)用最小二乘法估计 Y_t 关于 X_t 的回归方程；

(2)用 D.W.检验分析随机项的一阶自相关性；

(3)用 Durbin 两步法估计回归模型的参数；

(4)直接用差分法估计回归模型的参数。

4-20. 下表是被解释变量 Y 及解释变量 X_1 、 X_2 、 X_3 、 X_4 的时间序列观测值：

Y	6.0	6.0	6.5	7.1	7.2	7.6	8.0	9.0	9.0	9.3
X_1	40.1	40.3	47.5	49.2	52.3	58.0	61.3	62.5	64.7	66.8
X_2	5.5	4.7	5.2	6.8	7.3	8.7	10.2	14.1	17.1	21.3

X ₃	108	94	108	100	99	99	101	97	93	102
X ₄	63	72	86	100	107	111	114	116	119	121

要求：

- (1) 采用适当的方法检验多重共线性；
- (2) 多重共线性对参数估计值有何影响？
- (3) 用修正 Frisch 法确定一个较好的回归模型。

4-21. 下表是某种商品的需求量、价格以及消费者收入的统计资料：

年份	1	2	3	4	5	6	7	8	9	10
需求量 Y	3.5	4.3	5.0	6.0	7.0	9.0	8.0	10	12	14
价格 X ₁	16	13	10	7	7	5	4	3	3.5	2
收入 X ₂	15	20	30	42	50	54	65	72	85	90

要求：

- (1) 检验 X₁ 和 X₂ 是否存在严重的多重共线性？
- (2) 如何解决或减轻多重共线性的影响，并给出这一问题的回归方程。

4-22. 对于模型： $Y_t = \beta_1 + \beta_2 X_t + u_t$

要求：

- (1) 如果用变量的一次差分估计该模型，采用何种自相关形式？
- (2) 用差分估计时，并不删除截距，其含义是什么？

- (3) 假设模型存在一阶自相关，如果用 OLS 法估计，试证明其估计式： $\beta_2 = \frac{\sum x_i y_i}{\sum x_i^2}$ 仍然

是无偏的，式中的 $x_i = X_i - \bar{X}$ ， $y_i = Y_i - \bar{Y}$ 。

- (4) 试证明 $Var(\beta_2) = \sigma^2 \frac{1}{\sum x_i^2}$ 不是有效的。

4-23. 某国的政府税收 T（单位：百万美元）、国内生产总值 GDP（单位：10 亿美元）和汽车数量 Z（单位：百万辆）的观测数据如下表所示：

序号	T	GDP	Z
1	3	4	5
2	2	1	2
3	5	7	6
4	6	8	7
5	4	5	5
6	5	7	6

7	7	8	6
8	9	11	7
9	8	10	7

要求：试以汽车数量 Z 作为国内生产总值 GDP 的工具变量，估计税收函数：

$$T_t = \beta_0 + \beta_1 GDP_t + \varepsilon_t$$

4-24. 继续习题 3-21 的讨论。问题如下：

(1) 假定做 GMAT 分数对 GPA 的回归分析，并且发现两变量之间显著正相关。那么，你对多重共线性问题有何看法？

(2) 对习题 3-21 的 (1) 建立方差 (ANOVA) 分析表并检验假设：所有偏回归系数均为零。

(3) 用 R^2 值，对本题 (2) 建立 ANOVA 表进行分析。

4-25. 如果解释变量之间的相关系数为 0，则称它们是正交的。对于模型：

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

若 X_1 与 X_2 是正交的，证明下列结论：

(1) 多元线性回归的最小二乘估计量 $\hat{\beta}_1$ 、 $\hat{\beta}_2$ 分别等于 Y 对 X_1 、 Y 对 X_2 的一元线性回归的最小二乘估计量；

(2) 多元回归的回归平方和为两个一元回归的回归平方和的和。

4-26. 假设 Y 为内生变量， X 为外生变量，以下各组方程中哪些方程可以用 Durbin—Watson 方法检验一阶自相关：

$$(1) Y_{1t} = \alpha_1 X_{1t} + u_{1t}$$

$$Y_{2t} = \alpha_2 X_{2t} + \beta Y_{1t} + u_{2t}$$

$$(2) Y_{1t} = \beta_1 Y_{t-1} + u_{1t}$$

$$Y_{2t} = \beta_2 Y_{(t-1)} + \alpha_2 X_{2(t-1)} + u_{2t}$$

$$(3) Y_{1t} = \alpha_1 X_{1t} + u_{1t}$$

$$Y_{1t} = \alpha_2 X_{2t} + u_{2t}$$

4-27. 有 5 个解释变量的多元线性回归模型，用容量为 93 的样本数据进行回归分析。若根据回归残差序列计算的 D.W. 值为 1.1，应得出什么结论？若 D.W. 值为 2.35 呢？

4-28. 若已知线性回归模型 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ 的误差项的方差为 $\sigma_i^2 = 2X_{1i} - 3$ ，

问处理该模型的方法是什么？

4-29. 一个两变量线性回归模型的回归残差序列如下表所示：

n	残差 e	n	残差 e	n	残差 e
1	0.013	8	-0.082	15	0.198
2	0.054	9	-0.053	16	0.103
3	-0.014	10	0.041	17	0.000
4	-0.042	11	-0.151	18	-0.063
5	-0.078	12	-0.054	19	-0.058
6	-0.056	13	0.042		
7	0.083	14	0.117		

要求：请分析该模型的误差项是否存在什么问题？若存在一些问题，说明有哪些处理方法可以考虑？

4-30. 在研究生产中的劳动在增加值中所占的份额（即劳动份额）的变动时，有以下模型：

模型 A: $Y_t = \beta_0 + \beta_1 t + u_t$

模型 B: $Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$

其中，Y 为劳动份额，t 为劳动时间。根据该研究时期内的 15 年数据进行参数估计，得到模型结果为：

模型 A: $\hat{Y}_t = 0.4528 - 0.0041t$ $R^2 = 0.5284$ $D.W. = 0.8252$
(-3.9608)

模型 B: $\hat{Y}_t = 0.4786 - 0.0127t + 0.0005t^2$
(-3.2724) (2.7777)

$R^2 = 0.6629$ $D.W. = 1.82$

其中：括号中的数字是 t 检验值。

要求：（1）模型 A 中有没有自相关？模型 B 呢？

（2）如何解释自相关的存在？

（3）你会怎样区分“纯粹”自相关和模型形式设定错误？

四、习题解答

4-1. 答:

(1)异方差性指对于不同的样本值, 随机扰动项的方差不再是常数, 而是互不相同的。

(2)序列相关性指对于不同的样本值, 随机扰动项之间不再是完全相互独立, 而是存在某种相关性。

(3) 多重共线性指两个或多个解释变量之间不再彼此独立, 而是出现了相关性。

(4)偏回归系数指: 在三变量线性回归模型中, 当其中一个解释变量为常量时, 另一个解释变量对被解释变量均值的影响。

(5)完全多重共线性指: 在有多解释变量模型中, 其中一个变量可以表示为其他多个变量的完全线性函数, 即 $X_1 = B_2X_2 + B_3X_3 + \cdots + B_kX_k$, 其中至少有一个 $B_i \neq 0, (i = 2, 3, \cdots, k)$, X_1 与等式右边线性组合的相关系数为 1, 则这种情况被称为完全多重共线性。在此情况下, 不能估计解释变量各自对被解释变量的影响。

(6)不完全多重共线性指: 在实际经济活动中, 多个解释变量之间存在多重共线性问题, 但 X_1 与等式右边线性组合的相关系数不为 1。

(7)随机解释变量指: 在现实经济现象中, 解释变量是不可控的, 即解释变量的观测值具有随机性, 并且与模型的随机误差项有相关关系, 这样的解释变量称为随机解释变量。

(8)差分法是一类克服序列相关性的有效方法。它是将原计量经济模型变换为差分模型, 分为一阶差分法和广义差分法。

(9)广义最小二乘法 (GLS) 即最具有普遍意义的最小二乘法。

(10)D.W.检验: 全称杜宾-瓦森检验, 适用于一阶自相关的检验。该法构造一个统计量:

$$D.W. = \frac{\sum_{i=2}^n (\tilde{e}_i - \tilde{e}_{i-1})^2}{\sum_{i=1}^n \tilde{e}_i^2}, \text{ 计算该统计量的值, 根据样本容量 } n \text{ 和解释变量数目 } k \text{ 查 D.W. 分布表, 得到临界值 } d_l \text{ 和 } d_u, \text{ 然后按照判断准则考察计算得到的 D.W. 值, 以判断模型的自相关状态。}$$

4-2. 答:

(1)错。当存在异方差情况下，OLS 法估计量是无偏的但不具有有效性。

(2)对。如果存在异方差，通常使用的 t 检验和 F 检验是无效的。

(3) 错。实际情况可能是高估也可能是低估。

(4)对。通过将残差对其相应的观察值描图，了解变量与残差之间是否存在可以观察到的系统模式，就可以判断数据中是否存在异方差。

(5)错。当存在序列相关时，OLS 法估计量是无偏的但不具有有效性。

对。即假设误差项之间是完全正序列相关的，这样广义差分方程就转化为一阶差分方程。

(7)对。

(8)对。

(9)错。仍是无偏的。

4-3. 答：由于异方差的存在，使得：(1)OLS 估计量仍是线性无偏但不再具有最小方差，即不再有效；相应的置信区间和 t 检验、F 检验都是不可靠的。

4-4. 答：在存在 AR(1)自相关的情况下，使用广义最小二乘法能够产生 BLUE 估计量。具体步骤简述如下：

4-5. 答：在存在 AR(1)的情况下，估计自相关参数 ρ 有下述几种方法：

4-6. 答：存在；不存在；不存在；存在；存在。

4-7. 答：

$$\begin{aligned} \text{(1)模型两边同时除以 } x_{1t} \text{ 进行变换，得：} & \frac{y_t}{x_{1t}} = \frac{\alpha_0}{x_{1t}} + \alpha_1 + \alpha_2 \frac{x_{2t}}{x_{1t}} + \frac{u_t}{x_{1t}} \\ & = \alpha_0 \left(\frac{1}{x_{1t}} \right) + \alpha_1 + \alpha_2 \left(\frac{x_{2t}}{x_{1t}} \right) + v_t \end{aligned}$$

其中： $v_t = \frac{u_t}{x_{1t}}$ ，可以证明误差项 $v_t = \frac{u_t}{x_{1t}}$ 是同方差的。证明如下：

$$\text{已知： } v_t = \frac{u_t}{x_{1t}}, \quad v_t^2 = \frac{u_t^2}{x_{1t}^2}, \quad E(v_t^2) = E\left(\frac{u_t^2}{x_{1t}^2}\right) = E\left(\frac{\sigma^2 x_{1t}^2}{x_{1t}^2}\right) = E(\sigma^2) = \sigma^2 \quad (\text{根据已知条件}$$

σ^2 为常数)，证得变换后的误差项是同方差的。

(2)

4-8. 答：对于模型 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i \quad (i = 1, 2, \cdots, n)$ ，如果出现 $\text{Var}(u_i) = \sigma_i^2, (i = 1, 2, \cdots, n)$ ，即对于不同的样本点，随机误差项的方差不再是常数，而

且互不相同，则认为出现了异方差性。在现实经济运行中，异方差性经常出现，尤其是采用截面数据作样本的计量经济学问题。如：工业企业的研究与发展费用支出同企业的销售和利润之间关系的函数模型；服装需求量与季节、收入之间关系的函数模型；个人储蓄量与个人可支配收入之间关系的函数模型等。检验异方差性的思路即检验随机误差项的方差与解释变量观察值之间是否存在相关性。

4-9. 答：对于模型 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$ ($i = 1, 2, \dots, n$)，如果出现 $Cov(u_i, u_j) \neq 0, (i \neq j, i, j = 1, 2, \dots, n)$ ，即对于不同的样本点，随机误差项之间不再是完全互相独立，而是存在某种相关性，则认为出现了序列相关性。在现实经济运行中，序列相关性经常出现，尤其是采用时间序列数据作样本的计量经济学问题。如：以时间序列数据作为样本建立的行业生产函数模型；以时间序列数据作样本建立的居民总消费函数模型等。检验序列相关性的方法思路即先采用 OLS 法估计模型，以求得随机误差项的“近似估计量 $\tilde{e}_i$ ”，然后通过分析这些“近似估计量 $\tilde{e}_i$ ”之间的相关性以达到判断随机误差项是否具有序列相关性的目的。

4-10. 答：对于模型 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$ ($i = 1, 2, \dots, n$)，如果某两个或多个解释变量之间出现了相关性，则称为多重共线性。