

# 练习题

## 一、判断题

1. 线性回归模型中，解释变量是原因，被解释变量是结果。（）
  2. 多元回归模型统计显著是指模型中每个变量都是统计显著的。（）
  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. 总体回归线是当解释变量取给定值时因变量的条件均值的轨迹。（）
  5. 线性回归是指解释变量和被解释变量之间呈现线性关系。（）
  6. 判定系数  $R^2$  的大小不受到回归模型中所包含的解释变量个数的影响。（）
  7. 多重共线性是一种随机误差现象。（）
  8. 当存在自相关时，OLS 估计量是有偏的并且也是无效的。（）
  9. 在异方差的情况下，OLS 估计量误差放大的原因是从属回归的  $R^2$  变大。（）
  10. 任何两个计量经济模型的  $R^2$  都是可以比较的。（）
1. 随机误差项  $u_i$  和残差项  $e_i$  是一回事。（）
  2. 给定显著性水平  $\alpha$  及自由度，若计算得到的  $|t|$  值超过临界的  $t$  值，我们将接受零假设（）
  3. 利用 OLS 法求得的样本回归直线  $\hat{Y}_i = b_1 + b_2 X_i$  通过样本均值点  $(\bar{X}, \bar{Y})$ 。（）
  4. 判定系数  $R^2 = TSS/ESS$ 。（）
  5. 整个多元回归模型在统计上是显著的意味着模型中任何一个单独的变量均是统计显著的。（）
  6. 双对数模型的  $R^2$  值可以与对数线性模型的相比较，但不能与线性对数模型的相比较。（）
  7. 为了避免陷入虚拟变量陷阱，如果一个定性变量有  $m$  类，则要引入  $m$  个虚拟变量。（）
  8. 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。（ ）

9. 识别的阶条件仅仅是判别模型是否可识别的必要条件而不是充分条件。  
( )
10. 如果零假设  $H_0: B_2=0$ , 在显著性水平 5%下不被拒绝, 则认为  $B_2$  一定是 0。  
( )
1. 回归分析用来处理一个因变量与另一个或多个自变量之间的因果关系。( )
2. 拟合优度  $R^2$  的值越大, 说明样本回归模型对总体回归模型的代表性越强。( )
3. 线性回归是指解释变量和被解释变量之间呈现线性关系。( )
4. 引入虚拟变量后, 用普通最小二乘法得到的估计量仍是无偏的。( )
5. 多重共线性是总体的特征。( )
6. 任何两个计量经济模型的  $R^2$  都是可以比较的。( )
7. 异方差会使 OLS 估计量的标准误差高估, 而自相关会使其低估。( )
8. 杜宾-瓦尔森检验能够检验出任何形式的自相关。( )
9. 异方差值存在于横截面数据中, 而自相关值存在于时间序列数据中。( )
10. 内生变量的滞后值仍然是内生变量。( )
- 1、随机变量的条件均值与非条件均值是一回事。( )
- 2、线性回归模型意味着变量是线性的。( )
- 3、 $ESS = TSS + RSS$ 。( )
- 4、对于多元回归模型, 如果联合检验结果是统计显著的则意味着模型中任何一个单独的变量均是统计显著的。( )
- 5、双对数模型中的斜率表示因变量对自变量的弹性。( )
- 6、为了避免陷入虚拟变量陷阱, 如果一个定性变量有  $m$  类, 则要引入  $m$  个虚拟变量。  
( )
- 7、如果回归模型违背了同方差假定, 最小二乘估计量是有偏无效的。( )
- 8、在存在接近多重共线性的情况下, 回归系数的标准差会趋于变小, 相应的  $t$  值会趋于变大。( )
- 9、在任何情况下 OLS 估计量都是待估参数的最优线性无偏估计。( )

10、一个联立方程模型中的外生变量在另一个联立方程模型中可能是内生变量。( )

## 二、选择题

1. 在同一时间不同统计单位的相同统计指标组成的数据组合，是( )

- A. 原始数据    B. Pool 数据    C. 时间序列数据    D. 截面数据

2. 下列模型中属于非线性回归模型的是( )

- A.  $Y = \beta_0 + \beta_1 \ln X + u$     B.  $Y = \beta_0 + \beta_1 X + \beta_2 Z + u$   
C.  $Y = \beta_0 + X^{\beta_1} + u$     D.  $Y = \beta_0 + \beta_1 / X + u$

3. 半对数模型  $Y = \beta_0 + \beta_1 \ln X + u$  中，参数  $\beta_1$  的含义是( )

- A. X 的绝对量变化，引起 Y 的绝对量变化  
B. Y 关于 X 的边际变化  
C. X 的相对变化，引起 Y 的期望值绝对量变化  
D. Y 关于 X 的弹性

4. 模型中其数值由模型本身决定的变量是( )

- A、外生变量    B、内生变量    C、前定变量    D、滞后变量

5. 在模型  $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$  的回归分析结果报告中，F 统计量的

p 值 = 0.0000，则表明( )

- A. 解释变量  $X_{2t}$  对  $Y_t$  的影响是显著的  
B. 解释变量  $X_{3t}$  对  $Y_t$  的影响是显著的  
C. 解释变量  $X_{2t}$  和  $X_{3t}$  对  $Y_t$  的联合影响是显著的  
D. 解释变量  $X_{2t}$  和  $X_{3t}$  对  $Y_t$  的联合影响不显著

6. 根据样本资料估计人均消费支出 Y 对人均收入 X 的回归模型为

$\ln \hat{Y}_i = 2.00 + 0.75 \ln X_i$ ，这表明人均收入每增加 1%，人均消费支出将增加

( )

- A. 0.2%    B. 0.75%    C. 2%    D. 7.5%

7. 如果回归模型违背了同方差假定，最小二乘估计量是( )

- A. 无偏的，非有效的    B. 有偏的，非有效的

- C. 无偏的，有效的      D. 有偏的，有效的
8. 在回归模型满足 DW 检验的前提条件下，当  $d$  统计量等于 2 时，表明（    ）
- A. 存在完全的正自相关      B. 存在完全的负自相关
- C. 不存在自相关      D. 不能判定
9. 将一年四个季度对被解释变量的影响引入到包含截距项的回归模型当中，则需要引入虚拟变量的个数为（    ）
- A. 5      B. 4      C. 3      D. 2
10. 在联立方程结构模型中，对模型中的每一个随机方程单独使用普通最小二乘法得到的估计参数是（    ）
- A. 有偏但一致的    B. 有偏且不一致的    C. 无偏且一致的    D. 无偏但不一致的

### 三. 简答题

1. 计量经济模型分析经济问题的基本步骤。
2. 举例说明如何引进加法模式和乘法模式建立虚拟变量模型。
3. 下面是我国 1990-2003 年 GDP 对 M1 之间回归的结果。

$$\begin{array}{lcl} \ln(GDP) = 1.37 & + & 0.76 \ln(M1) \\ \text{se} & (0.15) & ( \quad ) \\ t & ( \quad ) & ( 23 ) \end{array}$$

$$P(|t| > 1.782) = 0.05, \text{自由度} = 12;$$

- (1) 求出空白处的数值，填在括号内。
- (2) 系数是否显著，给出理由。
4. 试述异方差的后果及其补救措施。
5. 多重共线性的后果及修正措施。
6. 试述 D-W 检验的适用条件及其检验步骤？

7. 下面是利用 1970-1980 年美国数据得到的回归结果。其中 Y 表示美国咖啡消费（杯/日. 人），X 表示平均零售价格（美元/磅）。（15 分） 注： $t_{\alpha/2}(9) = 2.262$ ， $t_{\alpha/2}(10) = 2.228$

$$\begin{aligned}\hat{Y}_t &= 2.6911 - 0.4795X_t \\ se &= (0.1216) \quad ( \quad ) \\ t\text{值} &= ( \quad ) \quad 42.06 \quad R^2 = 0.6628\end{aligned}$$

(1) 写空白处的数值。

2. 对模型中的参数进行显著性检验。四

8、考虑下面的模型： $Y_t = B_0 + B_1X_t + B_2D_{2t} + B_3D_{3t} + B_4D_{4t} + u_t$  其中，Y 表示大学教师的年薪收入，X 表示工龄。为了研究大学教师的年薪是否受到性别、学历的影响。按照下面的方式引入虚拟变量：（15 分）

$$D_2 = \begin{cases} 1, & \text{男教师} \\ 0, & \text{女教师} \end{cases} \quad D_3 = \begin{cases} 1, & \text{硕士} \\ 0, & \text{其他} \end{cases} \quad D_4 = \begin{cases} 1, & \text{博士} \\ 0, & \text{其他} \end{cases}$$

(1) 基准类是什么？

(2) 解释各系数所代表的含义，并预期各系数的符号。

(3) 若  $B_4 > B_3$ ，你得出什么结论？

9、下表给出了三变量模型的回归的结果：

方差来源	平方和	自由度 (d.f)	平方和的均值 (MSS)
来自回归(ESS)	106.58		
来自残差(RSS)			
总离差(TSS)	108.38	19	_____

注：保留 3 位小数，可以使用计算器。在 5% 的显著性水平下，本题的  $F_{\alpha} = 4.45$ 。

1. 完成上表中空白处内容。

2. 求  $R^2$  与  $\bar{R}^2$ 。

3. 利用 F 统计量检验  $X_2$  和  $X_3$  对 Y 的联合影响，写出简要步骤。

10、根据所给数据利用做小二乘法计算参数。

## 八、应用题

利用美国 1980-1995 年间人均消费支出（PCE）和人均可支配收入（PDPI）的数据，得到

了如下回归分析结果：

Dependent Variable: LOG(PCE)

Method: Least Squares

Date: 06/09/05 Time: 23:43

Sample: 1980 1995

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(PDPI)	1.205281	0.028891	41.71870	0.0000
C	-2.092664	0.281286	-7.439640	0.0000
R-squared	0.992020	Mean dependent var	9.641839	
Adjusted R-squared	0.991450	S.D. dependent var	0.096436	
S.E. of regression	0.008917	Akaike info criterion	-6.485274	
Sum squared resid	0.001113	Schwarz criterion	-6.388701	
Log likelihood	53.88219	F-statistic	1740.450	
Durbin-Watson stat	0.81	Prob(F-statistic)	0.000000	

若  $k = 2, n = 19, d_L = 1.074, d_U = 1.536$ , 显著性水平  $\alpha = 0.05$

- (1) 根据以上结果，写出回归分析结果报告。（10 分）
- (2) 对模型中解释变量系数  $B_2$  进行显著性检验。（10 分）
- (3) 如何解释解释变量的系数和综合判定系数？（10 分）
- (5) 模型可能存在什么问题？如何检验？
- (6) 如何就模型中所存在的问题，对模型进行改进？

# 联系题一答案

## 一、判断题（20 分）

1. 线性回归模型中，解释变量是原因，被解释变量是结果。（F）
2. 多元回归模型统计显著是指模型中每个变量都是统计显著的。（F）
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。（F）
4. 总体回归线是当解释变量取给定值时因变量的条件均值的轨迹。（Y）
5. 线性回归是指解释变量和被解释变量之间呈现线性关系。（F）
6. 判定系数  $R^2$  的大小不受回归模型中所包含的解释变量个数的影响。（F）
7. 多重共线性是一种随机误差现象。（F）
8. 当存在自相关时，OLS 估计量是有偏的并且也是无效的。（F）
9. 在异方差的情况下，OLS 估计量误差放大的原因是从属回归的  $R^2$  变大。（F）
10. 任何两个计量经济模型的  $R^2$  都是可以比较的。（F）

## 二、简答题（10）

1. 计量经济模型分析经济问题的基本步骤。（4 分）

答：

- 1) 经济理论或假说的陈述
- 2) 收集数据
- 3) 建立数理经济学模型
- 4) 建立经济计量模型
- 5) 模型系数估计和假设检验
- 6) 模型的选择
- 7) 理论假说的选择
- 8) 经济学应用

2. 举例说明如何引进加法模式和乘法模式建立虚拟变量模型。（6 分）

答案：设 Y 为个人消费支出；X 表示可支配收入，定义

$$D_{2t} = \begin{cases} 1 & \text{2季度} \\ 0 & \text{其他} \end{cases} \quad D_{3t} = \begin{cases} 1 & \text{3季度} \\ 0 & \text{其他} \end{cases} \quad D_{4t} = \begin{cases} 1 & \text{4季度} \\ 0 & \text{其他} \end{cases}$$

如果设定模型为

$$Y_t = B_1 + B_2 D_{2t} + B_3 D_{3t} + B_4 D_{4t} + B_5 X_t + u_t$$

此时模型仅影响截距项，差异表现为截距项的和，因此也称为加法模型。

如果设定模型为

$$Y_t = B_1 + B_2 D_{2t} + B_3 D_{3t} + B_4 D_{4t} + B_5 X_t \\ + B_6 (D_{2t} X_t) + B_7 (D_{3t} X_t) + B_8 (D_{4t} X_t) + u_t$$

此时模型不仅影响截距项，而且还影响斜率项。差异表现为截距和斜率的双重变化，因此也称为乘法模型。

三. 下面是我国 1990-2003 年 GDP 对 M1 之间回归的结果。(5 分)

$$\ln(GDP) = 1.37 + 0.76 \ln(M1) \\ \text{se} \quad (0.15) \quad (0.033) \\ \text{t} \quad (9.13) \quad (23)$$

$$P(|t| > 1.782) = 0.05, \text{自由度} = 12;$$

1. 求出空白处的数值，填在括号内。(2 分)

2. 系数是否显著，给出理由。(3 分)

答：根据 t 统计量，9.13 和 23 都大于 5% 的临界值，因此系数都是统计显著的。

四. 试述异方差的后果及其补救措施。(10 分)

答案：

后果：OLS 估计量是线性无偏的，不是有效的，估计量方差的估计有偏。建立在 t 分布和 F 分布之上的置信区间和假设检验是不可靠的。

补救措施：加权最小二乘法 (WLS)

1. 假设  $\sigma_i^2$  已知，则对模型进行如下变换：

$$\frac{Y_i}{\sigma_i} = \frac{B_1}{\sigma_i} + B_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

2. 如果  $\sigma_i^2$  未知

(1) 误差与  $X_i$  成比例：平方根变换。

$$\frac{Y_i}{\sqrt{X_i}} = \frac{B_1}{\sqrt{X_i}} + B_2 \frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}}$$

可见，此时模型同方差，从而可以利用 OLS 估计和假设检验。

(2) 误差方差和  $X_i^2$  成比例。即  $E(u_i^2) = \sigma^2 X_i^2$



$$\frac{Y_i}{X_i} = \frac{B_1}{X_i} + B_2 \frac{X_i}{X_i} + \frac{u_i}{X_i}$$

3. 重新设定模型：

五. 多重共线性的后果及修正措施。(10 分)

1) 对于完全多重共线性，后果是无法估计。

对于高度多重共线性，理论上不影响 OLS 估计量的最优线性无偏性。但对于个别样本的估计量的方差放大，从而影响了假设检验。

实际后果：联合检验显著，但个别系数不显著。估计量的方差放大，置信区间变宽，t 统计量变小。

对于样本内观测值得微小变化极敏感。某些系数符号可能不对。难以解释自变量对应变量的贡献程度。

2) 补救措施：剔出不重要变量；增加样本数量；改变模型形式；改变变量形式；利用先验信息。

六. 试述 D-W 检验的适用条件及其检验步骤？(10 分)

答案：

使用条件：

1) 回归模型包含一个截距项。

2) 变量 X 是非随机变量。

3) 扰动项的产生机制： $u_t = \rho u_{t-1} + v_t$   $-1 \leq \rho \leq 1$ 。

4) 因变量的滞后值不能作为解释变量出现在回归方程中。

检验步骤

1) 进行 OLS 回归，并获得残差。

2) 计算 D 值。

3) 已知样本容量和解释变量个数，得到临界值。

4) 根据下列规则进行判断：

零假设	决策	条件
无正的自相关	拒绝	$0 < d < d_L$
无正的自相关	无法确定	$d_L < d < d_U$
无负的自相关	拒绝	$4 - d_L < d < 4$
无负的自相关	无法决定	$4 - d_U < d < 4 - d_L$
无正的或者负的自相关	接受	$d_U < d < 4 - d_U$

七. (15 分) 下面是宏观经济模型

$$M_t = C(1) * P_t + C(2) * Y_t + C(3) * I_t + C(4) * M_{t-1} + u_t^D$$

$$I_t = C(5) * M_t + C(6) * Y_t + u_t^C$$

$$Y_t = C(7) * I_t + u_t^A$$

变量分别为货币供给  $M$ 、投资  $I$ 、价格指数  $P$  和产出  $Y$ 。

1. 指出模型中哪些是内生变量，哪些是外生变量。（5 分）

答：内生变量为货币供给  $M_t$ 、投资  $I_t$  和产出  $Y_t$ 。

外生变量为滞后一期的货币供给  $M_{t-1}$  以及价格指数  $P_t$

2. 对模型进行识别。（4 分）

答：根据模型识别的阶条件

方程（1）： $k=0 < m-1=2$ ，不可识别。

方程（2）： $k=2=m-1$ ，恰好识别。

方程（3）： $k=2=m-1$ ，恰好识别。

3. 指出恰好识别方程和过度识别方程的估计方法。（6 分）

答：

对于恰好识别方程，采用间接最小二乘法。首先建立简化方程，之后对简化方程进行最小二乘估计。

对于过度识别方程，采用两阶段最小二乘法。首先求替代变量（工具变量），再把这个工具变量作为自变量进行回归。

#### 八、（20 分）应用题

为了研究我国经济增长和国债之间的关系，建立回归模型。得到的结果如下：

Dependent Variable: LOG(GDP)

Method: Least Squares

Date: 06/04/05 Time: 18:58

Sample: 1985 2003

Included observations: 19

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.03	0.14	43.2	0
LOG(DEBT)	0.65	0.02	32.8	0
R-squared	0.981	Mean dependent		10.53
		var		
Adjusted R-squared	0.983	S.D. dependent var		0.86
S.E. of regression	0.11	Akaike info criterion		-1.46
Sum squared resid	0.21	Schwarz criterion		-1.36
Log likelihood	15.8	F-statistic		1075.5

若  $k=1, n=19, d_L=1.074, d_U=1.536$ , 显著性水平  $\alpha=0.05$

其中, GDP 表示国内生产总值, DEBT 表示国债发行量。

(1) 写出回归方程。(2 分)

答:  $\text{Log}(\text{GDP}) = 6.03 + 0.65 \text{ LOG}(\text{DEBT})$

(2) 解释系数的经济学含义? (4 分)

答:

截距项表示自变量为零时, 因变量的平均期望。不具有实际的经济学含义。

斜率系数表示 GDP 对 DEBT 的不变弹性为 0.65。或者表示增发 1% 国债, 国民经济增长 0.65%。

(3) 模型可能存在什么问题? 如何检验? (7 分)

答:

可能存在序列相关问题。

因为  $d.w = 0.81$  小于  $d_L = 1.074$ , 因此落入正的自相关区域。由此可以判定存在序列相关。

(4) 如何就模型中所存在的问题, 对模型进行改进? (7 分)

答: 利用广义最小二乘法。根据  $d.w = 0.81$ , 计算得到  $\rho = 0.6$ , 因此回归方程滞后一期后, 两边同时乘以 0.6, 得

$$0.6 \log(\text{GDP}_{t-1}) = 0.4B_1 + 0.6B_2 \log(\text{DEBT}_{t-1}) + 0.6u_{t-1}$$

方程

$$\log(\text{GDP}_t) = B_1 + B_2 \log(\text{DEBT}_t) + u_t$$

减去上面的方程, 得到

$$\log(\text{GDP}_t) - 0.6 \log(\text{GDP}_{t-1}) = 0.6B_1 + B_2 (\log(\text{DEBT}_t) - 0.6 \log(\text{DEBT}_{t-1})) + v_t$$

利用最小二乘估计, 得到系数。

# 练习题二答案

## 一、判断正误（20 分）

1. 随机误差项  $u_i$  和残差项  $e_i$  是一回事。（ F ）
2. 给定显著性水平  $\alpha$  及自由度，若计算得到的  $|t|$  值超过临界的  $t$  值，我们将接受零假设。（ F ）
3. 利用 OLS 法求得的样本回归直线  $\hat{Y}_i = b_1 + b_2 X_i$  通过样本均值点  $(\bar{X}, \bar{Y})$ 。（ T ）
4. 判定系数  $R^2 = TSS/ESS$ 。（ F ）
5. 整个多元回归模型在统计上是显著的意味着模型中任何一个单独的变量均是统计显著的。（ F ）
6. 双对数模型的  $R^2$  值可以与对数线性模型的相比较，但不能与线性对数模型相比较。（ T ）
7. 为了避免陷入虚拟变量陷阱，如果一个定性变量有  $m$  类，则要引入  $m$  个虚拟变量。（ F ）
8. 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。（ T ）
9. 识别的阶条件仅仅是判别模型是否可识别的必要条件而不是充分条件。（ T ）
10. 如果零假设  $H_0: B_2=0$ ，在显著性水平 5% 下不被拒绝，则认为  $B_2$  一定是 0。（ F ）

## 二、以一元回归为例叙述普通最小二乘回归的基本原理。(10 分)

解：依据题意有如下的一元样本回归模型：

$$Y_i = b_1 + b_2 X_i + e_i \quad (1)$$

普通最小二乘原理是使得残差平方和最小，即

$$\min Q = \min \sum e_i^2 = \min \sum (Y_i - b_1 - b_2 X_i)^2 \quad (2)$$

根据微积分求极值的原理，可得

$$\frac{\partial Q}{\partial b_1} = 0 \Leftrightarrow \frac{\partial Q}{\partial b_1} = -2 \sum (Y_i - b_1 - b_2 X_i) = 0 \quad (3)$$

$$\frac{\partial Q}{\partial b_2} = 0 \Leftrightarrow \frac{\partial Q}{\partial b_2} = -2 \sum (Y_i - b_1 - b_2 X_i) X_i = 0 \quad (4)$$

将(3)和(4)式称为正规方程，求解这两个方程，我们可得到：

$$\begin{aligned} \sum Y_i &= nb_1 + b_2 \sum X_i \\ \sum Y_i X_i &= b_1 \sum X_i + b_2 \sum X_i^2 \end{aligned} \quad (5)$$

解得：

$$b_1 = \bar{Y} - b_2 \bar{X}$$

$$b_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

其中  $x_i = X_i - \bar{X}$ ,  $y_i = Y_i - \bar{Y}$ , 表示变量与其均值的离差。

三、下面是利用 1970-1980 年美国数据得到的回归结果。其中 Y 表示美国咖啡消费（杯/日·人），X 表示平均零售价格（美元/磅）。（15 分） 注： $t_{\alpha/2}(9) = 2.262$ ,  $t_{\alpha/2}(10) = 2.228$

$$\hat{Y}_t = 2.6911 - 0.4795 X_t$$

$$se = (0.1216) \quad (a)$$

$$t\text{值} = (b) \quad 42.06 \quad R^2 = 0.6628$$

1. 写空白处的数值啊  $a$ ,  $b$ 。（0.0114, 22.066）
2. 对模型中的参数进行显著性检验。
3. 解释斜率系数  $B_2$  的含义，并给出其 95% 的置信区间。

解：1. （0.0114, 22.066）

2.  $B_1$  的显著性检验： $t = 22.066 > t_{\alpha/2}(9) = 2.262$ ，所以  $B_1$  是显著的。

$B_2$  的显著性检验： $t = 42.06 > t_{\alpha/2}(9) = 2.262$ ，所以  $B_2$  是显著的。

3.  $B_2$  表示每磅咖啡的平均零售价格每上升 1 美元，每人每天的咖啡消费量减少 0.479 杯。

$$P(-2.262 \leq t \leq 2.262) = 0.95$$

$$P\left(-2.262 \leq \frac{b_2 - B_2}{se(b_2)} \leq 2.262\right) = 0.95$$

$$P(b_2 - 2.262se(b_2) \leq B_2 \leq b_2 + 2.262se(b_2)) = 0.95$$

$B_2$  的 95% 的置信区间为：

$$[-0.479 - 0.026, -0.479 + 0.026]$$

$$[-0.505454, -0.454]$$

四、若在模型： $Y_t = B_1 + B_2 X_t + u_t$  中存在下列形式的异方差： $\text{var}(u_t) = \sigma^2 X_t^3$ ，你如何估计参数  $B_1, B_2$ （10 分）

解：对于模型

$$Y_t = B_1 + B_2 X_t + u_t$$

存在下列形式的异方差： $\text{var}(u_t) = \sigma^2 X_t^3$ ，我们可以在(1)式左右两端同时除以  $\sqrt{X_t^3}$ ，可

得

$$\begin{aligned}\frac{Y_t}{\sqrt{X_t^3}} &= B_1 \frac{1}{\sqrt{X_t^3}} + B_2 \frac{X_t}{\sqrt{X_t^3}} + \frac{u_t}{\sqrt{X_t^3}} \\ &= B_1 \frac{1}{\sqrt{X_t^3}} + B_2 \frac{X_t}{\sqrt{X_t^3}} + v_t\end{aligned}\quad (2)$$

其中

$$v_t = \frac{u_t}{\sqrt{X_t^3}}$$

代表误差修正项，可以证明

$$\text{var}(v_t) = \text{var}\left(\frac{u_t}{\sqrt{X_t^3}}\right) = \frac{1}{X_t^3} \text{var}(u_t) = \frac{1}{X_t^3} \sigma^2 X_t^3 = \sigma^2$$

即  $v_t$  满足同方差的假定，对(2)式使用 OLS，即可得到相应的估计量。

五、考虑下面的模型： $Y_t = B_0 + B_1 X_t + B_2 D_{2t} + B_3 D_{3t} + B_4 D_{4t} + u_t$  其中， $Y$  表示大学教师的年薪收入， $X$  表示工龄。为了研究大学教师的年薪是否受到性别（男、女）、学历（本科、硕士、博士）的影响。按照下面的方式引入虚拟变量：（15 分）

$$D_2 = \begin{cases} 1, & \text{男教师} \\ 0, & \text{女教师} \end{cases} \quad D_3 = \begin{cases} 1, & \text{硕士} \\ 0, & \text{其他} \end{cases} \quad D_4 = \begin{cases} 1, & \text{博士} \\ 0, & \text{其他} \end{cases}$$

1. 基准类是什么？

2. 解释各系数所代表的含义，并预期各系数的符号。

3. 若  $B_4 > B_3$ ，你得出什么结论？解：1. 基准类为本科女教师。

2.  $B_1$  表示工龄对年薪的影响，即工龄每增加 1 单位，平均而言，年薪将增加  $B_1$  个单位。预期符号为正，因为随着年龄的增加，工资应该增加。

$B_2$  体现了性别差异。

$B_3$  和  $B_4$  体现了学历差异，预期符号为正。

3.  $B_4 > B_3$  说明，博士教师的年薪高于硕士教师的年薪。

六、什么是自相关？杜宾—瓦尔森检验的前提条件和步骤是什么？（15 分）

解：自相关，在时间（如时间序列数据）或者空间（如在截面数据中）上按顺序排列的序列的各成员之间存在着相关关系。在计量经济学中指回归模型中随机扰动项之间存在相关关系。用符号表示：

$$\text{cov}(u_i, u_j) = E(u_i u_j) \neq 0 \quad i \neq j$$

杜宾—瓦尔森检验的前提条件为：

- (1) 回归模型包括截距项。
- (2) 变量  $X$  是非随机变量。

(3) 扰动项  $u_t$  的产生机制是

$$u_t = \rho u_{t-1} + v_t \quad (-1 \leq \rho \leq 1, \text{表示自相关系数})$$

上述这个描述机制我们称为一阶自回归模型，通常记为 AR(1)。

(4) 在回归方程的解释变量中，不包括把因变量的滞后变量。即检验对于自回归模型是不使用的。

杜宾—瓦尔森检验的步骤为：

(1) 进行 OLS 的回归并获得  $e_t$ 。

(2) 计算 d 值。

(3) 给定样本容量  $n$  和解释变量  $k$  的个数，从临界值表中查得  $d_L$  和  $d_U$ 。

(4) 根据相应的规则进行判断。

七、考虑下面的联立方程模型：
$$\begin{cases} Q_t = A_1 + A_2 P_t + A_3 X_t + u_{1t} \\ Q_t = B_1 + B_2 P_t + u_{2t} \end{cases} \quad \text{其中，} P, Q \text{ 是内生变量，} X \text{ 是外生变量，} u \text{ 是随机误差项 (15 分)}$$

1、求简化形式回归方程？

2、判定哪个方程是可识别的（恰好或过度）？

3、对可识别方程，你将用哪种方法进行估计，并简述基本过程？

解 1.

$$P_t = \Pi_1 + \Pi_2 X_t + v_{1t}$$

其中：

$$\Pi_1 = \frac{B_1 - A_1}{A_2 - B_2}, \quad \Pi_2 = -\frac{A_3}{A_2 - B_2}, \quad v_{1t} = \frac{u_{2t} - u_{1t}}{A_2 - B_2} \quad (1)$$

$$Q_t = \Pi_3 + \Pi_4 X_t + v_{2t}$$

其中：

$$\Pi_3 = \frac{A_2 B_1 - A_1 B_2}{A_2 - B_2}, \quad \Pi_4 = -\frac{A_3 B_2}{A_2 - B_2}, \quad v_{1t} = \frac{A_2 u_{2t} - B_2 u_{1t}}{A_2 - B_2} \quad (2)$$

2. 根据阶判断条件， $m = 2$ ,

对于第一个方程， $k=0$ ， $k < m-1$ ，所以第一个方程不可识别。

对于第二个方程， $k=1$ ， $k = m-1$ ，所以第二个方程恰好识别。

3. 对于恰好识别的方程，可以采用二阶段最小二乘法，也可以使用间接最小二乘法。

下面将简单介绍间接最小二乘法的基本过程：

步骤 1：从结构方程导出简化方程；

步骤 2：对简化方程的每个方程用 OLS 方法回归；

步骤 3：利用简化方程系数的估计值求结构方程系数的估计值。

# 联系题三答案

## 一、判断正误（20 分）

1. 回归分析用来处理一个因变量与另一个或多个自变量之间的因果关系。（ F ）
2. 拟合优度  $R^2$  的值越大，说明样本回归模型对总体回归模型的代表性越强。（ T ）
3. 线性回归是指解释变量和被解释变量之间呈现线性关系。（ F ）
4. 引入虚拟变量后，用普通最小二乘法得到的估计量仍是无偏的。（ T ）
5. 多重共线性是总体的特征。（ F ）
6. 任何两个计量经济模型的  $R^2$  都是可以比较的。（ F ）
7. 异方差会使 OLS 估计量的标准误差高估，而自相关会使其低估。（ F ）
8. 杜宾—瓦尔森检验能够检验出任何形式的自相关。（ F ）
9. 异方差问题总是存在于横截面数据中，而自相关则总是存在于时间序列数据中。（ F ）
10. 内生变量的滞后值仍然是内生变量。（ F ）

## 二、选择题（20 分）

1. 在同一时间不同统计单位的相同统计指标组成的数据组合，是（ D ）  
A. 原始数据    B. Pool 数据    C. 时间序列数据    D. 截面数据
2. 下列模型中属于非线性回归模型的是（ C ）  
A.  $Y = \beta_0 + \beta_1 \ln X + u$     B.  $Y = \beta_0 + \beta_1 X + \beta_2 Z + u$   
C.  $Y = \beta_0 + X^{\beta_1} + u$     D.  $Y = \beta_0 + \beta_1 / X + u$
3. 半对数模型  $Y = \beta_0 + \beta_1 \ln X + u$  中，参数  $\beta_1$  的含义是（ C ）  
A. X 的绝对量变化，引起 Y 的绝对量变化  
B. Y 关于 X 的边际变化  
C. X 的相对变化，引起 Y 的期望值绝对量变化  
D. Y 关于 X 的弹性
4. 模型中其数值由模型本身决定的变量是（ B ）



A、外生变量 B、内生变量 C、前定变量 D、滞后变量

5. 在模型  $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$  的回归分析结果报告中,  $F$  统计量的  $p$  值 = 0.0000, 则表明 ( C )
- A. 解释变量  $X_{2t}$  对  $Y_t$  的影响是显著的  
B. 解释变量  $X_{3t}$  对  $Y_t$  的影响是显著的  
C. 解释变量  $X_{2t}$  和  $X_{3t}$  对  $Y_t$  的联合影响是显著的  
D. 解释变量  $X_{2t}$  和  $X_{3t}$  对  $Y_t$  的联合影响不显著
6. 根据样本资料估计人均消费支出  $Y$  对人均收入  $X$  的回归模型为  $\ln \hat{Y}_i = 2.00 + 0.75 \ln X_i$ , 这表明人均收入每增加 1%, 人均消费支出将增加 ( B )
- A. 0.2%    B. 0.75%    C. 2%    D. 7.5%
7. 如果回归模型违背了同方差假定, 最小二乘估计量是 ( A )
- A. 无偏的, 非有效的    B. 有偏的, 非有效的  
C. 无偏的, 有效的    D. 有偏的, 有效的
8. 在回归模型满足 DW 检验的前提条件下, 当  $d$  统计量等于 2 时, 表明 ( C )
- A. 存在完全的正自相关    B. 存在完全的负自相关  
C. 不存在自相关    D. 不能判定
9. 将一年四个季度对被解释变量的影响引入到包含截距项的回归模型当中, 则需要引入虚拟变量的个数为 ( C )
- A. 5    B. 4    C. 3    D. 2
10. 在联立方程结构模型中, 对模型中的每一个随机方程单独使用普通最小二乘法得到的估计参数是 ( B )
- A. 有偏但一致的    B. 有偏且不一致的    C. 无偏且一致的    D. 无偏但不一致的

三、下表给出了三变量模型的回归的结果: (10 分)

方差来源	平方和	自由度 (d.f)	平方和的均值 (MSS)
来自回归(ESS)	106.58	2	53.29
来自残差(RSS)	1.8	17	0.106

总离差(TSS)	108.38	19	_____
----------	--------	----	-------

注：保留 3 位小数，可以使用计算器。在 5% 的显著性水平下，本题的  $F_\alpha = 4.45$ 。

1. 完成上表中空白处内容。

2. 求  $R^2$  与  $\bar{R}^2$ 。

3. 利用 F 统计量检验  $X_2$  和  $X_3$  对  $Y$  的联合影响，写出简要步骤。

答案： 1. 见题

$$2. R^2 = \frac{ESS}{TSS} = \frac{106.58}{108.38} = 0.982$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0.982) \frac{19}{17} = 0.980$$

3. 可以利用  $F$  统计量检验  $X_2$  和  $X_3$  对  $Y$  的联合影响。

$$F = \frac{ESS/2}{RSS/17} = \frac{53.29}{0.106} = 502.736 \quad \left( \text{或} \quad F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \right)$$

因为  $F > F_\alpha = 4.45$ ， $X_2$  和  $X_3$  对  $Y$  的联合影响是显著的。

四、考虑下面的模型： $Y_t = B_0 + B_1 X_t + B_2 D_{2t} + B_3 D_{3t} + B_4 D_{4t} + u_t$  其中， $Y$  表示大学教师的年薪收入， $X$  表示工龄。为了研究大学教师的年薪是否受到性别、学历的影响。按照下面的方式引入虚拟变量：（10 分）

$$D_2 = \begin{cases} 1, & \text{男} \\ 0, & \text{女} \end{cases} \quad D_3 = \begin{cases} 1, & \text{硕士} \\ 0, & \text{本科} \end{cases} \quad D_4 = \begin{cases} 1, & \text{博士} \\ 0, & \text{本科} \end{cases}$$

1. 基准类是什么？

2. 解释各系数所代表的含义，并预期各系数的符号。

3. 若  $B_4 > B_3$ ，你得出什么结论？

答案： 1. 基准类是本科学历的女教师。

2.  $B_0$  表示刚参加工作的本科学历女教师的收入，所以  $B_0$  的符号为正。

$B_1$  表示在其他条件不变时，工龄变化一个单位所引起的收入的变化，所以  $B_1$  的符号为正。

$B_2$  表示男教师与女教师的工资差异，所以  $B_2$  的符号为正。

$B_3$  表示硕士学历与本科学历对工资收入的影响，所以  $B_3$  的符号为正。

$B_4$  表示博士学历与本科学历对工资收入的影响，所以  $B_4$  的符号为正。

3. 若  $B_4 > B_3$ ，说明博士学历的大学教师比硕士学历的大学教师收入要高。

五、若在模型： $Y_t = B_1 + B_2 X_t + u_t$  中存在下列形式的异方差： $Var(u_t) = \sigma^2 X_t^3$ ，你如何估计参数  $B_1, B_2$ （10 分）

答案：使用加权最小二乘法估计模型中的参数  $B_1$ ,  $B_2$ 。

在模型  $Y_t = B_1 + B_2 X_t + u_t$  的两边同时除以  $\sqrt{X_t^3}$ ，我们有：

$$\frac{Y_t}{\sqrt{X_t^3}} = B_1 \frac{1}{\sqrt{X_t^3}} + B_2 \frac{1}{\sqrt{X_t}} + \frac{u_t}{\sqrt{X_t^3}}$$

令  $Y_t^* = \frac{Y_t}{\sqrt{X_t^3}}$ ,  $v_t = \frac{u_t}{\sqrt{X_t^3}}$ ，则上面的模型可以表示为：

$$Y_t^* = B_1 \frac{1}{\sqrt{X_t^3}} + B_2 \frac{1}{\sqrt{X_t}} + v_t \quad (1)$$

$$Var(v_t) = Var\left(\frac{u_t}{\sqrt{X_t^3}}\right) = \frac{Var(u_t)}{X_t^3} = \frac{\sigma^2 X_t^3}{X_t^3} = \sigma^2$$

，即变换后的模型（1）的随机误差项满足同方差假定，可以使用 OLS 估计出  $B_1$ ,  $B_2$ 。上述方法称为加权最小二乘法。

六、简述自相关后果。对于线性回归模型  $Y_t = B_1 + B_2 X_{1t} + B_3 X_{2t} + u_t$ ，如果存在  $u_t = \rho u_{t-1} + v_t$  形式的自相关，应该采取哪些补救措施？（15 分）

答案：自相关就是指回归模型中随机误差项之间存在相关。用符号表示：

$$Cov(u_i, u_j) = Eu_i u_j = 0 \quad i \neq j$$

对于线性回归模型  $Y_t = B_1 + B_2 X_{1t} + B_3 X_{2t} + u_t$ ，若在模型中存在  $u_t = \rho u_{t-1} + v_t$  形式的自相关问题，我们使用广义差分变换，使得变换后的模型不存在自相关问题。

对于模型：  $Y_t = B_1 + B_2 X_{1t} + B_3 X_{2t} + u_t$

(1)

取模型的一阶滞后：  $Y_{t-1} = B_1 + B_2 X_{1t-1} + B_3 X_{2t-1} + u_{t-1}$

(2)

在（2）式的两边同时乘以相关系数  $\rho$ ，则有：

$$\rho Y_{t-1} = \rho B_1 + \rho B_2 X_{1t-1} + \rho B_3 X_{2t-1} + \rho u_{t-1}$$

(3)

用（1）式减（3）式并整理得：

$$Y_t - \rho Y_{t-1} = B_1(1 - \rho) + B_2(X_{1t} - \rho X_{1t-1}) + B_3(X_{2t} - \rho X_{2t-1}) + u_t - \rho u_{t-1}$$

$$\text{令 } Y_t^* = Y_t - \rho Y_{t-1}, \quad B_1^* = B_1(1 - \rho), \quad X_{1t}^* = X_{1t} - \rho X_{1t-1}, \quad X_{2t}^* = X_{2t} - \rho X_{2t-1}$$

则有：

$$Y_t^* = B_1^* + B_2 X_{1t}^* + B_3 X_{2t}^* + v_t$$

(4)

在 (4) 中  $v_t$  满足古典假定, 我们可以使用普通最小二乘法估计 (4) 式, 得到  $B_1^*$ ,  $B_2$ ,  $B_3$  的估计量, 再利用  $B_1$  和  $B_1^*$  的对应关系得到  $B_1$  的估计值。

# 练习题答案四

## 一、判断正误（10 分）

- 1、随机变量的条件均值与非条件均值是一回事。(F)
- 2、线性回归模型意味着变量是线性的。(F)
- 3、 $ESS = TSS + RSS$ 。(F)
- 4、对于多元回归模型，如果联合检验结果是统计显著的则意味着模型中任何一个单独的变量均是统计显著的。(F)
- 5、双对数模型中的斜率表示因变量对自变量的弹性。(T)
- 6、为了避免陷入虚拟变量陷阱，如果一个定性变量有  $m$  类，则要引入  $m$  个虚拟变量。(F)
- 7、如果回归模型违背了同方差假定，最小二乘估计量是有偏无效的。(F)
- 8、在存在接近多重共线性的情况下，回归系数的标准差会趋于变小，相应的  $t$  值会趋于变大。(F)
- 9、在任何情况下 OLS 估计量都是待估参数的最优线性无偏估计。(F)
- 10、一个联立方程模型中的外生变量在另一个联立方程模型中可能是内生变量。(T)

## 二、用经济计量方法研究经济问题时有哪些主要步骤？（10 分）

答：书中第二页，经济计量学方法论中的八个步骤。

## 三、回归模型中的随机误差项主要包括哪些因素的影响？（10 分）

答：见 PPT

## 四、古典线性回归模型具有哪些基本假定。（10 分）

答：1 解释变量与随机误差项不相关。

2 随机误差项的期望或均值为零。

3 随机误差项具有同方差，即每个随机误差项的方差为一个相等的常数。

4 两个随机误差项之间不相关，即随机误差项无自相关。

## 五、以二元回归为例简述普通最小二乘法的原理？（10 分）

答：书中第 88 页的最小二乘原理。

六、若在模型： $Y_t = B_1 + B_2 X_t + u_t$  中存在下列形式的异方差： $Var(u_t) = \sigma^2 X_t^2$ ，你如何估计参数  $B_1, B_2$ （10 分）

答：将原模型左右两边同时除以  $X_t$ ，原模型变形为：

$$\frac{Y_t}{X_t} = \frac{B_1}{X_t} + B_2 + \frac{u_t}{X_t}$$

(1)

令  $Y_t^* = \frac{Y_t}{X_t}$ ,  $X_t^* = \frac{1}{X_t}$ ,  $v_t = \frac{u_t}{X_t}$ ，则式 (1) 可以写为：

$$Y_t^* = B_2 + B_1 X_t^* + v_t$$

(2)

由于  $Var(v_t) = Var(\frac{u_t}{X_t}) = \frac{1}{X_t^2} Var(u_t) = \sigma^2$ ，所以式(2)所表示的模型不再存在异方差问题，故可利用普通最小二乘法对其进行估计，求得参数  $B_1, B_2$  的估计值。

七、应用题（共 30 分）

利用美国 1980-1995 年间人均消费支出（PCE）和人均可支配收入（PDPI）的数据，得到了如下回归分析结果：

Dependent Variable: LOG(PCE)

Method: Least Squares

Date: 06/09/05 Time: 23:43

Sample: 1980 1995

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(PDPI)	1.205281	0.028891	41.71870	0.0000
C	-2.092664	0.281286	-7.439640	0.0000
R-squared	0.992020	Mean dependent var	9.641839	
Adjusted R-squared	0.991450	S.D. dependent var	0.096436	
S.E. of regression	0.008917	Akaike info criterion	-6.485274	
Sum squared resid	0.001113	Schwarz criterion	-6.388701	
Log likelihood	53.88219	F-statistic	1740.450	
Durbin-Watson stat	2.322736	Prob(F-statistic)	0.000000	

(1) 根据以上结果，写出回归分析结果报告。（10 分）

答： $\log(\hat{PCE}) = -2.09 + 1.21\log(PDPI)$   
 $se=(0.28) \quad (0.029)$   
 $t=(-7.44) \quad (41.72)$   
 $p=(0.0000) \quad (0.0000) \quad R^2=0.992$

(2) 如何解释解释变量的系数和综合判定系数？（10 分）

答：由于该模型是双对数模型，因此，解释变量的系数为因变量对自变量的弹性，在本例中为消费收入弹性，表示收入每增加 1%，消费将平均增加 1.2%。

(3) 对模型中解释变量系数  $B_2$  进行显著性检验。（10 分）

答：1、  $H_0: B_2=0$   $H1: B_2 \neq 0$

$$T = \frac{b_2 - B_2}{se(b_2)} \sim t(14)$$

2、构造统计量

$$T = \frac{b_2 - B_2}{se(b_2)} = \frac{1.21 - 0}{0.029} = 41.72$$

3、计算相应的 T 值，

4、查显著性水平为  $\alpha$  的临界值  $t_{\alpha/2}(14) = 2.145$

$$T = \frac{b_2 - B_2}{se(b_2)} = \frac{1.21 - 0}{0.029} = 41.72 > t_{\alpha/2}(15) = 2.131$$

由于

所以拒绝原假设，认为解释变量系数  $B_2$  是统计显著的。