

## 多元线性回归模型练习

### 一、单项选择题

1. 在由  $n=30$  的一组样本估计的、包含 3 个解释变量的线性回归模型中，计算得可决系数为 0.8500，则调整后的可决系数为 ( D )

A. 0.8603      B. 0.8389      C. 0.8655      D. 0.8327

2. 用一组有 30 个观测值的样本估计模型  $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + u_t$  后，在 0.05 的显著性水平上对  $b_1$  的显著性作 t 检验，则  $b_1$  显著地不等于零的条件是其统计量 t 大于等于 ( C )

A.  $t_{0.05}(30)$     B.  $t_{0.025}(28)$     C.  $t_{0.025}(27)$     D.  $F_{0.025}(1,28)$

3. 线性回归模型  $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + u_t$  中，检验

$$t = \frac{\hat{\beta}_i}{\sqrt{\text{var}(\hat{\beta}_i)}}$$

$H_0: b_i = 0 (i = 0, 1, 2, \dots, k)$  时，所用的统计量 服从 ( C )

A.  $t(n-k+1)$       B.  $t(n-k-2)$       C.  $t(n-k-1)$

D.  $t(n-k+2)$

4. 调整的可决系数  $\bar{R}^2$  与多元样本判定系数  $R^2$  之间有如下关系 ( D )

A.  $\bar{R}^2 = \frac{n-1}{n-k-1} R^2$       B.  $\bar{R}^2 = 1 - \frac{n-1}{n-k-1} R^2$

C.  $\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 + R^2)$       D.  $\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2)$

5. 对模型  $Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \mu_i$  进行总体显著性 F 检验，检验的零假设是 ( A )

A.  $\alpha_1 = \alpha_2 = 0$     B.  $\alpha_1 = 0$     C.  $\alpha_2 = 0$     D.  $\alpha_0 = 0$  或  $\alpha_1 = 0$

6. 设 k 为回归模型中的参数个数，n 为样本容量。则对多元线性回归方程进行显著性检验时，所用的 F 统计量可表示为 ( B )

A.  $\frac{\text{ESS}/(n-k)}{\text{RSS}/(k-1)}$       B.  $\frac{R^2/k}{(1-R^2)/(n-k-1)}$

C.  $\frac{R^2/(n-k)}{(1-R^2)/(k-1)}$       D.  $\frac{\text{ESS}/(k-1)}{\text{TSS}/(n-k)}$

7. 多元线性回归分析中 ( 回归模型中的参数个数为 k ), 调整后的可决系数  $\bar{R}^2$  与可决系数  $R^2$  之间的关系 ( A )

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

A.

B.  $\bar{R}^2 = R^2$

C.  $R^2 > 0$

D.

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

8. 已知五元线性回归模型估计的残差平方和为  $\sum e_t^2 = 800$ ，样本容量为 46，则随机误差项  $u_t$  的方差估计量  $\sigma^2$  为( D )

A. 33.33      B. 40      C. 38.09      D. 20

9. 多元线性回归分析中的 ESS 反映了( C )

A. 因变量观测值总变差的大小      B. 因变量回归估计值总变差的大小

C. 因变量观测值与估计值之间的总变差      D. Y 关于 X 的边际变化

23. 在古典假设成立的条件下用 OLS 方法估计线性回归模型参数，则参数估计量具有( C )的统计性质。

A. 有偏特性      B. 非线性特性      C. 最小方差特性      D. 非一致性特性

10. 关于可决系数  $R^2$ ，以下说法中错误的是( D )

A. 可决系数  $R^2$  的定义为被回归方程已经解释的变差与总变差之比

B.  $R^2 \in [0, 1]$

C. 可决系数  $R^2$  反映了样本回归线对样本观测值拟合优劣程度的一种描述

D. 可决系数  $R^2$  的大小不受到回归模型中所包含的解释变量个数的影响

11. 下列说法中正确的是：( D )

A 如果模型的  $R^2$  很高，我们可以认为此模型的质量较好

B 如果模型的  $R^2$  较低，我们可以认为此模型的质量较差

C 如果某一参数不能通过显著性检验，我们应该剔除该解释变量

D 如果某一参数不能通过显著性检验，我们不应该随便剔除该解释变量

## 二、多项选择题

1. 调整后的判定系数  $\bar{R}^2$  与判定系数  $R^2$  之间的关系叙述正确的有( CDE )

A.  $\bar{R}^2$  与  $R^2$  均非负      B.  $\bar{R}^2$  有可能大于  $R^2$

C. 判断多元回归模型拟合优度时，使用  $\bar{R}^2$

D. 模型中包含的解释变量个数越多， $\bar{R}^2$  与  $R^2$  就相差越大

E. 只要模型中包括截距项在内的参数的个数大于 1，则  $\bar{R}^2 < R^2$

2. 对模型  $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + u_t$  进行总体显著性检验，如果检验结果总体线性关系显著，则有( BCD )

- A.  $b_1 = b_2 = 0$     B.  $b_1 \neq 0, b_2 = 0$     C.  $b_1 = 0, b_2 \neq 0$     D.  $b_1 \neq 0, b_2 \neq 0$     E.  
 $b_1 = b_2 \neq 0$

3. 回归变差（或回归平方和）是指（ BCD ）

- A. 被解释变量的实际值与平均值的离差平方和  
 B. 被解释变量的回归值与平均值的离差平方和  
 C. 被解释变量的总变差与剩余变差之差  
 D. 解释变量变动所引起的被解释变量的变差  
 E. 随机因素影响所引起的被解释变量的变差

4. 剩余变差是指（ ACDE ）

- A. 随机因素影响所引起的被解释变量的变差  
 B. 解释变量变动所引起的被解释变量的变差  
 C. 被解释变量的变差中，回归方程不能做出解释的部分  
 D. 被解释变量的总变差与回归平方和之差  
 E. 被解释变量的实际值与回归值的离差平方和

### 三、计算题

1. 根据某地 1961—1999 年共 39 年的总产出  $Y$ 、劳动投入  $L$  和资本投入  $K$  的年度数据，运用普通最小二乘法估计得出了下列回归方程：

$$\ln \hat{Y} = -3.938 + 1.451 \ln L + 0.3841 \ln K$$

(0.237) (0.083) (0.048)

$$R^2 = 0.9946, \text{DW} = 0.858$$

式下括号中的数字为相应估计量的标准误差。

(1) 解释回归系数的经济含义；

(2) 系数的符号符合你的预期吗？为什么？

解答：（1）这是一个对数化以后表现为线性关系的模型， $\ln L$  的系数为 1.451 意味着资本投入  $K$  保持不变时劳动—产出弹性为 1.451； $\ln K$  的系数为 0.384 意味着劳动投入  $L$  保持不变时资本—产出弹性为 0.384。

（2）系数符号符合预期，作为弹性，都是正值。

2. 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：

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

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

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？

解答：（1）第2个方程更合理一些，因为某天慢跑者的人数同该天日照的小时数应该是正相关的。

（2）出现不同符号的原因很可能是由于 $X_2$ 与 $X_3$ 高度相关而导致出现多重共线性的缘故。从生活经验来看也是如此，日照时间长，必然当天的最高气温也就高。而日照时间长度和第二天需交学期论文的班级数是没有相关性的。

3. 设有模型  $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + u_t$ ，试在下列条件下：

$b_1 + b_2 = 1$        $b_1 = b_2$ 。分别求出  $b_1$ ， $b_2$  的最小二乘估计量。

解答：当  $b_1 + b_2 = 1$  时，模型变为  $y_t - x_{2t} = b_0 + b_1(x_{1t} - x_{2t}) + u_t$ ，可作为一元回归模型来

$$\text{对待 } b_1 = \frac{n \sum (x_{1t} - x_{2t})(y_t - x_{2t}) - \sum (x_{1t} - x_{2t}) \sum (y_t - x_{2t})}{n \sum (x_{1t} - x_{2t})^2 - (\sum (x_{1t} - x_{2t}))^2}$$

当  $b_1 = b_2$  时，模型变为  $y_t = b_0 + b_1(x_{1t} + x_{2t}) + u_t$ ，同样可作为一元回归模型来对待

$$b_1 = \frac{n \sum (x_{1t} + x_{2t}) y_t - \sum (x_{1t} + x_{2t}) \sum y_t}{n \sum (x_{1t} + x_{2t})^2 - (\sum (x_{1t} + x_{2t}))^2}$$

4. 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：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）：

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

(2.6) (6.3) (0.61) (5.9)

$$\bar{R}^2 = 0.63 \quad n = 35$$

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

解答：（1） $X_{1i}$  是盒饭价格， $X_{2i}$  是气温， $X_{3i}$  是学校当日的学生数量， $X_{4i}$  是附近餐厅的盒饭价格。

（2）在四个解释变量中，盒饭价格同校园内食堂每天卖出的盒饭数量应该是负相关关系，其符号应该为负，应为  $X_{4i}$ ；学校当日的学生数量每变化一个单位，盒饭相应的变化数量不会是 28.4 或者 12.7，应该是小于 1 的，应为  $X_{3i}$ ；至于其余两个变量，从一般经验来看，

被解释变量对价格的反应会比对气温的反应更灵敏一些，所以  $x_{1i}$  是附近餐厅的盒饭价格，  
 $x_{2i}$  是气温。