第二章　简单线性回归模型

一、单项选择题（每题 2 分）：

1、回归分析中定义的（　　）。

A、解释变量和被解释变量都是随机变量

B、解释变量为非随机变量，被解释变量为随机变量

C、解释变量和被解释变量都为非随机变量

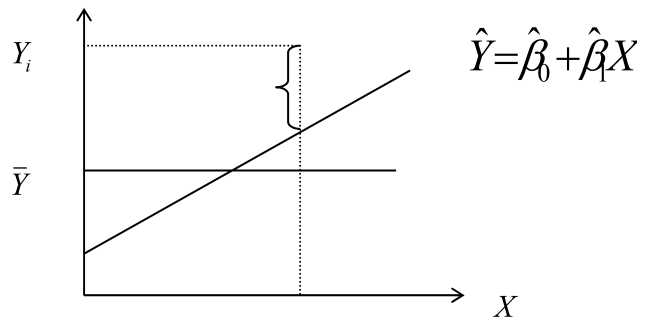
D、解释变量为随机变量，被解释变量为非随机变量

2、最小二乘准则是指使（　　　　）达到最小值的原则确定样本回归方程。

A、$\left|\sum_{t=1}^{n}(Y_t-\hat{Y}_t)\right|$　　　　B、$\sum_{t=1}^{n}\left|Y_t-\hat{Y}_t\right|$

C、$\max\left|Y_t-\hat{Y}_t\right|$　　　　D、$\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2$

3、下图中“{”所指的距离是（　　　　　　）。



A、随机误差项　　　　B、残差

C、Y_i的离差　　　　D、$\hat{Y}_i$的离差

4、参数估计量$\hat{\beta}$是Y_i的线性函数称为参数估计量具有(　　　　)的性质。

A、线性　　　　B、无偏性

C、有效性　　　　D、一致性

5、参数β的估计量$\hat{\beta}$具备最佳性是指（　　　）。

A、$Var(\hat{\beta})=0$　　　　B、$Var(\hat{\beta})$为最小

C、$\hat{\beta}-\beta=0$　　　　D、$(\hat{\beta}-\beta)$为最小

6、反映由模型中解释变量所解释的那部分离差大小的是(　　　)。

A、总体平方和　　　　B、回归平方和

C、残差平方和　　　　D、样本平方和

7、总体平方和 TSS、残差平方和 RSS 与回归平方和 ESS 三者的关系是（　　　）。

A、RSS=TSS+ESS　　　　B、TSS=RSS+ESS

C、ESS=RSS-TSS　　　　D、ESS=TSS+RSS

8、下面哪一个必定是错误的（　　　）。

A、$\hat{Y}_i = 30 + 0.2X_i$，$r_{XY} = 0.8$

B、$\hat{Y}_i = -75 + 1.5X_i$，$r_{XY} = 0.91$

C、$\hat{Y}_i = 5 - 2.1X_i$，$r_{XY} = 0.78$

D、$\hat{Y}_i = -12 - 3.5X_i$，$r_{XY} = -0.96$

9、产量(X,台)与单位产品成本(Y,元/台)之间的回归方程为$\hat{Y} = 356 - 1.5X$，这说明（　　　）。

A、产量每增加一台，单位产品成本增加 356 元

B、产量每增加一台，单位产品成本减少 1.5 元

C、产量每增加一台，单位产品成本平均增加 356 元

D、产量每增加一台，单位产品成本平均减少 1.5 元

10、回归模型$Y_i = \beta_0 + \beta_1 X_i + \mu_i$，i = 1，…，n 中，总体方差未知，检验$H_0$：$\beta_1 = 0$时，所用的检验统计量$\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}$服从（　　　）。

A、$\chi^2(n-2)$　　　　B、$t(n-1)$

C、$\chi^2(n-1)$　　　　D、$t(n-2)$

11、对下列模型进行经济意义检验，哪一个模型通常被认为没有实际价值的（　）。

A、C_i（消费）$= 500 + 0.8I_i$（收入）

B、Q_{di}（商品需求）$= 10 + 0.8I_i$（收入）$+ 0.9P_i$（价格）

C、Q_{si}（商品供给）$= 20 + 0.75P_i$（价格）

D、Y_i（产出量）$= 0.65K_i^{0.6}$（资本）$L_i^{0.4}$（劳动）

12、进行相关分析时，假定相关的两个变量(　　　)。

A、都是随机变量　　　　B、都不是随机变量

C、一个是随机变量，一个不是随机变量　　D、随机或非随机都可以

13、假设用OLS法得到的样本回归直线为$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$，以下说法不正确的是(　　　　)。

A、$\sum e_i = 0$　　　　B、$(\bar{X}, \bar{Y})$一定在回归直线上

C、$\bar{\hat{Y}} = \bar{Y}$　　　　D、$COV(X_i, e_i) \neq 0$

14、对样本的相关系数γ，以下结论错误的是（　　　　）。

A、$|\gamma|$越接近 0，X 和 Y 之间的线性相关程度越高

B、$|\gamma|$越接近 1，X 和 Y 之间的线性相关程度越高

C、$-1 \le \gamma \le 1$

D、$\gamma = 0$，则在一定条件下 X 与 Y 相互独立。

二、多项选择题（每题 3 分）：

1、利用普通最小二乘法求得的样本回归直线 $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ 的特点（　　　　　）。

A、必然过点$(\bar{X}, \bar{Y})$　　　　B、可能通过点$(\bar{X}, \bar{Y})$

C、残差e_i的均值为常数　　　　D、$\hat{Y}_i$的均值与Y_i的均值相等

E、残差e_i与解释变量之间有一定的相关性

2、古典线性回归模型的普通最小二乘估计量的特性有（　　　　　）。

A、无偏性　　B、线性　　C、最小方差　　D、不一致性　　E、有偏性

3、指出下列哪些现象是相关关系（　　　　）。

A、家庭消费支出与收入　　　　B、商品销售额和销售量、销售价格

C、物价水平与商品需求量　　　　D、小麦亩产量与施肥量

E、学习成绩总分与各门课程成绩分数

4、一元线性回归模型 $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$ 的经典假设包括（　　　　）。

A、$E(e_i) = 0$　　　　B、$Var(e_i) = \sigma^2$（常数）

C、$\mathrm{cov}(e_i, e_j) = 0$　　　　D、e_i～N(0，1)

E、X 为非随机变量，且$\mathrm{cov}(X_i, e_i) = 0$

5、以 Y 表示实际观测值，$\hat{Y}$表示回归估计值，e 表示残差，则回归直线满足（　　　　）。

A、通过样本均值点$(\bar{X}, \bar{Y})$　　　　B、$\sum Y_i = \sum \hat{Y}_i$

C、$\mathrm{cov}(X_i, e_i) = 0$　　　　D、$\sum (Y_i - \hat{Y}_i)^2 = 0$

E、$\sum (\hat{Y}_i - \bar{Y})^2 = 0$

6、反映回归直线拟合优度的指标有（　　　　）。

A、相关系数　　　　B、回归系数

C、样本决定系数　　　　D、回归方程的标准误差

E、剩余变差（或残差平方和）

三、名词解释（每题 4 分）：

1、回归平方和

2、拟合优度检验

3、相关关系

6、高斯－马尔可夫定理

四、简答（每题 5 分）：

1、给定一元线性回归模型$Y_t = \beta_0 + \beta_1 X_t + \mu_t \qquad t = 1,2,\cdots,n$

（1）叙述模型的基本假定；

（2）写出参数β_0和β_1的最小二乘估计公式；

（3）说明满足基本假定的最小二乘估计量的统计性质；

（4）写出随机扰动项方差的无偏估计公式。

2、随机误差项包含哪些因素影响？

3、普通最小二乘法参数估计量的统计性质及其含义。

4、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$kids = \beta_0 + \beta_1 educ + u$

（1）随机扰动项u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。

5、简要回答：为什么用可决系数R^2评价拟合优度，而不用残差平方和作为评价标准？

五、辨析（每题 5 分）：

1． 即使经典线性回归模型（CLRM）中的干扰项不服从正态分布的，OLS估计量仍然是无偏的。

2、随机扰动项的方差与随机扰动项方差的无偏估计没有区别。

3、在计量经济模型中，随机扰动项与残差项无区别。

六、计算分析（每题 12 分）：

1、某农产品试验产量Y（公斤/亩）和施肥量X（公斤/亩）7 块地的数据资料汇总如下：

$\sum X_i = 255$　　$\sum Y_i = 3050$

$\sum x_i^2 = 1217.71$　　$\sum y_i^2 = 8371.429$　　$\sum x_i y_i = 3122.857$

后来发现遗漏的第八块地的数据：$X_8 = 20$，$Y_8 = 400$。

要求汇总全部 8 块地数据后进行以下各项计算，并对计算结果的经济意义和统计意义做简要的解释。

（1）该农产品试验产量对施肥量 X（公斤/亩）回归模型$Y = a + bX + u$进行估计；

（2）对回归系数（斜率）进行统计假设检验，信度为 0.05；

（3）估计可决系数并进行统计假设检验，信度为 0.05。

1、首先汇总全部 8 块地数据：

$$\sum_{i=1}^{8} X_i = \sum_{i=1}^{7} X_i + X_8 \ \ =255+20 \ \ =275$$

$$\overline{X}_{(8)} = \sum_{i=1}^{8} X_i \Big/ n = \frac{275}{8} = 34.375$$

$$\sum_{i=1}^{7} X_i^2 = \sum_{i=1}^{7} x_i^2 + 7\overline{X}_{(7)}^2 \ \ =1217.71+7\times\left(\frac{255}{7}\right)^2=10507$$

$$\sum_{i=1}^{8} X_i^2 = \sum_{i=1}^{7} X_i^2 + X_8^2 \ \ =10507+20^2 \ = \ 10907$$

$$\sum_{i=1}^{8} x_i^2 = \sum_{i=1}^{8} X_i^2 + 8\overline{X}_{(8)}^2 \ \ = \ 10907-8\times\left(\frac{275}{8}\right)^2=1453.88$$

$$\sum_{i=1}^{8} Y_i = \sum_{i=1}^{7} Y_i + Y_8=3050+400=3450$$

$$\overline{Y}_{(8)} = \sum_{i=1}^{8} Y_i \Big/ n = \frac{3450}{8} = 431.25$$

$$\sum_{i=1}^{7} Y_i^2 = \sum_{i=1}^{7} y_i^2 + 7\overline{Y}_{(7)}^2 \ \ =8371.429+7\times\left(\frac{3050}{7}\right)^2=1337300$$

$$\sum_{i=1}^{8} Y_i^2 = \sum_{i=1}^{7} Y_i^2 + Y_8^2 \ \ =1337300+400^2 \ = \ 1497300$$

$$\sum_{i=1}^{8} y_i^2 = \sum_{i=1}^{8} Y_i^2 + 8\overline{Y}_{(8)}^2 \ \ =1497300 \ -8\times(\frac{3450}{8})^2== \ 9487.5$$

$$\sum_{i=1}^{7} X_iY_i = \sum_{i=1}^{7} x_iy_i + 7\overline{X}_{(7)}\overline{Y}_{(7)} \ \ ==3122.857+7\left(\frac{255}{7}\right)\times\left(\frac{3050}{7}\right)=114230$$

$$\sum_{i=1}^{8} X_iY_i = \sum_{i=1}^{7} X_iY_i + X_8Y_8 \ \ =114230+20\times400 \ \ =122230$$

$$\sum_{i=1}^{8} x_iy_i = \sum_{i=1}^{8} X_iY_i - 8\overline{X}_{(8)}\overline{Y}_{(8)} \ \ =122230-8\times34.375\times431.25 \ \ =3636.25$$

（1）该农产品试验产量对施肥量 X（公斤/亩）回归模型$Y = a + bX + u$进行估计

$$\hat{b}=\frac{\sum x_i y_i}{\sum x_i^2}=\frac{3636.25}{1453.88}=2.5011$$

$$\hat{a}=\overline{Y}-\hat{b}\overline{X}\quad =431.25-34.375*2.5011=345.28$$

$$\hat{Y}=\hat{a}+\hat{b}X\quad =345.28+2.5011X$$

统计意义：当 X 增加 1 个单位，Y 平均增加 2.5011 个单位。

经济意义：当施肥量增加 1 公斤，亩产量平均增加 2.5011 公斤。

（2）对回归系数（斜率）进行统计假设检验，置信度为 0.05。

$$\hat{\sigma}^2=\frac{\sum y_i^2-\hat{b}^2\sum x_i^2}{n-k-1}=\frac{9487.5-2.5011^2\times 1453.88}{8-(1+1)}=65.495$$

$$S_{\hat{b}}=\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}\quad =\sqrt{\frac{65.495}{1453.88}}\quad = 0.2122$$

H_0: b = 0　　H_1: b≠0

$$t=\frac{\hat{b}-b}{S_{\hat{b}}}\ =\ \frac{2.5011-0}{0.2122}\ =\ 11.7839$$

∵ $|t|$> 2.447（=$t_{0.025,6}$）

∴拒绝假设 H_0: b = 0,　接受对立假设 H_1: b≠0

统计意义：在 95%置信概率下，$\hat{b}$＝2.5011 与 b=0 之间的差异不是偶然的，$\hat{b}$＝2.5011 不是由 b=0 这样的总体所产生的。

经济意义：在 95%置信概率下，施肥量对亩产量的影响是显著的。

（3）估计可决系数并进行统计假设检验，信度为 0.05。

$$R^2=\frac{\hat{b}^2\sum x_i{}^2}{\sum y_i{}^2}=\frac{2.5011^2\times 1453.88}{9487.5}=0.9586$$

统计意义：在 Y 的总变差中，有 95.86%可以由 X 做出解释。回归方程对于样本观测点拟合良好。

经济意义：在亩产量的总变差中，有 95.86%是可以由施肥量做出解释的。

$H_0:\rho^2=0\quad H_1:\rho^2\neq 0$

$$F=\frac{R^2/k}{(1-R^2)/[n-(k+1)]}=\frac{0.9586/1}{(1-0.9586)/[8-(1+1)]}=138.859>(5.99=F_{0.05,1,6})$$

∴拒绝假设$\mathrm{H}_0:\rho^2=0$　接受对立假设$\mathrm{H}_1:\rho^2\neq 0$

统计意义：在 95%的置信概率下，回归方程可以解释的方差与未被解释的方差之间的差异不是偶然的，$R^2=0.9586$不是由$\rho^2=0$这样的总体产生的。

经济意义：在 95%的置信概率下，施肥量对亩产量的影响显著。

所需临界值在以下简表中选取：

$t_{0.025,6}=2.447$	$t_{0.025,7}=2.365$	$t_{0.025,8}=2.306$
$t_{0.005,6}=3.707$	$t_{0.005,7}=3.499$	$t_{0.005,8}=3.355$
$F_{0.05,1,7}=5.59$	$F_{0.05,2,7}=4.74$	$F_{0.05,3,7}=4.35$
$F_{0.05,1,6}=5.99$	$F_{0.05,2,6}=5.14$	$F_{0.05,3,6}=4.76$

2、试将下列非线性函数模型的线性化：

（1）$y=1/(\beta_0+\beta_1 e^{-x}+u)$；

（2）$y=\beta_1\sin x+\beta_2\cos x+\beta_3\sin 2x+\beta_4\cos 2x+u$

3、利用《中国统计年鉴（2006）》中提供的有关数据，可以对 2005 年国内各地区居民消费进行分析。如果以各省（自治区、直辖市）居民可支配收入（X，单位：元）作为解释变量，以居民消费性支出（Y，单位：元）作为被解释变量，利用 Eviews 软件，可以得到以下估计结果：

Dependent Variable: Y

Method: Least Squares

Sample: 1 31

Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	346.0459	(a)	1.131693	
X	0.728453	0.028858	(b)	

R-squared	0.956468	Mean dependent var	7773.217
Adjusted R-squared	0.954966	S.D. dependent var	2183.308
S.E. of regression	463.3222	Akaike info criterion	15.17706
Sum squared resid	6225356.	Schwarz criterion	15.26958

Log likelihood	-233.2445	F-statistic	637.1699
Durbin-Watson stat	1.372727	Prob(F-statistic)	0.000000

要求：（1）将表中(a)和(b)两项空缺的数字填出（2 分）；

（2）已知$t_{0.025}(29)=2.045, t_{0.05}(29)=1.699, t_{0.05}(30)=1.607, t_{0.025}(30)=2.042$；$\chi^2_{0.05}(29)=42.5569, \chi^2_{0.05}(30)=43.77, \chi^2_{0.025}(29)=45.72, \chi^2_{0.025}(30)=46.98$。请对模型参数的显著性做出判断（5 分）；

（3）利用回归结果进行简要分析（5 分）。

七、填空（每题 2 分）：

1、与数学中的函数关系相比，计量经济模型的显著特点是引入随机误差项u，u包含了丰富的内容，主要包括四方面，在解释变量中被忽略掉的因素的影响、变量观测值的观测误差的影响、______________________，以及其他随机因素的影响。

2、被解释变量的观测值Y_i与其回归估计值$\hat{Y}_i$之间的偏差，称为___________。

3、对线性回归模型$Y=\beta_0+\beta_1X+\mu$进行最小二乘估计，最小二乘准则是______________________。

4、高斯—马尔可夫定理证明在总体参数的各种无偏估计中，普通最小二乘估计量具有___________的特性，并由此才使最小二乘法在数理统计学和计量经济学中获得了最广泛的应用。

5、普通最小二乘法得到的参数估计量具有线性、无偏性和___________统计性质。

6、相关系数 r 的取值范围是_____________________。

7、已知某一直线回归方程的可决系数为 0.64，则解释变量与被解释变量间的相关系数为________。

8、可决系数r^2的取值范围是_____________________。

9、用一组有 30 个观测值的样本估计模型$y_t=\beta_0+\beta_1x_t+u_t$，在 0.05 的显著性水平下对$\beta_1$的显著性作 t 检验，则$\beta_1$显著地不等于零的条件是其统计量 t 大于_____________________。

10、对回归模型$Y_i=\beta_0+\beta_1X_i+u_i$进行统计检验时，通常假定$u_i$服从的分布类型为_____________________。

第二章　简单线性回归模型

一、

1、B 2、D 3、B 4、A 5、B 6、B 7、B 8、C 9、D 10、D 11、B 12、A 13、D 14、A

二、

1、ACD 2、ABC 3、ABCD 4、ABCE 5、ABC 6、C

三、

1、回归平方和用ESS表示，是被解释变量的样本估计值与其平均值的离差平方和。

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

3、当一个或若干个变量X取一定数值时，与之相对应的另一个变量Y的值虽然不确定，但却按某种规律在一定范围内变化，变量之间的这种关系，称为不确定性的统计关系或相关关系，可表示为Y=f(X，u)，其中u为随机变量。

4、在古典假定条件下，OLS 估计式是其总体参数的最佳线性无偏估计式。

四、

1、(1）零均值假定，同方差假定，无自相关假定，随机扰动项与解释变量不相关，正态性假定。

（2）$\hat{\beta}_1 = \dfrac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$，$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

（3）无偏性，最小方差性，线性。

（4）$\hat{\sigma}^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-2}$

2、随机误差项主要包括下列因素的影响：

（1）未知因素的影响；

（2）无法取得数据的已知因素的影响；

（3）众多细小因素的综合影响；

（4）模型的设定误差的影响；

（5）变量的观测误差的影响；

（6）经济现象的内在随机性的影响。

3、普通最小二乘法参数估计量的统计性质主要有线性、无偏性和最小方差性。所谓线性是指参数估计量 $\hat{\beta}$ 是 Y_i 的线性函数；所谓无偏性是指参数估计量 $\hat{\beta}$ 的均值（期望）等于模型参数值，即 $E(\hat{\beta}_0) = \beta_0$，$E(\hat{\beta}_1) = \beta_1$；参数估计量的最

小方差性是指在所有线性、无偏估计量中，该参数估计量的方差最小。

4、(1) 收入、年龄、家庭状况、政府的相关政策等也是影响生育率的重要的因素，在上述简单回归模型中，它们被包含在了随机扰动项之中。有些因素可能与教育水平相关，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。

5、可决系数 $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$，含义为样本回归做出解释的离差平方和在总离差平方和中占的比重，如果拟合程度越好，各样本观测点与回归线靠得越近，R^2 越接近 1，拟合程度越差，R^2 越小。而残差平方和不能反映拟合程度的优劣。

五、

1. 正确。$E(\hat{\beta}_2) = E(\beta_2 + \sum K_i u_i) = \beta_2$，该表达式成立与否与正态性无关。

2、错。随机扰动项的方差反映总体的波动情况，对一个特定的总体而言，是一个确定的值。

在最小二乘估计中，由于总体方差在大多数情况下并不知道，所以用样本数据去估计 $\sigma^2 : \hat{\sigma}^2 = \sum e_i^2 \big/ (n-k)$：。其中n为样本数，k为待估参数的个数。$\hat{\sigma}^2$ 是 σ^2 线性无偏估计，为一个随机变量。

3、错。它们均为随机项，但随机误差项表示总体模型的误差，残差项表示样本模型的误差；另外，残差=随机误差项+参数估计误差。

六、

2、解：(1) 由 $y = 1/(\beta_0 + \beta_1 e^{-x} + u)$　　可得 $1/y = \beta_0 + \beta_1 e^{-x} + u$，

令 $Y = 1/y, X = e^{-x}$，则可得线性模型

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

(2) 令 $X_1 = \sin x, X_2 = \cos x, X_3 = \sin 2x, X_4 = \cos 2x,$ 则原模型可化为线性模型

$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$

3、(1) (a) 为 305.7770；(b) 为 25.24223

(2) 需要使用 t 检验。由于 $t_{0.025}(29) = 2.045, t_{0.05}(29) = 1.699$；而模型中截距项和斜率项的 t 值分别为 1.131693 和 25.24223，前者不能通过 10%水平的显著性检验，后者则可以通过 5%的显著性检验。实际上二者的 p 值分别为 0.2670 和 0.0000。当然，截距项的实际价值不大。

（3）要点如下：第一，模型总体显著，拟合优度较高；第二，边际消费倾向为 0.73 左右；第三，由于模型考虑因素较少、形式过于简单，部分检验（如 DW 检验）不太理想，需做进一步完善。

七、

1、模型关系的设定误差的影响 2、剩余项（或者残差）3、$\min\sum e_i^2 = \min\sum (Y_i - \hat{Y}_i)^2$ 4、最佳线性无偏 5、最小方差性 6、$-1\leq r\leq 1$ 7、± 0.8 8、$0\leq r^2\leq 1$ 9、$t_{0.025}$（28） 10、N（0，σ^2）