

第 11 章回归分析习题解答

一. 选择题

1. 变量之间的关系可以分为两大类, 它们是 (A).

- A. 函数关系与相关关系. B. 线性相关关系和非线性相关关系.
C. 正相关关系和负相关关系. D. 简单相关关系和复杂相关关系.

2. 进行相关分析时的两个变量 (A).

- A. 都是随机变量. B. 随机的或非随机都可以.
C. 都不是随机变量. D. 一个是随机变量, 一个不是随机变量.

3. 进行回归分析时的两个变量 (D).

- A. 都是随机变量. B. 随机的或非随机都可以.
C. 都不是随机变量. D. 一个是随机变量, 一个不是随机变量.

4. 回归分析中使用的距离是点到直线的垂直坐标距离. 最小二乘准则是指 (D).

- A. 使 $\sum_{t=1}^n (Y_t - \hat{Y}_t)$ 达到最小值. B. 使 $\sum_{t=1}^n |Y_t - \hat{Y}_t|$ 达到最小值.
C. 使 $\max |Y_t - \hat{Y}_t|$ 达到最小值. D. 使 $\sum_{t=1}^n (Y_t - \hat{Y}_t)^2$ 达到最小值.

5. 在一元线性回归模型 $\begin{cases} y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2) \end{cases}$ 中, 若记 $x = x_0$ 时相应的因变量 Y 的值为

y_0 , 则 y_0 为 (B).

- A. 是一个尚不知晓的确定的数.
B. 是随机变量, 且有 $y_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$.
C. 当 β_0, β_1 确知时等于 $\beta_0 + \beta_1 x_0$.
D. 等于 $\hat{\beta}_0 + \hat{\beta}_1 x_0$.

6. 在回归分析中, 检验线性相关显著性常用的三种检验方法, 不包含 (D).

- A. 相关系数显著性检验法. B. t 检验法.
C. F 检验法 (即方差分析法). D. χ^2 检验法.

7. 在线性模型 $Y = \beta_0 + \beta_1 x + \varepsilon$ 的相关性检验中, 如果原假设 $H_0: \beta_1 = 0$ 被否定, 则表

明两个变量之间(D)。

- A. 不存在任何相关关系。
- B. 不存在显著的线性相关关系。
- C. 不存在一条曲线 $\hat{Y} = f(x)$ 能近似描述其关系。
- D. 存在显著的线性相关关系。

8. 在线性模型 $Y = \beta_0 + \beta_1 x + \varepsilon$ 的相关性检验中, 如果原假设 $H_0: \beta_1 = 0$ 没有被否定,

则表明(C)。

- A. 两个变量之间没有任何相关关系。
- B. 两个变量之间存在显著的线性相关关系。
- C. 两个变量之间不存在显著的线性相关关系。
- D. 不存在一条曲线 $\hat{Y} = f(x)$ 能近似地描述两个变量间的关系。

9. 对一元线性回归模型 $y_i = \beta_0 + \beta_1 x + e_i, i = 1, 2, \dots, n$; 诸 e_i 相互独立, 且服从

$N(0, \sigma^2)$, 作分解 $SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SS_e + SS_R$, 对检验

假设 $H_0: \beta_1 = 0$, 取显著性水平 α , 用 F 检验的拒绝域为(A)。

- A. $\left\{ \frac{SS_R}{SS_e} > \frac{1}{n-2} F_{\alpha}(1, n-2) \right\}$ 。
- B. $\left\{ \frac{SS_R}{SS_e} > \frac{1}{n-2} F_{\alpha/2}(1, n-2) \right\}$ 或 $\left\{ \frac{SS_R}{SS_e} < \frac{1}{n-2} F_{\alpha}(1, n-2) \right\}$ 。
- C. $\left\{ \frac{SS_R}{SS_T} > \frac{1}{n-2} F_{\alpha}(1, n-2) \right\}$ 。
- D. $\left\{ \frac{SS_R}{SS_e} < \frac{1}{n-2} F_{\alpha}(1, n-2) \right\}$ 。

10. 对一元线性回归模型: $Y = \beta_0 + \beta_1 x + e, e \sim N(0, \sigma^2)$, 用 7 次观测数据算得

$SS_T = \sum_{i=1}^7 (y_i - \bar{y})^2 = 8.5, SS_R = \sum_{i=1}^7 (\hat{y}_i - \bar{y})^2 = 7.82$ 。现对检验假设 $H_0: \beta_1 = 0$ 进行 F 检

验, 下列的推测正确的是(A)。(其中 $F_{0.05}(1, 5) = 6.61, F_{0.05}(1, 7) = 5.59$)

A. 由 $F = \frac{5SS_R}{SS_e} = 57.5 > 6.61$, 可以认为 Y 与 x 有显著的线性相关关系 .

B. 由 $F = \frac{5SS_R}{SS_T} = 4.6 < 6.61$, 不可以认为 Y 与 x 有显著的线性相关关系 .

C. 由 $F = \frac{7SS_R}{SS_T} = 6.44 > 5.59$, 可以为 Y 与 x 有显著的线性相关关系 .

D. 由 $F = \frac{7SS_R}{SS_e} = 80.5 > 5.59$, 可以为 Y 与 x 有显著的线性相关关系 .

二 . 填空题

11. 设 $(x_1, y_1), \dots, (x_n, y_n)$ 是 (X, Y) 的一个样本 , 样本平均值记为 (\bar{x}, \bar{y}) , y 对 x 的回归方程为 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, 则可用样本表示出数 β_0 与 β_1 的估计为 $\hat{\beta}_0 = \underline{\bar{y} - \hat{\beta}_1 \bar{x}}$,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} .$$

12. 平方和分解公式是 $SS_T = SS_R + SS_E$ 其中 $SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$, $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$,

而 $SS_R = \underline{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}$ 被称为 回归 平方和 .

13. 设总体 X 的样本为 (x_1, x_2, \dots, x_n) 对应总体 Y 有样本 (y_1, \dots, y_n) , 则 X 和 Y 的样本相关系数为 $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$, 其中 $S_{xy} = \underline{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}$;

$S_{xx} = \underline{\sum_{i=1}^n (x_i - \bar{x})^2}$; $S_{yy} = \underline{\sum_{i=1}^n (y_i - \bar{y})^2}$; 若 $|r|$ 接近于 1 就表示 X 与 Y 之

间 线性相关关系显著 .

14. 测得 (x, Y) 的观测值为

x	-2	-1	0	1	2
y	1	0	2	3	4

则 y 对 x 的回归方程是 $\hat{y} = 2 + 0.9x$; x 和 y 的样本值相关系数

$r = 0.9$; 算得统计量 $T = \frac{(n-2)r}{\sqrt{1-r^2}}$ 的观测值 $t \approx 6.19$; 有

$t_{\alpha/2}(3) = 3.18$, 检验得 y 对 x 的线性相关关系 显著 .

15 . 一家保险公司十分关心其总公司营业部加班的程度 , 决定认真调查一下现状 . 经过 10 周时间 , 收集了每周加班工作时间的数据和签发的新保单数目 , x 为每周签发的新保单数目 , Y 为每周加班工作时间 (小时) . 利用 Excel 的数据分析功能得到统计分析如下表 .

	Coefficients	标准误差	t Stat	P-value
Intercept	0.118129	0.355148	0.33262	0.74797
X Variable 1	0.003585	0.000421	8.508575	2.79E-05

由此可见 , 回归方程为 $\hat{y} = 0.118129 + 0.003585x$; 在显著性水平 $\alpha = 0.05$ 下 ,

由于对 x 的系数的检验 P - 值 $2.79 \times 10^{-5} < 0.05$, 所以 , y 对 x 的线性相关关系

显著 ; 若新保单数 $x_0 = 1000$, 给出 Y 的估计值为

$\hat{y}_0 = 0.118129 + 0.003585 \times 1000 = 3.703129$.

16 . 下表是 16 只公益股票某年的每股帐面价值 x 和当年红利 y , 利用 Excel 的数据分析功能得到的统计分析结果如下 :

方差分析					
	df	SS	MS	F	Significance F
回归分析	1	48.54045	48.54045	144.5244	9.14E-09
残差	15	5.037949	0.335863		
总计	16	53.5784			
	Coefficients	标准误差	t Stat	P-value	
Intercept	0	#N/A	#N/A	#N/A	
X Variable 1	0.097409	0.008103	12.02183	4.22E-09	

由此可见 , 当年红利关于股票帐面价值的回归方程为 $\hat{y} = 0.097409x$; 在显著性

水平 $\alpha = 0.05$ 下, 对方程的显著性的 F 检验的 P -值 $9.14 \times 10^{-9} < 0.05$, 所以, 可以认为公益股票某年的每股帐面价值和当年红利的线性相关关系 显著; 回归系数的经济意义为 每股帐面价值每增加 1 个单位, 当年每股红利增加 0.097409 个单位; 若公司序号为 6 的股票每股帐面价值 20.25 元, 估计当年红利可能为 $\hat{y} = 0.097409 \times 20.25 = 1.97253225$.

三. 应用计算题

17. 在动物学研究中, 有时需要找出某种动物的体积与重量的关系, 因为重量相对容易测量, 而测量体积比较困难, 可以利用重量预测体积的值. 下面是某种动物的 18 个随机样本的体重 x (kg) 与体积 Y ($10^{-3} m^3$) 的数据如表:

x	17.1	10.5	13.8	15.7	11.9	10.4	15.0	16.0	17.8	15.8
Y	16.7	10.4	13.5	15.7	11.6	10.2	14.5	15.8	17.6	15.2
x	15.1	12.1	18.4	17.1	16.7	16.5	15.1	15.1		
Y	14.8	11.9	18.3	16.7	16.6	15.9	15.1	14.5		

- (1) 拟合回归直线方程 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$;
- (2) 对体重 x 与体积 Y 之间的线性相关关系的显著性检验;
- (3) 求相关系数;
- (4) 对体重 $x = 15.3$ 的这种动物, 试估计它的体积 y_0 .

解: (1) 参数估计的计算表如下表.

参数估计的计算表

序号	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	17.1	16.7	292.41	278.89	285.57
2	10.5	10.4	110.25	108.16	109.2
3	13.8	13.5	190.44	182.25	186.3
4	15.7	15.7	246.49	246.49	246.49
5	11.9	11.6	141.61	134.56	138.04
6	10.4	10.2	108.16	104.04	106.08
7	15.0	14.5	225	210.25	217.5

8	16.0	15.8	256	249.64	252.8
9	17.8	17.6	316.84	309.76	313.28
10	15.8	15.2	249.64	231.04	240.16
11	15.1	14.8	228.01	219.04	223.48
12	12.1	11.9	146.41	141.61	143.99
13	18.4	18.3	338.56	334.89	336.72
14	17.1	16.7	292.41	278.89	285.57
15	16.7	16.6	278.89	275.56	277.22
16	16.5	15.9	272.25	252.81	262.35
17	15.1	15.1	228.01	228.01	228.01
18	15.1	14.5	228.01	210.25	218.95
列和	270.1	265	4149.39	3996.14	4071.71

计算可得：

$$S_{yy} = \sum y_i^2 - n\bar{y}^2 = 94.75$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 96.39$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 95.24$$

由此计算得

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = 0.9881$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -0.1048$$

因此，由该样本估计的回归方程为 $\hat{y} = -0.1048 + 0.9881x$ 。

$$(2) H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$$

$$\text{计算可得 } SS_T = S_{yy} = 94.75$$

$$SS_R = \frac{S_{xy}^2}{S_{xx}} = 94.1037$$

$$SS_E = SS_T - SS_R = 0.6463$$

其中 $n=18$, 查表可知临界值 $F_{0.05}(1,16)=4.49$ 和 $F_{0.01}(1,16)=8.53$. 因此得方差分析表

方差分析表

方差来源	平方和	自由度	F 值	临界值
回归	94.1037	1	2329.66	$F_{0.05}(1,16)=4.49$ $F_{0.01}(1,16)=8.53$
残差	0.6463	16		
总计	94.75	17		

由表可知 $F=2329.66 > F_{0.01}(1,16)=8.53$, 拒绝 H_0 . 可认为体重 X 与体积 Y 之间的线性相关关系非常显著 .

$$(3) r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0.9966$$

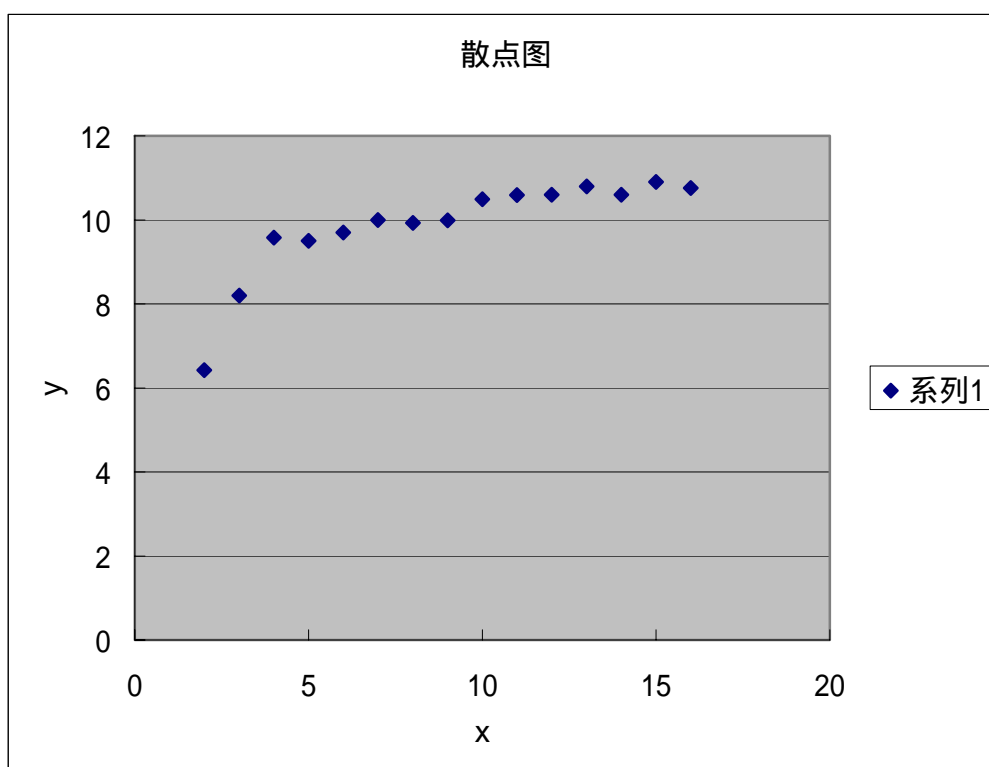
$$(4) \hat{y}_0 = -0.1048 + 0.9881 \times 15.3 = 15.01313$$

18 . 已知变量 x 与 Y 的样本数据如表:

x	Y	x	Y	x	Y
2	6.42	7	10.00	12	10.60
3	8.20	8	9.93	13	10.80
4	9.58	9	9.99	14	10.60
5	9.50	10	10.49	15	10.90
6	9.70	11	10.59	16	10.76

(1)画出散点图 ; (2)拟合回归模型 $y = \frac{x}{ax+b}$.

解 : (1) 散点图如下图



(2) $Y = \frac{x}{ax+b}$ 可以化为 $\frac{1}{Y} = a + b \cdot \frac{1}{x}$.

置换变量，设 $Y^* = \frac{1}{Y}$, $x^* = \frac{1}{x}$, 得

$$\hat{Y}^* = \hat{a} + \hat{b}x^*$$

为了检验 Y^* 与 x 的线性相关关系的显著性，并确定系数 a 及 b ，利用已给的数据 (x, Y)

写出对应的数据 (x^*, Y^*) 如下表：

变换后数据

x^*	Y^*	x^*	Y^*	x^*	Y^*
0.5	0.155763	0.142857	0.1	0.083333	0.09434
0.333333	0.121951	0.125	0.100705	0.076923	0.092593
0.25	0.104384	0.111111	0.1001	0.071429	0.09434
0.2	0.105263	0.1	0.095329	0.066667	0.091743
0.166667	0.103093	0.090909	0.094429	0.0625	0.092937

于是计算可得 $\sum x_i^* = 2.380729$, $\sum Y_i^* = 1.546969$, $\sum x_i^{*2} = 0.584347$, $\sum Y_i^{*2} = 0.163332$, $\sum x_i^* Y_i^* = 0.272624$, 且 $S_{y^* y^*} = 0.003791$, $S_{x^* x^*} = 0.206488$,

$$S_{x^*y^*} = 0.027096 .$$

$$\text{又可得 } SS_T = 0.003791 , SS_R = 0.0035556 , SS_E = SS_T - SS_R = 0.0002354 .$$

因此可得方差分析表为

方差分析表

方差来源	平方和	自由度	F 值	临界值
回归	0.0035556	1	196.359	$F_{0.05}(1,13) = 4.67$
残差	0.0002354	13		$F_{0.01}(1,13) = 9.07$
总计	0.003791	14		

因为 $F > 9.07$, 所以 Y^* 与 x^* 之间的线性相关关系特别显著 . 计算可得 :

$$\hat{b} = 0.131223 , \hat{a} = 0.082304$$

Y^* 关于 x^* 的线性回归方程为

$$Y^* = 0.082304 + 0.131223x^*$$

再换回原变量 , 得

$$\frac{1}{Y} = 4.575743 + 0.320525 \frac{1}{x}$$

$$\text{即 } Y = \frac{x}{4.575743x + 0.320525}$$

这就是所求的曲线回归方程 .