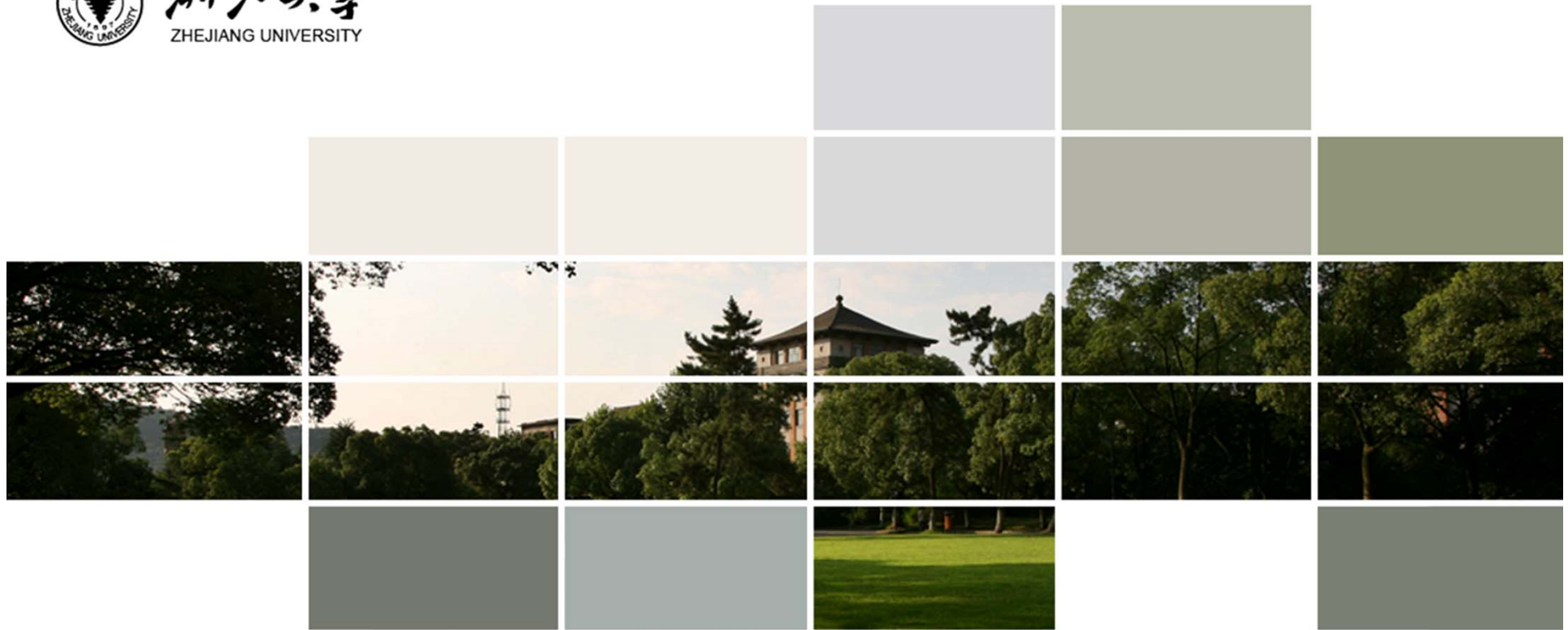




浙江大学  
ZHEJIANG UNIVERSITY



# 第64讲 一元线性回归分析 (模型检验及应用)



在第63讲已经建立了一元线性回归模型

$$\left\{ \begin{array}{l} Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n, \\ \varepsilon_i \sim N(0, \sigma^2), \text{且相互独立,} \\ \beta_0, \beta_1 (\text{回归系数}), \sigma^2 \text{未知.} \end{array} \right.$$

并根据样本  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

估计未知参数  $\beta_0, \beta_1, \sigma^2$ .

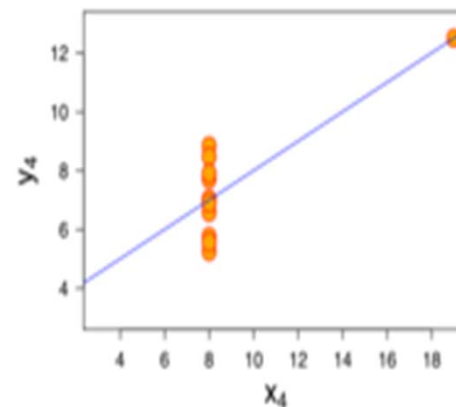
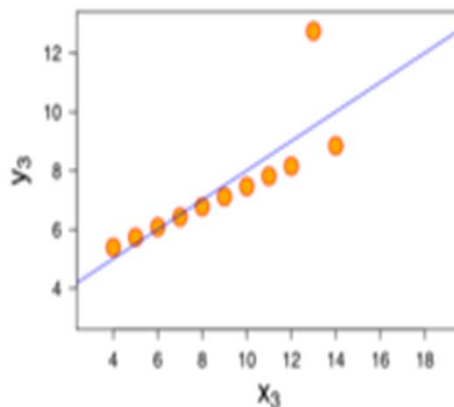
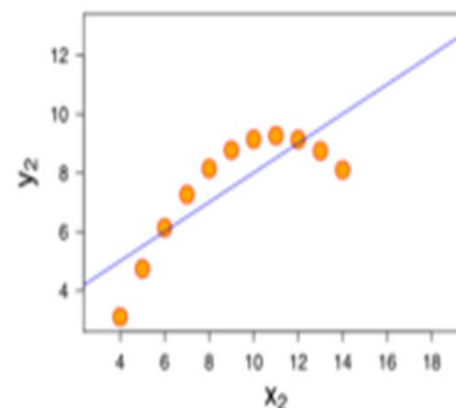
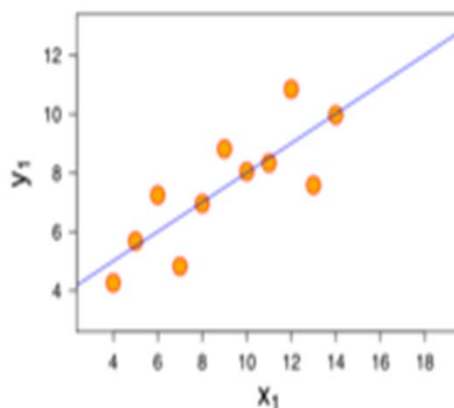


采用最小二乘法估计参数 $\beta_0, \beta_1$ ，并不需要事先知道 $Y$ 与 $x$ 之间一定具有相关关系。

因此 $\mu(x)$ 是否为 $x$ 的线性函数：

一要根据专业知识和实践来判断，

二要根据实际观察得到的数据用假设检验方法来判断。





### (3) 线性假设的显著性检验

检验假设  $H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$ .

若原假设被拒绝，说明回归效果是显著的。否则，若接受原假设，说明Y与x不是线性关系，回归方程无意义。回归效果不显著的原因可能有以下几种：

- (1) 影响Y取值的，除了x，还有其他不可忽略的因素；
- (2)  $E(Y)$  与x的关系不是线性关系，而是其他关系；
- (3) Y与x不存在关系。



检验假设  $H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$  有两种方法:

方法一: 回归方程的检验——F检验

采用方差分析法:

令  $S_T = \sum_{i=1}^n (y_i - \bar{y})^2$  —— 总平方和.

$S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  —— 回归平方和,

$S_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  —— 残差平方和.

则  $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$



定理:(1)  $\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / s_{xx})$ ,

$$\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right) \sigma^2\right).$$

$$\hat{\beta}_1 = s_{xy} / s_{xx},$$

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

一般地有

$$(2) \frac{S_E}{\sigma^2} = \frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2),$$

$S_R = \hat{\beta}_1^2 s_{xx}$  与  $S_E$  ( $\hat{\beta}_1$  与  $s^2$ ) 相互独立.

$$(3) \text{当 } \beta_1 = 0 \text{ 时, } \frac{S_R}{\sigma^2} \sim \chi^2(1).$$

$$\frac{\hat{\beta}_1 - \beta_1}{s / \sqrt{s_{xx}}} \sim t(n-2)$$

当  $\beta_1 = 0$  时, 
$$F = \frac{S_R / 1}{S_E / (n-2)} \sim F(1, n-2).$$



## 方差分析表

来源	平方和	自由度	均方	F
回归	$S_R$	1	$MS_R = S_R / 1$	$MS_R / MS_E$
残差	$S_E$	$n-2$	$MS_E = S_E / (n-2)$	
总的	$S_T$	$n-1$		

对于显著水平  $\alpha$ , 拒绝域  $W = \{F > F_\alpha(1, n-2)\}$



## 方法二：回归系数检验——t检验

当  $H_0: \beta_1 = 0$  成立时,  $t = \frac{\hat{\beta}_1 \sqrt{s_{xx}}}{s} \sim t(n-2)$

$$\frac{\hat{\beta}_1 - \beta_1}{s / \sqrt{s_{xx}}} \sim t(n-2)$$

对于显著水平  $\alpha$ , 拒绝域  $W = \{|t| > t_{\alpha/2}(n-2)\}$ .

### (4) 回归系数的置信区间

由此,  $\beta_1$  的置信水平为  $1-\alpha$  的置信区间为:

$$\left( \hat{\beta}_1 \pm t_{\alpha/2}(n-2) \times \frac{s}{\sqrt{s_{xx}}} \right)$$





## (5) 应用

对于  $y = \beta_0 + \beta_1 x + \varepsilon$ , 如果回归显著  
可以用在  $x = x_0$ , 求  $y_0$  的点预测.

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0.$$



例 1 (续): 分析可支配收入与消费支出  
之间关系. 采用Excel中的“数据分析”  
模块. 得到数据分析表.

地区	可支配收入 x(万元)	消费支出 y(万元)	地区	可支配收入 x(万元)	消费支出 y(万元)
北京	3.647	2.405	上海	4.019	2.625
天津	2.963	2.002	江苏	2.968	1.883
河北	2.054	1.253	浙江	3.455	2.155
山西	2.041	1.221	安徽	2.102	1.501
内蒙古	2.315	1.772	福建	2.806	1.859



地区	可支配收入 x(万元)	消费支出 y(万元)	地区	可支配收入 x(万元)	消费支出 y(万元)
辽宁	2.322	1.659	江西	1.986	1.278
吉林	2.021	1.461	山东	2.576	1.578
黑龙江	1.776	1.298	河南	2.044	1.373
湖北	2.084	1.450	云南	2.107	1.388
湖南	2.132	1.461	西藏	1.803	1.118
广东	3.023	2.240	陕西	2.073	1.533
广西	2.124	1.424	甘肃	1.716	1.285
海南	2.092	1.446	青海	1.757	1.235
重庆	2.297	1.657	宁夏	1.983	1.407
四川	2.031	1.505	新疆	1.792	1.389
贵州	1.870	1.259			



## 方差分析表

	自由度	平方和	均方	F值	P_值
回归	1	3.800452	3.800452	353.987	8.54E-18 显著!
误差	29	0.311348	0.010736		
总的	30	4.1118			

$S^2$

	Coef.	标准误差	t Stat	P value	Lower 95%	Upper 95%
Intercept	0.1707	0.0774	2.2046	0.0356	0.012	0.329
X	0.6089	0.0324	18.815	8.54E-18	0.543	0.675

$\hat{\beta}_0$

$\hat{\beta}_1$

$$\frac{\hat{\beta}_1 | \sqrt{s_{xx}}}{s}$$

$$\hat{\beta}_1 \pm t_{\alpha/2}(n-2) \times \frac{s}{\sqrt{s_{xx}^{12}}}$$



回归方程:  $\hat{y} = 0.1707 + 0.6089x$ .

$$x_0 = 2.5, \hat{y}_0 = 1.693.$$

关于回归方程的更多内容可以查阅  
有关回归分析教材。



《概率论与数理统计》全部结束，  
衷心感谢大家给予的支持与帮助。  
希望本次学习之旅你能有所收获，  
并能为你开启一条新的学习之路！

谢谢！

THE  
END