

Mathematical Laboratory

# 回归分析

— 多元线性回归和matlab实现



重庆大学数学与统计学院



多个变量共同影响输出变量时，采用模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{pmatrix} \quad \mathcal{E} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\beta = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_k)^T$$

$$Y = X\beta + \varepsilon$$

$$\hat{\beta} = \left[ (X^T X)^{-1} X^T \right] Y$$





$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

**线性关系成立时，具有性质：**

(1)  $\beta_i \neq 0$ ;

$$H_0: \beta_i = 0; \quad H_1: \beta_i \neq 0;$$

$$K = \{|\beta_i| > C\};$$

**单个参数的t检验。**

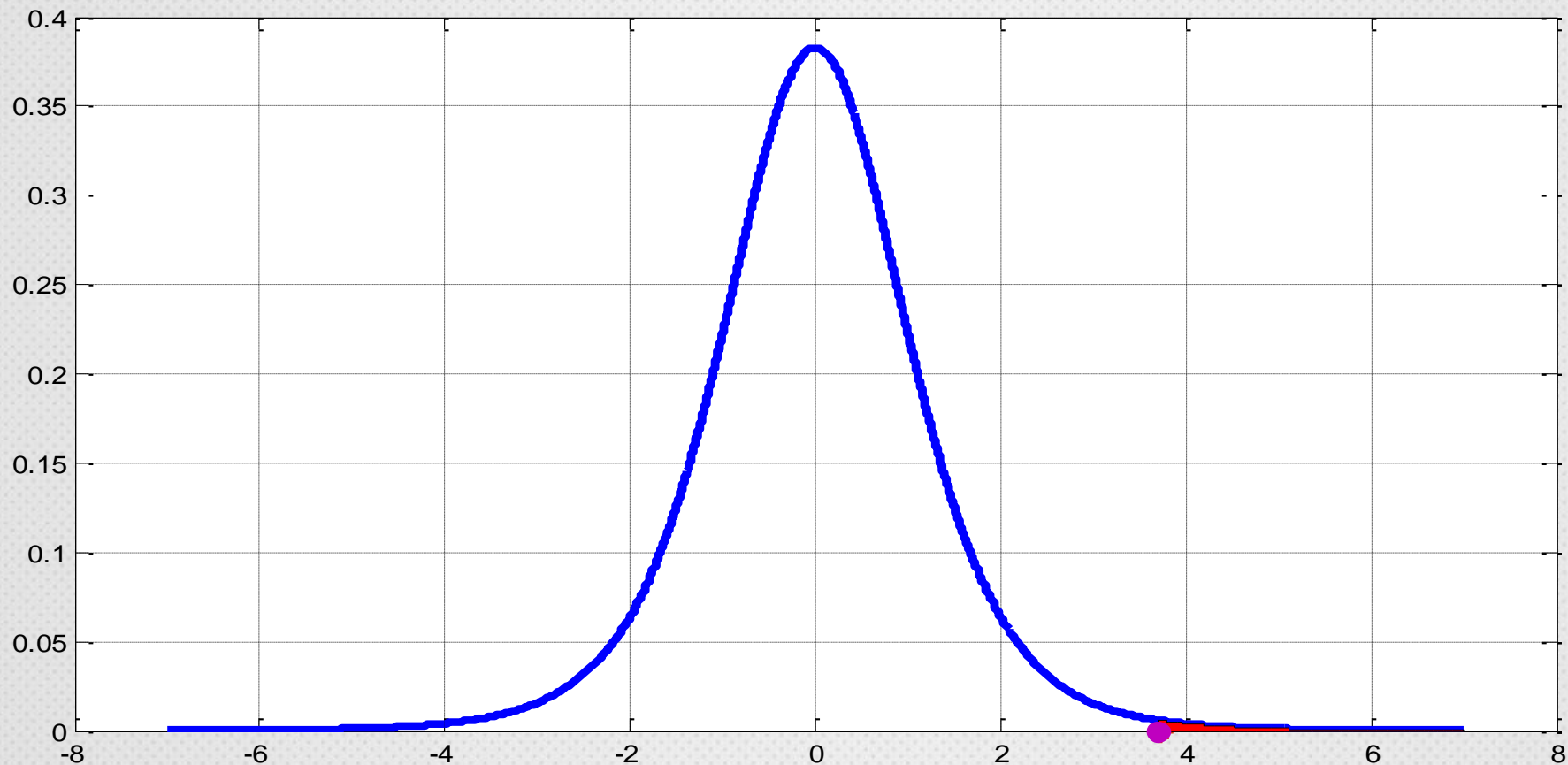
(2) **残差服从正态分布。**

**模型残差的F检验。**



**线性关系成立时，统计量服从特定的分布。**

**比统计量观测值更加异常的概率，称为p-value。**



**$n=6, \quad p=1-0.995=0.005$**



```
[b, bint, r, rint, stats] = regress(y, X)
```

**b**: 回归系数估计值。

**bint**: 回归系数95%区间估计，若区间中含有0，则表示该系数和零没有显著的差异。

**r**: 残差估计。

**rint**: 残差区间估计，若区间中不含有0，则表示该次观测值为“异常点”。

**stats**: 依次为 $R^2$ ，F-观测值，F- p-value，和扰动项方差估计值。





## 鸡肉的需求问题。

影响鸡肉需求量的潜在因素有：

可支配收入、鸡肉价格、猪肉价格、牛肉价格等。

$$\begin{aligned}\log(y) = & \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) \\ & + \beta_3 \log(x_3) + \beta_4 \log(x_4) + \varepsilon\end{aligned}$$

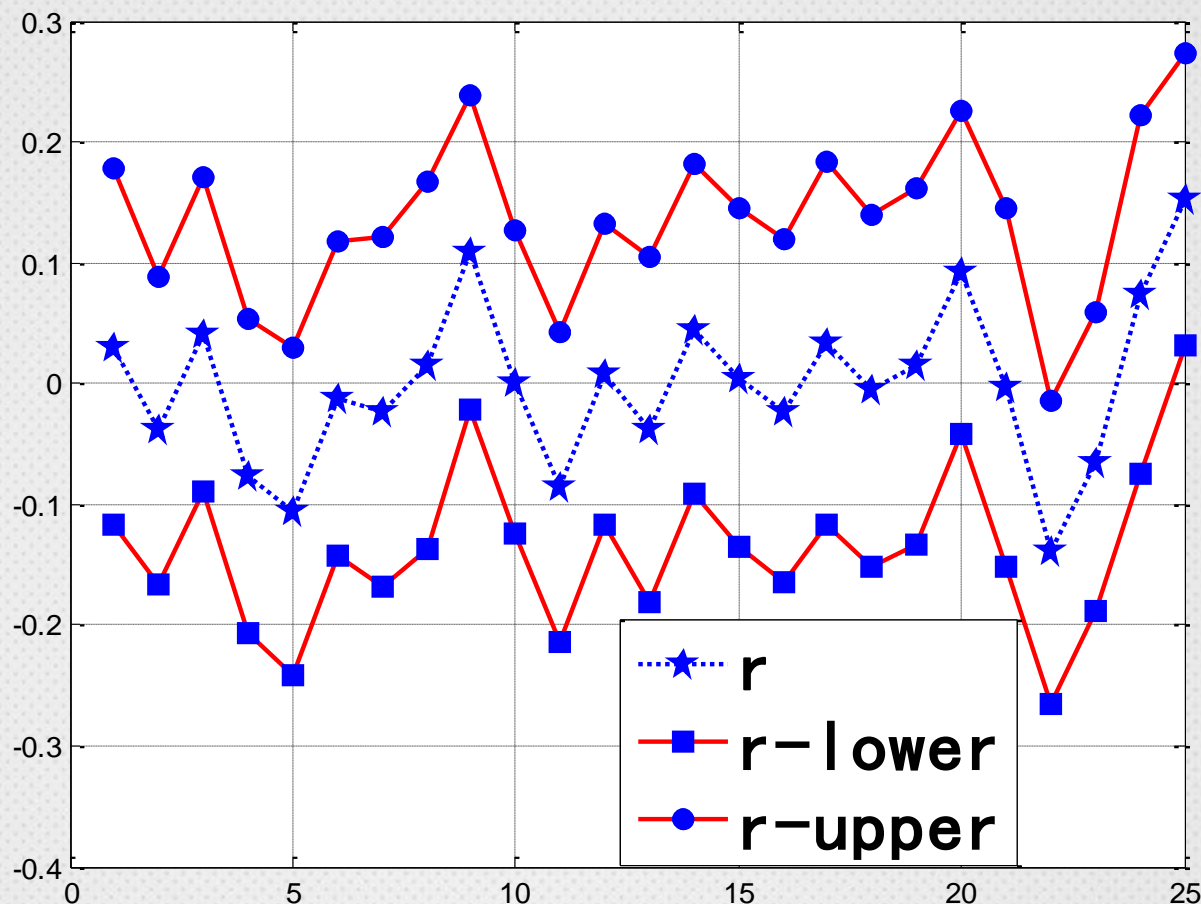
$y$  鸡肉需求量；  $x_1$  可支配收入；  $x_2$  鸡肉价格；  
 $x_3$  猪肉价格；  $x_4$  牛肉价格。

对数线性模型，其中  $\beta_i$  称为弹性系数。



MatLab输出：

B	B-int	
5.47	1.38	9.57
0.15	0.11	0.18
-0.51	-0.96	-0.05
-0.67	-1.01	-0.34
-0.22	-0.63	0.18



Stat=[0.8677, 32.8053, 0.0000, 0.0054]  
=[ $R^2$ , F, p,  $\sigma^2$ ]

# Thanks



重庆大学数学与统计学院