

# 一元线性回归模型的评价与检验

---



陈玉娟

浙江财经大学数据科学学院



# 一元线性回归模型

## 拟合程度的评价

### 1 拟合优度 (goodness of fit) :

指回归方程与各观测数据的接近程度。

### 2 度量拟合优度最常用的指标:

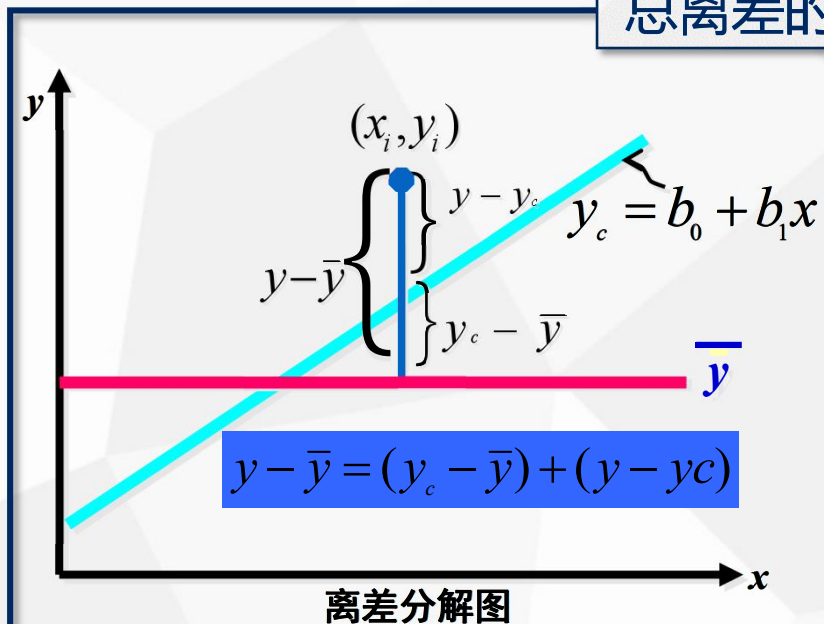
- 可决系数 (coefficient of determination , 又称决定系数、判定系数)



# 一元线性回归模型

## 判定系数

### 总离差的分解



$$\sum (y - \bar{y})^2 = \sum (y - y_c)^2 + \sum (y_c - \bar{y})^2$$

总变差  
(SST)

回归变差  
(SSR)

残差  
(SSE)

$$SST = SSR + SSE$$



# 一元线性回归模型

## 判定系数



判定系数（可决系数） $r^2$

计算公式

$$r^2 = \frac{SSR}{SST} = \frac{\sum (y_c - \bar{y})^2}{\sum (y - \bar{y})^2} = 1 - \frac{\sum (y - y_c)^2}{\sum (y - \bar{y})^2}$$



表明自变量X的方差对因变量y方差的解释程度，即y的方差中有多大程度是由x原因所引起的。判定系数是相关系数的平方。



判定系数 $r^2$ 取值范围在 $[0, +1]$ 之间，  
当 $r^2=1$ 时，剩余变差等于0，说明总变差可以完全由回归直线来解释，拟合是完全的。  
当 $r^2=0$ 时，回归变差等于0， $X$ 完全无法解释 $y$ 的变动。



# 一元线性回归模型

## 估计标准误



估计标准误 (standard error of estimate)

是对各观测数据在回归直线周围分散程度的一个度量值，反映了用估计的回归方程拟合因变量Y时平均误差的大小。可以证明， $S_{yx}$  是对误差项 $\varepsilon$ 的标准差 $\sigma$ 的无偏估计。



计算公式

$$s_{yx} = \sqrt{\frac{\sum (y - y_c)^2}{n - 2}}$$

各观测数据越靠近回归直线，估计标准误就越小，回归直线对各观测数据的代表性就越好。



## 估计标准误取值特性

显然， $SYX$ 的数值越小，说明估计值的代表性越大，观测点越靠近回归直线，离散程度越小。

特殊地，当 $syx=0$ ，表明实际值和估计值完全一致，观测点都落在回归直线上。

反之， $SYX$ 的数值越大，说明观测点的离散程度越大，直线方程的代表性越差，回归估计越不精确。



## 简捷计算公式

$$s_{yx} = \sqrt{\frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n - 2}}$$





# 一元线性回归模型

## 拟合优度的计算

以2000-2015年杭州市城镇居民年人均可支配收入与年人均消费支出数据资料为例，已经拟合城镇居民年人均消费支出倚年人均可支配收入的回归方程。

$$y_c = 1806.9 + 0.629x$$

计算判定系数和估计标准误。

年份	X	Y
2000	9668	7790
2001	10896	8968
2002	11778	9215
2003	12898	9950
2004	14565	11213
2005	16601	13438
2006	19027	14472
2007	21689	14896
2008	24104	16719
2009	26864	18595
2010	30035	20219
2011	34065	22642
2012	37511	22800
2013	39310	24833
2014	44632	32165
2015	48316	33818
合计	401959	281732



# 一元线性回归模型

## 前例拟合优度的计算

一元线性回归方程拟合值计算表

年份	X	Y	$y_c$	$(y - y_c)^2$
2000	9668	7790	7888.1	9618.12
2001	10896	8968	8660.5	94566.09
2002	11778	9215	9215.3	0.068644
2003	12898	9950	9919.7	901.08032
2004	14565	11213	10968.3	59777.805
2005	16601	13438	12248.9	1413889.8
2006	19027	14472	13774.8	485732.42
2007	21689	14896	15449.5	306647.3
2008	24104	16719	16968.1	61977.009
2009	26864	18595	18704.3	12003.825
2010	30035	20219	20698.9	230318.41
2011	34065	22642	23233.8	350209.49
2012	37511	22800	25401.3	6766860.5
2013	39310	24833	26532.9	2889626
2014	44632	32165	29880.4	5219269.2
2015	48316	33818	32197.7	2625488.8
合计	401959	281732	281742	20526886

每个 $y_c$ 与 $y$ 之间都有误差，利用这些样本数据，计算判定系数为：

$$r^2 = \frac{SSR}{SST} = 1 - \frac{\sum (y - y_c)^2}{\sum (y - \bar{y})^2} = 0.979$$

回归估计标准误为：

$$s_{yx} = \sqrt{\frac{\sum (y - y_c)^2}{n - 2}} = \sqrt{\frac{20526886}{14}} = 1210.87$$



# 一元线性回归模型

解：

一元线性回归方程计算表

年份	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
2000	9668	7790	93470224	60684100	75313720
2001	10896	8968	118722816	80425024	97715328
2002	11778	9215	138721284	84916225	108534270
2003	12898	9950	166358404	98997724	128332004
2004	14565	11213	212139225	125726435	163314141
2005	16601	13438	275593201	180579844	223084238
2006	19027	14472	362021401	209431259	275351771
2007	21689	14896	470428337	221883368	323079284
2008	24104	16719	580982569	279528305	402990164
2009	26864	18595	721670735	345764728	499528062
2010	30035	20219	902101225	408807961	607277665
2011	34065	22642	1160424225	512660164	771299730
2012	37511	22800	1407075121	519840000	855250800
2013	39310	24833	1545276100	616677889	976185230
2014	44632	32165	1992015424	1034587225	1435588280
2015	48316	33818	2334435856	1143657124	1633950488
合计	401959	281732	12481436148	5924167374	8576795176

人均消费支出倚人均可支配收入的一元线性回归模型为：

$$y_c = 1806.9 + 0.629x$$

相关系数为：  $r = 0.989$

判定系数为：

$$r^2 = 0.989^2 = 0.979$$



# 一元线性回归模型

解：

一元线性回归方程计算表

年份	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
2000	9668	7790	93470224	60684100	75313720
2001	10896	8968	118722816	80425024	97715328
2002	11778	9215	138721284	84916225	108534270
2003	12898	9950	166358404	98997724	128332004
2004	14565	11213	212139225	125726435	163314141
2005	16601	13438	275593201	180579844	223084238
2006	19027	14472	362021401	209431259	275351771
2007	21689	14896	470428337	221883368	323079284
2008	24104	16719	580982569	279528305	402990164
2009	26864	18595	721670735	345764728	499528062
2010	30035	20219	902101225	408807961	607277665
2011	34065	22642	1160424225	512660164	771299730
2012	37511	22800	1407075121	519840000	855250800
2013	39310	24833	1545276100	616677889	976185230
2014	44632	32165	1992015424	1034587225	1435588280
2015	48316	33818	2334435856	1143657124	1633950488
合计	401959	281732	12481436148	5924167374	8576795176

人均消费支出倚人均可支配收入的一元线性回归模型为：

$$y_c = 1806.9 + 0.629x$$

回归估计标准误为：

$$s_{yx} = \sqrt{\frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n-2}} \\ = 1210.87$$



# 一元线性回归模型

## 回归模型的统计检验

### 1 模型整体拟合效果的显著性检验:

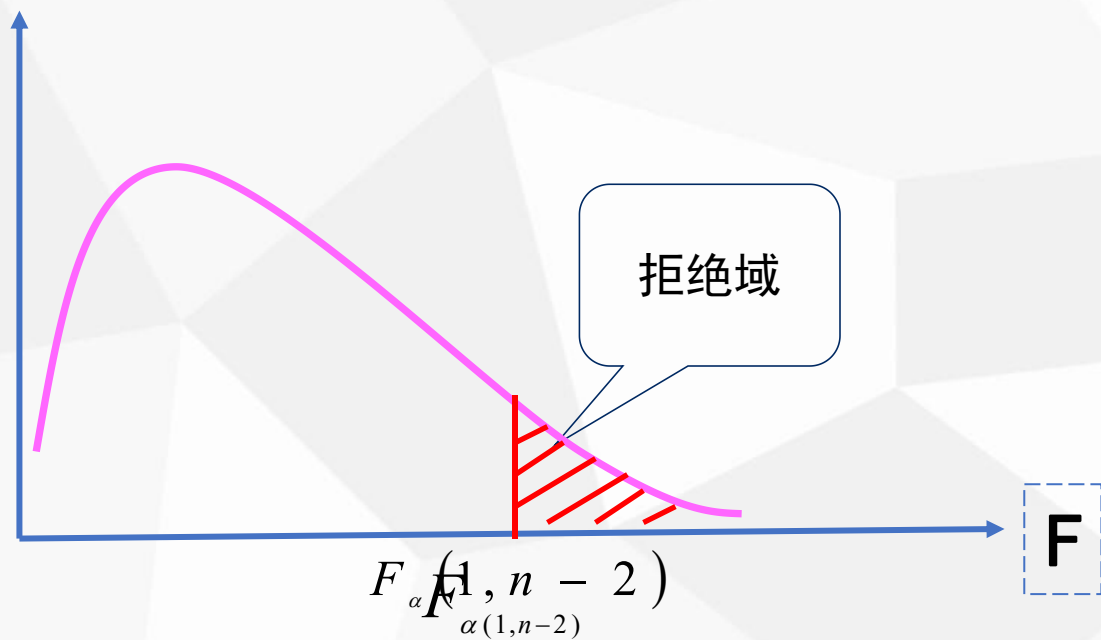
1 提出假设:  $H_0: \beta_0 = \beta_1 = 0$      $H_1: \beta_0 \neq 0$  或  $\beta_1 \neq 0$

2 构造检验统计量:  $F = \frac{SSR / 1}{SSE / (n - 2)} \sim F_\alpha(1, n - 2)$

3 根据给定的显著性水平 $\alpha$ , 确定临界值  $F_\alpha$ ;

4 确定原假设的拒绝规则, 决策:

若  $F > F_\alpha(1, n - 2)$ , 则拒绝 $H_0$ , 表示回归方程整体是显著的。





# 一元线性回归模型

## 回归模型的统计检验

### 2 回归系数的显著性检验：自变量对于回归模型是否必要

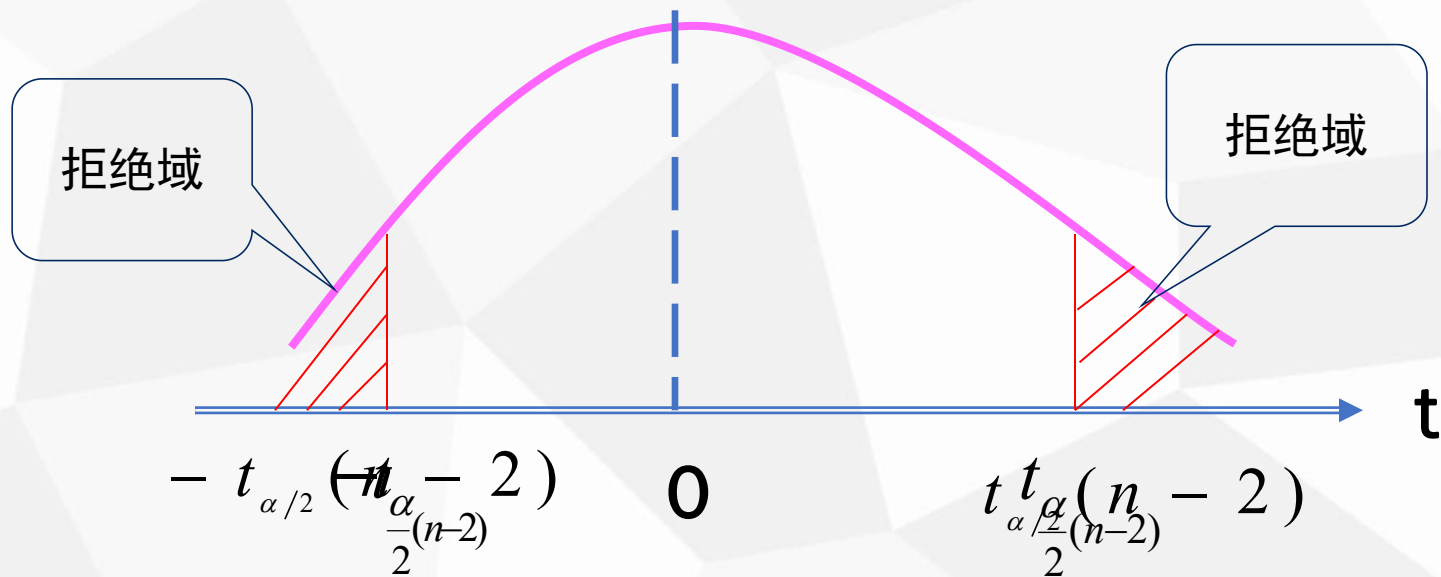
1 提出假设：  $H_0 : \beta_1 = 0$      $H_1 : \beta_1 \neq 0$

2 构造检验统计量： 
$$t = \frac{b_1}{S_{yx} \sqrt{\frac{1}{\sum (x - \bar{x})^2}}} \sim t_{\frac{\alpha}{2}}(n - 2)$$

3 根据给定的显著性水平 $\alpha$ ，确定临界值 $t_{\alpha/2}(n - 2)$ ；

4 确定原假设的拒绝规则，决策：

若  $|t| \geq t_{\alpha/2}(n - 2)$  则拒绝 $H_0$ ，表示回归系数对方程的影响是显著的。







对于已经拟合的2000-2015年杭州市城镇居民年人均消费支出倚年人均可支配收入的回归方程：

$$y_c = 1806.9 + 0.629x$$

对其进行相关f检验和t检验 ( $\alpha=0.05$ ) 。



# 一元线性回归模型

解：

## 模型整体拟合效果的显著性检验：

$$F = \frac{SSR / 1}{SSE / (n - 2)} = 643.05$$

在显著性水平 $\alpha=0.05$ 时，f分布的临界值

$$F_{\alpha}(1, n - 2) = 4.6$$

$$F > F_{\alpha}(1, n - 2)$$

拒绝原假设，表明拟合的回归方程整体是显著的

## 回归系数的显著性检验：

$$t = \frac{b_1}{S_{yx} \sqrt{\frac{1}{\sum (x - \bar{x})^2}}} = 25.36$$

在显著性水平 $\alpha=0.05$ 时，t分布的临界值

$$t_{\alpha/2}(n - 2) = 2.145$$

$$|t| \geq t_{\alpha/2}(n - 2)$$

拒绝 $H_0$ ，表示回归系数对方程的影响是显著的



浙江财经大学  
Zhejiang University of Finance & Economics

# 谢谢

---

日期：2017/08/23