

一元线性回归模型的构建



陈玉娟

浙江财经大学数据科学学院



一元线性回归模型

一元线性回归模型的构建

+ 理论模型

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon$$

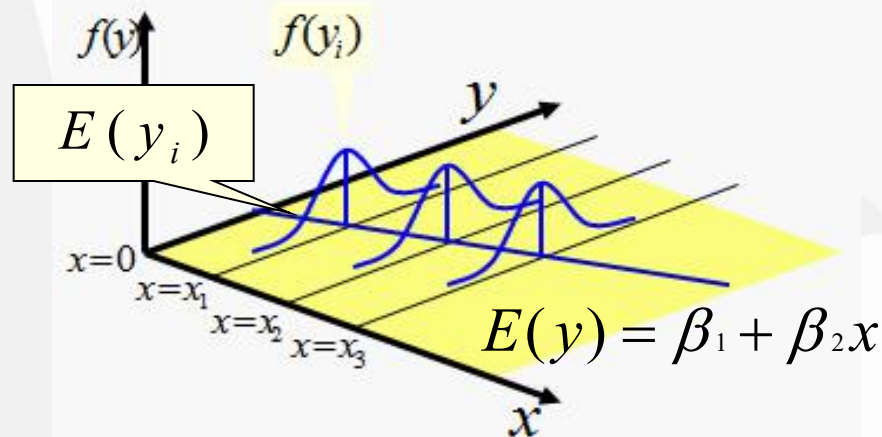
$$E(\varepsilon) = 0$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

$$\text{Var}(\varepsilon_i) = \text{Var}(y_i) = \sigma^2$$

$$\text{Cov}(\varepsilon_i, x_i) = 0$$

式中： β_0 与 β_1 为模型参数，



ε_i 为误差项随机变量



一元线性回归模型

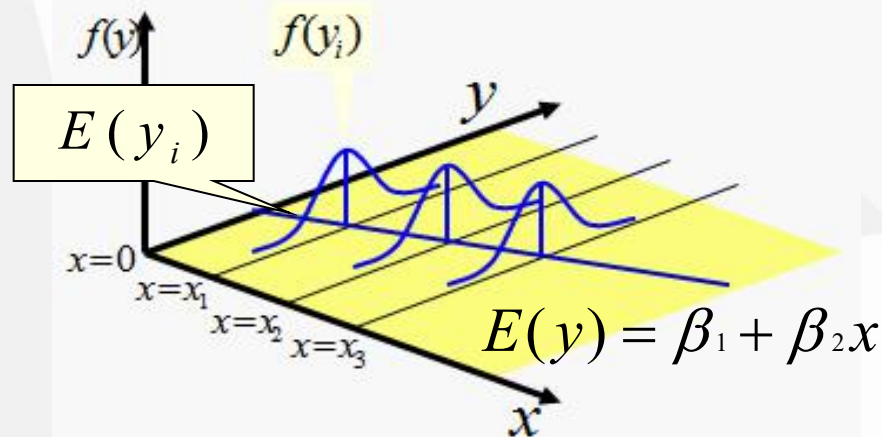
一元线性回归模型的构建

+ 估计模型

$$y_c = b_0 + b_1 x$$

b_0 称为截距

b_1 称为斜率（回归系数），经济含义是自变量变动一个单位，因变量平均变动 b_1 个单位。

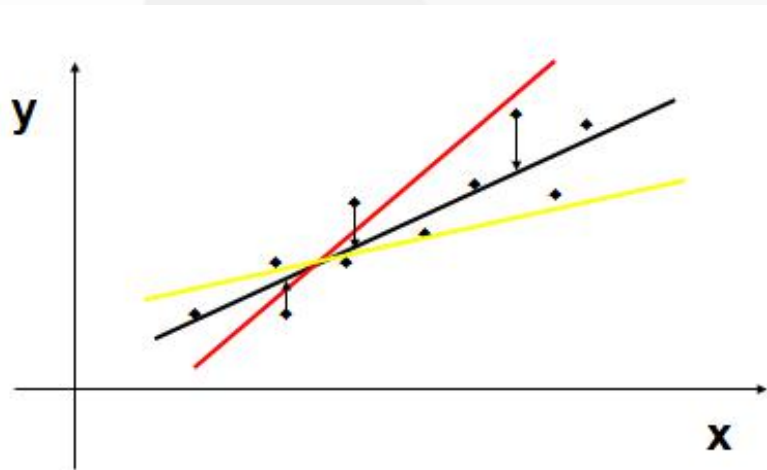




一元线性回归模型

参数估计

⊕ 最小二乘法 使得实际值 y 与估计值 y_c 的离差平方和最小。



基本原理

$$\sum (y_i - y_c)^2 = \min$$



一元线性回归模型

参数估计

⊕ 最小二乘法 使得实际值 y 与估计值 y_c 的离差平方和最小。

$$Q = \sum (y - y_c)^2 = \sum (y - b_0 - b_1 x)^2 = \min$$

$$\left\{ \begin{array}{l} \frac{\partial Q}{\partial b_0} = 0 \\ \frac{\partial Q}{\partial b_1} = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} 2 \sum (y - b_0 - b_1 x) \cdot (-1) = 0 \\ 2 \sum (y - b_0 - b_1 x) \cdot (-x) = 0 \end{array} \right.$$



一元线性回归模型

参数估计

⊕ 最小二乘法 使得实际值 y 与估计值 y_c 的离差平方和最小。

整理得:

$$\left\{ \begin{array}{l} \sum y = n \cdot b_0 + b_1 \cdot \sum x \\ \sum xy = b_0 \sum x + b_1 \sum x^2 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \hat{b}_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \\ \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \end{array} \right.$$



一元线性回归模型

例：

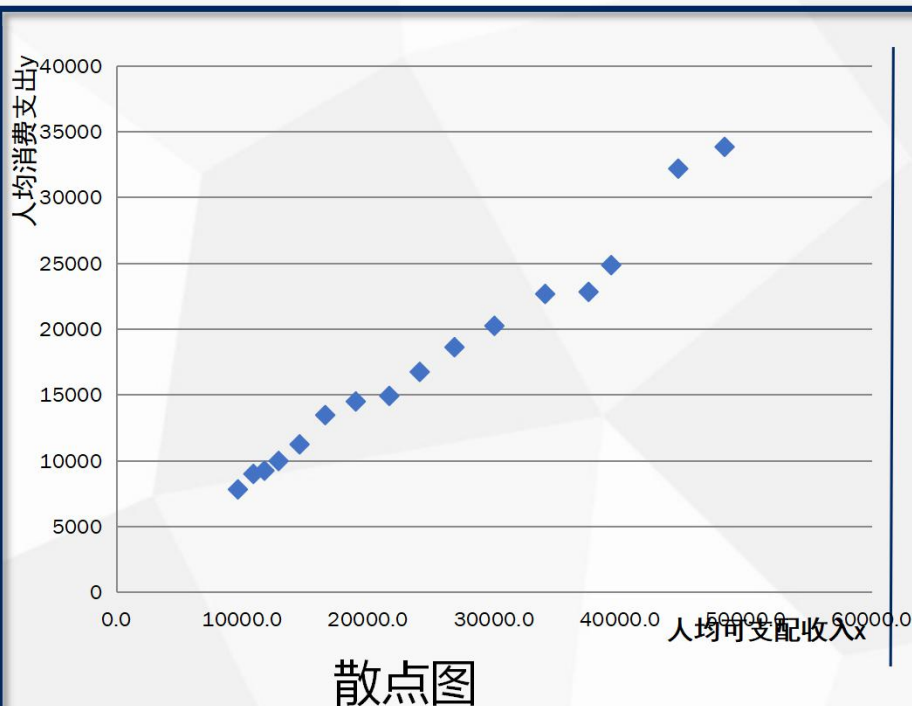
已知杭州市2000-2015年城镇居民人均可支配收入与人均消费支出数据资料如右表所示（单位：元），要求建立城镇居民人均消费支出倚人均可支配收入的回归模型。

年份	人均可支配收入	人均消费支出
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



一元线性回归模型

解：



人均消费支出随可支配收入变化而变化，因此，人均消费支出是因变量 y ，人均可支配收入为 x ，建立一元线性回归模型：

$$y_c = b_0 + b_1 x$$

利用最小二乘法估计参数，有

$$\sum y = n \cdot b_0 + b_1 \cdot \sum x$$

$$\sum xy = b_0 \sum x + b_1 \sum x^2$$



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 ²	Y ²	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

$$\hat{b}_1 = \frac{16 \times 8576795176 - 401959 \times 281732}{16 \times 12481436148 - 401959^2} = 0.629$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 1806.9$$

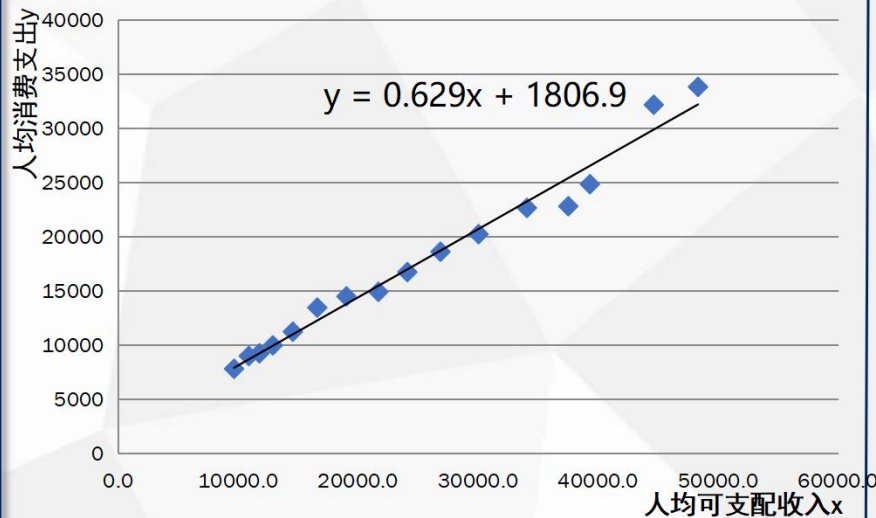
人均消费支出倚人均可支配收入的一元线性回归模型为：

$$y_c = 1806.9 + 0.629x$$



一元线性回归模型

解：



回归直线图

该回归直线的斜率为0.629，即杭州市城镇居民人均可支配收入每增加一元，人均消费支出平均增加0.629元。



浙江财经大学
Zhejiang University of Finance & Economics

谢谢

日期：2017/08/23