

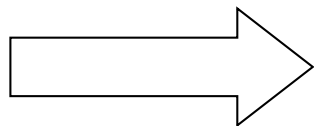
# 回归分析—相关分析

# 内容关联结构

一元回归  $y = a + bx$

直线回归

直线相关



$y \neq a + bx$

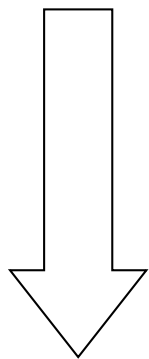
非线性回归

简单曲线

$y : x$

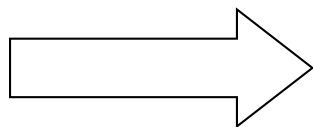
波浪曲线

$y : x^1, \dots, x^k$



多元回归

多元相关



多项式回归

$y : x_1, \dots, x_k$

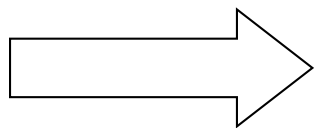
# 第九章 多元回归与多元相关分析

## 9—1 多元回归分析

## 9—2 多元相关分析

## 9—3 多项式回归分析

多元回归  
多元相关



多项式回归

$y : x$



$y : x^1, \dots, x^k$



$y : x_1, \dots, x_k$

## **9—1 多元回归分析**

- 一、多元线性回归模型
- 二、多元线性回归方程的建立
- 三、多元回归的假设检验和置信区间
- 四、应用
- 五、逐步回归

## **9—2 多元相关分析**

- 一、多元相关
- 二、偏相关

## **9—3 多项式回归分析**

- 一、多项式回归的数学模型
- 二、多项式回归方程的建立
- 三、多项式回归方程的假设检验
- 四、相关指数

# 9-1 多元回归分析

## 一、多元线性回归模型

多元线性回归的数学模型：

$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon_i$$

式中， $\alpha$ 是截距或常数项

$\beta_1$ 是 $x_2, x_3, \dots, x_m$ 固定不变时 $x_1$ 对 $y$ 的偏回归系数

$\varepsilon_i$ 为随机误差，服从正态分布 $N(0, \sigma^2_{y/x_1, x_2, \dots, x_m})$

$\sigma^2_{y/x_1, x_2, \dots, x_m}$ 为离回归方差

样本多元线性回归方程：

$$y_i = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e_i$$

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m$$

样本多元线性回归方程：

$$y_i = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + e_i$$
$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$$

数据结构：① 教材第四版P216 表12—1  
② 大学生提供数据

体重	身高	年龄	性别	省份	城乡	海拔	兄弟姐妹
Y	x1	x2	x3	x4	x5	x6	x7
60	176	21	0	0	0	0	0
58	170	21	0	0	2	0	0
52	170	21	0	0	3	0	0
80	178	21	0	0	2	1	0
68	177	20.5	0	0	3	0	0
65	169	20.6	0	1	3	0	0
50	169	19.7	0	0	1	0	0
70	164	21	0	0	3	0	0
51	165	22	0	1	2	0	0
65	181	20	0	1	3	1	1
60	170	20	0	0	3	0	0
80	171	20	0	1	2	2	1
55	167	21	0	1	0	0	0
60	172	21	0	0	2	2	0
44	154	20	1	0	2	0	0

## 二、多元线性回归方程的建立

与直线回归方程一样，也采用最小二乘估计：

$$Q = \sum (y - \hat{y})^2 = \text{最小值}$$

正规方程组（用矩阵、向量表示）：

$$A b = K$$

解方程组，得：

$$b = A^{-1} K$$

$A^{-1}$ 为矩阵A的逆矩阵： $A^{-1} = (c_{ij})_{m \times m}$

例如： $3b = 1$

$$b = 3^{-1} \times 1 = 1/3$$

$$\begin{aligned} an + b \sum x &= \sum y \\ a \sum x + b \sum x^2 &= \sum xy \end{aligned}$$

$b_0$   
 $b_1$   
 $b_2$   
 $\dots$   
 $b_m$

$$a = b_0$$

$$\begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

=

$$\begin{pmatrix} \sum y \\ \sum xy \end{pmatrix}$$

### 三、多元回归的假设检验和置信区间

#### 1、多元线性回归方程的估计标准误

总平方和 (**SSy**) = 回归平方和(**U**) + 离回归平方和(**Q**)

$$SSy = U_{y/x_1, x_2, \dots, x_m} + Q_{y/x_1, x_2, \dots, x_m}$$

自由度:  $(n-1) = (m) + (n - m - 1)$

$$U_{y/x_1, x_2, \dots, x_m} = b_1 SP_{1y} + \dots + b_m SP_{my}$$

其df = m

$$Q_{y/x_1, x_2, \dots, x_m} = SSy - U_{y/x_1, x_2, \dots, x_m}$$

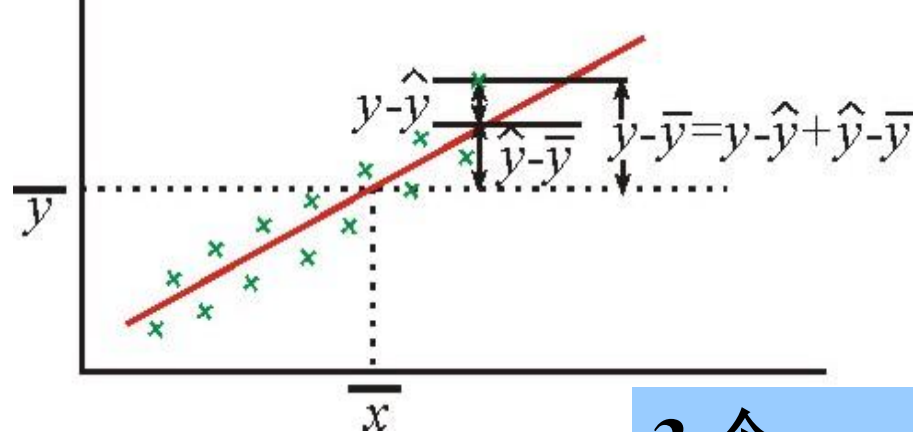
其df = n - m - 1

由多元回归方程的离回归平方和  $Q_{y/x_1, x_2, \dots, x_m}$ ，及自由度  $df = n - m - 1$ ，定义多元线性回归方程的估计标准误：

$$s_{y/x_1, x_2, \dots, x_m} = \sqrt{[Q_{y/x_1, x_2, \dots, x_m} / (n - m - 1)]}$$

# 仿 直线回归的变异来源

P128



3 个 y

平方和分解:

$$\begin{aligned} SS_y &= \sum (y - \bar{y})^2 = \sum (y - \bar{y} + \hat{y} - \hat{y})^2 \\ &= \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2 \end{aligned}$$

$$\text{回归平方和 } U = \sum (\hat{y} - \bar{y})^2$$

$$= b_1 SP_{1y} + \dots + b_m SP_{my}$$

$$\text{离回归平方和 } Q = \sum (y - \hat{y})^2 = SS_y - U$$

$$\text{自由度分解: } (n-1) = m + (n - m - 1)$$

$$\text{两个方差: 回归方差} = U / m$$

$$\text{离回归方差} = s_{y/12\dots m}^2 = Q / (n - m - 1)$$

$$\text{离回归标准差, 或回归估计标准差: } s_{y/12\dots m} = \sqrt{[Q / (n - m - 1)]}$$



# 公式推导

回归平方和：

$$\begin{aligned} U &= \sum (\hat{y} - \bar{y})^2 \\ &= \sum \{ \bar{y} + b_1(x_1 - \bar{x}_1) + \dots + b_m(x_m - \bar{x}_m) - \bar{y} \}^2 \\ &= \sum \{ b_1(x_1 - \bar{x}_1) + \dots + b_m(x_m - \bar{x}_m) \}^2 \\ &= b_1^2 \sum (x_1 - \bar{x}_1)^2 + \dots + b_m^2 \sum (x_m - \bar{x}_m)^2 \\ &= b_1 SP_{1y} + \dots + b_m SP_{my} \end{aligned}$$

离均差乘积和等于 0 ？

**P214： 公式12.3**

$$\hat{y} = \bar{y} + b_1(x_1 - \bar{x}_1) + \dots + b_m(x_m - \bar{x}_m)$$

## 2、多元线性回归方程的假设检验——F 测验

### 方差分析

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_A: \beta_1, \beta_2, \dots, \beta_m \text{不全为} 0$$

不包括 $\beta_0$

$$F = \frac{U_{y/x_1, x_2, \dots, x_m} / m}{Q_{y/x_1, x_2, \dots, x_m} / (n - m - 1)} \sim F(m, n - m - 1)$$

**P220: 表12-3 例12.1资料多元线性回归关系的方差分析**

变异来源	df	SS	$s^2$	F	$F_{0.05}$	$F_{0.01}$
回 归	3	3.8576	1.2859	23.4653**	6.59	16.69
离回归	4	0.2193	0.0548			
总变异	7	4.0769				

$$F = \frac{U_{y/123}/m}{Q_{y/123}/(n - m - 1)} = \frac{3.8576/3}{0.2193/(8 - 3 - 1)} = 23.4653^{**}$$

### 3、偏回归系数的假设检验——t 检验、F检验

EXCEL和SPSS等统计软件多用 t 检验:

$$H_0: \beta_i = 0$$

$$H_A: \beta_i \neq 0$$

包括 $\beta_0$

偏回归系数的标准误:  $s_{b_i} = s_{y/x_1, x_2, \dots, x_m} \sqrt{c_{ii}}$

$$t = \frac{(b_i - \beta_i)}{s_{b_i}} \\ = \frac{b_i}{s_{b_i}} \\ \sim t(n - m - 1)$$

$i=0, 1, \dots, m$

$$\beta_0 = \alpha$$

其  $c_{ii}$  为逆矩阵 $A^{-1}$ 对角线上的元素

	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-31.4	11.7206	-2.679	0.02007	-56.937	-5.8625
比叶重x1	15.3518	5.32497	2.88299	0.01375	3.74973	26.954
气孔x2	0.65299	0.27906	2.33996	0.03739	0.04497	1.26101
叶绿素x3	0.93254	0.86949	1.07252	0.30458	-0.9619	2.82699

# 4、多元线性回归的区间估计

回归截距的区间估计:  $\alpha$   $\beta_0$   
回归系数的区间估计:  $\beta_1, \beta_2, \dots, \beta_m$

依变量y的区间估计:

$\mu_{y/x_1, x_2, \dots, x_m}$ 的区间估计: 教材P223公式 (12.30)  
单个 y 的区间估计: 教材P223公式 (12.31)

Coefficient		标准误差	t Stat	P-value	lower 95%	Upper 95%
Intercept	-31.4	11.7206	-2.679	0.02007	-56.937	-5.8625
比叶重x1	15.3518	5.32497	2.88299	0.01375	3.74973	26.954
气孔x2	0.65299	0.27906	2.33996	0.03739	0.04497	1.26101
叶绿素x3	0.93254	0.86949	1.07252	0.30458	-0.9619	2.82699

## 四、应用：

教材第二版P197 例12.1

红薯光合速率  $y$  与 3 个自变量的关系

$$Y = -31.4 + 15.35 x_1 + 0.65 x_2 + 0.93 x_3$$

比叶重 $x_1$	气孔 $x_2$	叶绿素 $x_3$	光合 $y$
1.9993	11.4	4.0575	11.7161
2.0254	8.1	3.775	6.9862
2.001	10.7	3.3733	11.3444
2.1072	11.2	3.1352	12.477
1.8941	9	3.519	5.9618
2.0188	12.5	3.4278	11.221
1.9362	10.1	3.8518	8.8416
2.1072	8.5	4.1373	7.9483
1.9843	8.3	4.2719	9.8014
1.9904	10.8	4.9872	11.0765
1.7836	10.7	3.0091	6.3744
1.973	8.8	4.3073	9.3993
1.9414	10.2	4.3965	9.842
2.0519	9	4.1673	8.251
1.9626	11.1	4.0186	10.64
1.8651	14.2	3.4175	6.6433

### 方差分析

	df	SS	MS	F	Significance F
回归分析	3	32.8846	10.9615	4.15342	0.03108
残差	12	31.67	2.63916		
总计	15	64.5546			

### 回归统计

R	0.71373
R Square	0.50941
Adjusted	0.38676
标准误差	1.62455
观测值	16

	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-31.4	11.7206	-2.679	0.02007	-56.937	-5.8625
比叶重 $x_1$	15.3518	5.32497	2.88299	0.01375	3.74973	26.954
气孔 $x_2$	0.65299	0.27906	2.33996	0.03739	0.04497	1.26101
叶绿素 $x_3$	0.93254	0.86949	1.07252	0.30458	-0.9619	2.82699

## 五、逐步回归

### 回归分析的用途：

- (1)、因素分析：例如大学生体重的影响因素分析
- (2)、预测预报：例如天气预报、虫情预报
- (3)、统计控制：见倪宗瓚《卫生统计学》P121

### 最优回归方程：

将不具有显著效应的自变量舍去，使所得到的多元线性回归方程中的自变量对依变量  $y$  均具有显著效应。这样的回归方程叫最优回归方程。

### 最优回归方程的获得途径：

- ①、逐个淘汰不显著自变量的回归方法：从多到少
  - ②、逐个选入显著自变量的回归方法：从少到多
- 通常称此法为逐步回归方法

# 逐步回归:

红薯光合速率  $y$  与 3 个自变量的关系:

$$Y = -31.4 + 15.35 x_1 + 0.65 x_2 + 0.93 x_3$$

删掉 $X_3$ ,

求出最优回归方程:

$$Y = -28.8 + 16.35 x_1 + 0.56 x_2$$

	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-31.4	11.7206	-2.679	0.02007	-56.937	-5.8625
比叶重 $x_1$	15.3518	5.32497	2.88299	0.01375	3.74973	26.954
气孔 $x_2$	0.65299	0.27906	2.33996	0.03739	0.04497	1.26101
叶绿素 $x_3$	0.93254	0.86949	1.07252	0.30458	-0.9619	2.82699

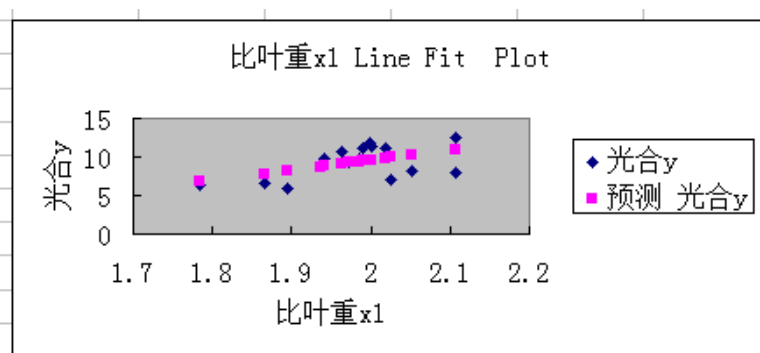
回归统计								
Multiple R	0.679986							
R Square	0.462381							
Adjusted R Square	0.37967							
标准误差	1.633914							
观测值	16							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	2	29.84881	14.92441	5.590348	0.017705			
残差	13	34.70576	2.669674					
总计	15	64.55457						
		Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%	
Intercept		-28.8054	11.53443	-2.49734	0.026722	-53.724	-3.88677	-53.724 -3.88677
比叶重 $x_1$		16.35058	5.273132	3.100734	0.008435	4.958672	27.74249	4.958672 27.74249
气孔 $x_2$		0.559256	0.266549	2.098135	0.056002	-0.01659	1.135101	-0.01659 1.135101

# 删掉X2

## 红薯光合速率 y 与 1 个x的关系:

$$Y = -16.7 + 13.14 x_1$$

SUMMARY OUTPUT								
回归统计								
Multiple R	0.52946							
R Square	0.280328							
Adjusted R Square	0.228923							
标准误差	1.821658							
观测值	16							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	1	18.09645	18.09645	5.453306	0.034932			
残差	14	46.45812	3.318437					
总计	15	64.55457						
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	-16.6969	11.13443	-1.49958	0.155934	-40.5779	7.184032	-40.5779	7.184032
比叶重x1	13.13703	5.625581	2.335232	0.034932	1.071363	25.20271	1.071363	25.20271





补充实例：1

2002年

大学生体重的

影响因素分析

方差分析						
	df	SS	MS	F	Significance F	
回归分析	13	5990.15	460.781	11.8866695	4.5E-12	
残差	59	2287.1	38.7645			
总计	72	8277.25				
	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-54.486	50.422	-1.0806	0.28427271	-155.38	46.4083
身高 x1	0.91641	0.18303	5.00689	5.3135E-06	0.55017	1.28265
年龄 x2	-0.8114	0.97301	-0.8339	0.40769652	-2.7584	1.13559
性别 x3	-5.8133	2.52278	-2.3043	0.02474253	-10.861	-0.7652
省份 x4	0.44626	1.67639	0.2662	0.79101109	-2.9082	3.80072
城乡 x5	1.42812	0.94763	1.50705	0.1371348	-0.4681	3.32432
海拔高x6	2.88534	1.25041	2.30752	0.02455191	0.38328	5.38739
兄弟姊妹x7	0.11075	1.17812	0.094	0.92542536	-2.2467	2.46816
父年龄x8	0.16032	0.33507	0.47847	0.63408467	-0.5102	0.8308
体重x9	0.06174	0.09412	0.65601	0.51436717	-0.1266	0.25008
身高x10	-0.0524	0.23015	-0.2278	0.82062307	-0.513	0.40812
年龄x11	0.14151	0.39863	0.355	0.72385518	-0.6561	0.93918
体重x12	0.11109	0.11581	0.95922	0.3413641	-0.1206	0.34282
身高x13	-0.2646	0.1864	-1.4197	0.16094912	-0.6376	0.10835

回归统计	
Mult R	0.8507
R Square	0.72369
Adjusted	0.66281
标准误差	6.22611
观测值	73

补充实例： 1

2002年

大学生身高的

影响因素分析

		方差分析							
			df	SS	MS	F	ificance F		
		回归分析	12	3144. 53	262. 044	13. 5872969	5. 6E-13		
		残差	60	1157. 16	19. 286				
		总计	72	4301. 68					
回归统计		Coefficient		标准误差	t Stat	P-value	lower 95%	pper 95%	
		Intercept	26. 3857	35. 4015	0. 74533	0. 45898461	-44. 428	97. 1993	
		年龄 x2	x2	0. 58871	0. 68209	0. 86311	0. 39151643	-0. 7757	1. 95309
		性别 x3	x3	-10. 796	1. 10622	-9. 7596	5. 3254E-14	-13. 009	-8. 5835
		省份 x4	x4	2. 34025	1. 14319	2. 04712	0. 04503349	0. 05352	4. 62697
		城乡 x5	x5	1. 27943	0. 64768	1. 97542	0. 05282571	-0. 0161	2. 57498
		海拔高x6	x6	0. 18206	0. 88166	0. 2065	0. 837102	-1. 5815	1. 94564
		兄弟姊妹x7	x7	0. 14131	0. 83078	0. 17009	0. 86550776	-1. 5205	1. 80313
		父年龄x8	x8	-0. 3824	0. 23113	-1. 6547	0. 10321305	-0. 8448	0. 07988
		体重x9	x9	-0. 0823	0. 06553	-1. 2563	0. 21388103	-0. 2134	0. 04876
		身高x10	x10	0. 3806	0. 15472	2. 45987	0. 01679684	0. 07111	0. 69009
		m年龄x11	x11	0. 2203	0. 27973	0. 78753	0. 43407077	-0. 3392	0. 77985
		体重x12	x12	0. 1133	0. 08037	1. 40979	0. 16376417	-0. 0475	0. 27406
		身高x13	x13	0. 45472	0. 11764	3. 86534	0. 00027512	0. 2194	0. 69004

## 补充实例：

### 2、《科学》文章：女性的长寿与家庭大小无关 与孩子性别有关

贵州日报, 2002. 5. 13.

# 英芬学者一项研究称 生男减寿 生女延年

别, 男孩多的母亲寿命最短, 每个儿子平均减去她 34 周的寿命, 反而女儿则平均增加母亲 23 周的寿命。”

研究人员特

一项研究称, 多生男孩可能会使女性减寿, 但生女孩却会使人长寿。每生一个儿子令母亲“减寿”34 周, 生女则令母亲“添寿”23 周。

长久以来, 科学家一直怀疑大家庭会导致为人父母者减寿, 为了照顾儿女, 做父母的往往劳心劳力。但根据英国剑桥大学和芬兰芬兰大学在《科学》杂志发表的研究报告, 这一想法只对了一半。根据研究, 生男孩特别令

母亲“伤身”。研究人员认为, 男孩除了体型较大, 需要母亲较多体力生产外, 在成长过程中也花费母亲较多时间和精神。此外, 女性在怀男胎期间, 体内睾丸素会增加, 其免疫系统可能被削弱。相比之下, 女儿会令母亲添寿, 因为女儿多数比较听话, 她们会分担家务, 减轻母亲负担。

一名研究员说: “我们的研究结果显示, 女性的长寿与家庭大小无关, 而是决定于孩子的性

别。芬兰北欧在工业革命前(1640 年至 1870 年)的教会记录来研究, 避免现代社会中因医疗素质或家庭收入等造成太大差异。芬兰研究员赫拉称: “在男孩多于女孩的家庭, 生男对母亲造成的负面影响最明显。”研究人员称, 发展中国家的母亲也可能出现上述现象, 但对于发达国家, 由于社会倾向小家庭和营养充足, 生男是否有碍母亲健康则未能下结论。(东行)

# 生养儿子或可缩短母亲寿命

2012年07月30日 09:51 来源：凤凰网亲子论坛

- 对母亲来说，生儿子是一件艰苦的事。**首先**，生产时，一般男婴会比女婴更重一些；**其次**，生儿子时，母亲体内睾丸激素含量会上升；**当然**，男孩长大后比女孩更加顽皮吵闹，这也让母亲们头疼不已。英国谢菲尔德大学的维尔皮·卢玛在查阅了过去两百年来芬兰的一项记录后证实：**生儿子会缩短母亲的寿命，平均长度为34周。**
- 今年33岁的维尔皮·卢玛是芬兰的进化生物学家，在系谱专家的帮助下，她查阅了几百年前的资料卷宗和几十年前的缩影胶片，试图找出进化对人类繁衍的影响。历史学家、经济学家，甚至社会学家都曾用这种方法在各自的领域进行研究，而卢玛则是首批将现代人类作为一种动物来进行研究的生物学家，只不过我们这种动物的数量有资料可查。
- 对萨米人的研究结果显示，生育儿子的母亲，寿命比生育女儿的母亲短，这和婴儿的体重以及睾丸激素水平有关。卢玛表示，睾丸激素会危及免疫系统，影响健康。生了儿子的母亲们对流行性传染病(如肺结核)尤为敏感；“生儿子的代价”比生女儿大，还因为男孩会从母体吸取更多养分，这和其他哺乳动物(如赤鹿)的研究也是吻合的。当然，男孩长大以后也不太会像女儿一样帮妈妈忙。
- 有兄长并非是好事
- 最近，卢玛和她的同事研究发现，生儿子不仅对母亲来说有负面影响，对同胞兄弟姐妹也一样。如果前一胎是男孩，那么接下来出生的孩子个头就会比较矮小，长大后家庭成员会比较少，而且更容易染病致死。即使先出生的那个男孩夭折，对后出生的孩子产生的影响还是一样，这就是说，对这种负面影响相互作用产生的，同胞间的相互作用包括争食、打架及长子继承权等。因此，卢玛说：“如果第五胎是个男孩，那第六个出生的会更糟。”
- 这一现象在龙凤胎中更为明显。对150年间芬兰5个乡镇出生的754个双胞胎的资料进行统计发现，龙凤胎中的女性结婚率比双凤胎低15%，生孩子的几率低25%，而且平均比后者少生两个孩子。





# 3、你幸福吗？ 你快乐吗？

## 英国心理学家推出

### 幸福公式

水,那就更好。我弟弟的幸福观和我有点不一样,他认为最幸福的事莫过于过年的时候吃上韭黄炒年糕,加点火腿肉片更好。在那个时候,幸福的目标具体而实在,稍微努力一把,也能够着。后来看书多了,接触了一些哲学知识和逻辑知识后,幸福的概念开始抽象起来,人也就慢慢变得迷茫而困顿。我不知道原因,不知道是不是人人都像我这样。

我写文章有个毛病,就是喜欢引用名人名言,如果我在一篇讲幸福的文章里继续引用的话,就有卖注水肉的嫌疑。因为从古希腊城邦时代开始,一直有人对幸福问题喋喋不休,名人名言的资源相当丰富,引用起来,毫不费力。但说的人多了,各种观点旁逸斜出,众说纷纭,根本不知道该听谁的,只好作罢。我所了解到的情况是,在最早的时候,亚里士多德他们认为幸福源于肉体的快乐,也就是热被窝和炒年糕阶段。后来哲学界的克鲁伊夫·康德提出了他的意见,

热被窝和炒年糕,在半空中虚无飘渺得无法捉摸,让一干追求幸福又急需理论依据的人心急如焚。

为了解决这个问题,最近英国科学家(其实我认为是社会学家才对)给出了关于幸福的一个公式,幸福= $P+(5\times E)+(3\times H)$ 。在这里,P代表个性;E代表生存状况;H代表更高一层的需要,包括自尊心、期望、雄心和幽默感。作为一个前理科学生,我看不出它比康德的理论哪里更具操作性,我只发现公式里H和E的比例接近黄金分割(0.618)。除此之外,一无是处。另外一个解释是拆字法,把幸福的“幸”字拆开,发现“土”字下面有个人民币符号“¥”,这样的唯金钱论,相信也不会得人心。

想来想去,还是觉得最初的幸福最实在,但最实在的幸福现在已经找不到。成长的过程,其实就是一个“谁动了我的热被窝”的过程。要找回热被窝,先睡一阵凉炕再说。

$$\text{幸福} = P + 5E + 3H$$

P: 个性

E: 生存状况

H: 更高层次的需求

2013年4月25日,俄罗斯总统普京与民众进行电视连线互动,回答了民众提出的85个问题,当被记者提问:“你幸福吗?”普京似乎没有意料到会有这个问题,听后迟疑了一下,他说:这是一个哲学问题.....他表示,他感谢俄罗斯人民选他当国家元首,“这是我的全部生活,这对幸福来说是否足够,我不知道”。



# 你幸福吗？

- 一位清徐县北营村务工人员，在面对记者提问时，首先推脱了一番：“我是外地打工的，不要问我。”记者继续追问道：“您幸福吗？”这位务工人员认真答道：“我姓曾”。
- 因为这名务工人员的无厘头回答，网友们举一反三在微博中发明了一段记者与尔康的对话。  
“您幸福吗？”尔康答：“对，我姓福。”“您满足吗？”尔康答：“恩，我是满族的。”

一位在郑州就读的大学生与记者的对话，  
“你觉得幸福吗？”“幸福啊。”“你觉得幸福是什么呢？”“每天把该做的事做完之后，舒舒服服地玩就是幸福。”“最想要什么？”“女朋友！”“那最坏的事呢？”“最坏的事是接受你采访，队被人插了！”

记者在清华大学采访一位在校大学生，这位大学生的搞笑回答乐翻网友，“你幸福吗”“我天，我不会还会上新闻联播吧！我今天不幸福”“为什么啊？”微笑回答：“女朋友刚分手。”





# 中国人的幸福感

- 美国南加州大学的研究人员发表了题为“**China's life satisfaction, 1990-2010**”的文章，指出在过去的二十年里，**处于国内收入下层的三分之一人群对其生活满意度越来越低，而收入最上层的三分之一人群则越来越满意**。这说明经济增长并不意味着更高的生活满意度，幸福感。
- 相关成果公布在美国《国家科学院院刊》（**PNAS**）杂志上。
- 这项研究由著名人口经济学家**Richard Easterlin**领导完成，他指出，尽管中国经济正在以前所未有的速度向前发展，但是国内人民却比二十年前幸福感更低。
- **Easterlin**教授是幸福经济学（**happiness economics**）研究领域的奠基者。

## 金钱与幸福并不存在正比关系

- 金钱能换来幸福吗？这个问题一直引起人们争议。美国研究人员在对全球众多国家进行调查后认为，从长远观点看，金钱并不能换来幸福。
- 该研究由美国南加州大学的研究人员完成，研究共涉及54个国家，其中包括17个拉美国家、17个发达国家、多个经历了体制转型的东欧国家，以及9个新兴工业化国家。研究人员分析了与幸福有关的各种数据，如经济增长率、收入情况和生活水平等，然后把这些数据与人们的幸福感进行对比，**但研究并没有发现足以证明经济增长能提高人们幸福感的有关证据**。
- 研究得出的结论是：从长远观点看，金钱的多少与幸福指数的高低并不存在正比关系。这一结论既适用于发达国家，也适用于那些刚摆脱了贫困而实现经济快速增长的发展中国家。这项研究成果刊登在美国《国家科学院院刊》（PNAS）上。
- 负责这项研究的经济学家教授理查德·伊斯特林指出：“幸福不会随着经济增长而增长，无论是在新兴工业化国家还是在体制转型国家。我们早就知道，在发达国家，金钱买不到幸福，现在通过比对，这一判断也适用于收入水平较低的国家。”
- 伊斯特林说，经济增长可能会提高人们短暂的幸福感，但随着时间推移，这种幸福感会很快消失。他解释说，**随着收入增加，人们的欲望也会增加，而且在很多情况下，收入的增加赶不上欲望的上升，正是这种落差抵消了人们的幸福感**。
- 他进一步分析说，影响人们幸福指数的另外一个因素是：人们很难遏制已经产生的欲望，这就是为什么当收入降低后，人们的欲望并不会随之降低，因为谁也不会放弃已经习惯的好日子。

## • 补充实例

### 4、疾病、死亡 爱情、婚姻 容貌、微笑

据新华社 10 月 29 日电 一场秋雨一场寒，武汉最近气温骤降，染上感冒的人不少，打针吃药之余，不妨试试英国爱丁堡睡眠研究中心负责人克里斯·伊齐科斯基发明的全新治疗感冒的公式： $I4 + (X \times T3) + (Y \times I1) - A1 - T4 + T2 - I3 + (2 \times (P + P2)) + L1 =$  治愈感冒。

用公式来治感冒可是个新鲜事，不过这个公式也不是随手写出来的。据悉，这一感冒治疗公式是伊齐科斯基和同事调查 2000 人，综合 4 组数据后得出的结论。这 4 组数据分别是影响病人睡眠质量的 4 个因素，分别为温度、光线、位置和

睡前饮食。

可别看花了眼。公式中字母分别代表感冒病人睡前的一些习惯行为。公式解释了各项行为是否有利于病人入睡。

I4 为感冒药。这些药剂，特别是含有扑热息痛成分的药剂能减缓头痛等症状。

I1 代表睡前喝些香甜热酒。伊

英国科学家提出全新治病概念

## 想治好感冒 依公式行事

齐科斯基认为，这能减轻病人痛苦，但注意不要超过两杯。

Y 指的是打开窗户通风，这有助睡眠时呼吸畅通。

T2 代表睡前洗个热水澡。这有助于皮肤毛细血管舒张。

病人睡前最好不要看电视 (A1)，这会妨碍病人入睡，尤其在鼻塞头痛时。

穿着袜子上床 (T4) 和抱着热水袋反而会让人睡不着觉。伊齐科斯基说：“人脑想要降低体温以便进入睡眠状态。你不应营造过暖环境，当然也不要冻得发抖。”

伊齐科斯基认为，人们另一个常犯错误是睡前吃镇静药，这会影响人呼吸。

中老年人 4 年内死亡率可测试

## 美国推出“生命公式”

本报综合报道 未来 4 年内，你的健康状况会对生命构成威胁吗？要回答这个耸人听闻的问题，做个小测试就行了。不过，这个由美国一个研究小组推出的测试只适用于 50 岁以上的人。

《美国医学会协会》15 日发表了这份测试卷。测试共列出 12 项对健康构成威胁的因素，测试者根据自身情况选择对应分数，总得分越高表明越危险。



# 爱情方程式

$$\text{爱} = 1.7A + 1.5B + 1.5C + 1.5D + 1.3E$$

**A:** 你对朋友的总体感觉

**B:** 朋友令你感受心灵慰藉的程度

**C:** 在身体上接近朋友的渴望

**D:** 你希望那人对你的需要与关怀程度

**E:** 对关系破裂的恐惧

每个因素评分10分为满分

美心理学家新著推出

## 真假爱情方程式

本报综合消息 罗密欧和朱丽叶或许不会察觉,爱情可能真有方程式。美国顶尖心理学家纽约大学教授伯恩斯透露,他已找出这一方程式,能分辨那些爱情是真心,那些是假意。

伯恩斯将于今年7月推出的最新著作《社会心理学》中,详细公开了这条“爱情方程式”,他试图对力量巨大然而却神秘难测的男女感情世界作出科学的诠释。他表示,真爱包含5种元素:性吸引、心灵慰藉刺激、渴求亲密关系、强烈渴求被爱和被认同、经常害怕失去所爱的人。

他补充,由于爱情属于个人主观,只有当事人才能决定哪个元素最重要,而在他所定出的爱情方程式中,不同元素所占的比重,主要是通过

对北美、英国和荷兰一连串有关男女爱情的心理测验和访问而定出的。

将人际关系“量化”,是社会心理学家的热门课题之一,尽管很多人都认为这种研究“浪费时间”。不过,社会科学家近年在美国婚礼上进行的一连串访问和资料收集,却为量化研究注入一剂强心针。专家按照这些访问数据,预测哪些婚姻可维持两年以上,准确率高达80%。

爱情方程式如下:

爱 = (1.7 × A) + (1.5 × B) + (1.5 × C) + (1.5 × D) + (1.3 × E) 每个元素的评分 10 分为满分。A—你对朋友的总体感觉。B—朋友令你所感受到的心灵慰藉振奋程度。C—在身体上接近朋友的渴望。D—你希望那人对你的需要和关怀程度。E—对关系破裂的恐惧。

长江时报 99.4.27

# 数学家发明公式可算出最佳求婚时间<http://en.eol.cn> 2010-03-02



- 担心男友不向你求婚？给他买个计算器吧！
- 数学家研究出一个“求婚公式”，让男士们可以计算出求婚的最佳时间。
- 计算过程中，只需提供最初考虑结婚的年龄和最迟结婚期限，剩下的都交给公式来处理。
- 数学教授安东尼·杜利说：“把数学运用到感情问题中通常不保险。在生活中，你得处理感情问题，而且需要更仔细地思量。“但如果你想算出最佳结婚时间，这个公式可为你提供的一个参考。”
- 该公式基于一个被称为“最优停止”的统计学技术，可计算出做某事的最佳时间。
- 杜利教授的同事布鲁斯·布朗教授说，该公式是一个“合理的方法”，很多年轻男士都在不知不觉中遵循着这一公式，包括他本人。
- 杜利教授在研究报告中以第三人称写道：“至于作者，他可以告诉你，回过头去看看，再做番计算，他的确遵循了这一婚姻公式——尽管是偶然的巧合，而且应用得很完美。”
- 尽管这一公式是为男士设计的，但对女士同样适用，比如那些正在犹豫是否接受求婚的女性。该公式还可以帮助紧张不安的男士计算出避免求婚的时间。
- 不过发现自己已经错过最佳求婚年龄的人也不要恐慌——就直接向下一个可能结为连理的对象求婚吧。
- 悉尼新南威尔士大学的杜利教授说：“概率不是婚姻最为浪漫的基础，但尽管这一公式不适用于每一个人，它确实适用于很多夫妻，不论是巧合还是精心设计的。”



# 瑞士发表 “夫妻感情走势图”

本报综合报道 爱的归宿是婚姻，夫妻相处则是一门大学问，日前，瑞士研究员通过调研数据编制出了一份“夫妇感情走势图”。

瑞士苏黎世大学对1.5万名德国人进行了长达16年的跟踪调查。调查显示，新婚且没有孩子的二人世界是婚姻生活中最快乐的时光。如果两人合适，这段“蜜月期”将持续一年。但那些注定离婚的怨偶在结婚头一年愉快感就会随时间骤降。

蜜月一过，即使有望白头到老的夫妻也会出现两年左右的“降温期”。而往后婚姻的愉快感则会“一

一份感情，  
一段婚姻，  
能否天长地久？

美国科学家称：

用数学方法可改变婚姻走向

作者：周欣宇 发表时间：2004-2-27 摘自：新华网 原载：中国青年报

日前，来自美国的婚姻研究者和应用数学家宣布，他们创造了一个数学模型，可以用来准确预测哪些夫妻将不能白头偕老。他们还表示，这套模型可以帮助夫妻克服那些可能使他们走上离婚之路的行为。据公式的设计者称，这套公式的准确率高达94%。

“当牛顿把数学方法引入科学，物理学才真正起飞”，华盛顿大学心理学教授和人际关系研究所主任约翰·戈特曼说，“而心理学研究中，数学的方法往往被忽视。”

华盛顿大学应用数学系教授詹姆斯·穆雷说，我们做的是提取婚姻中的关键因素到模型中，以使它具有解释性和预测性，“虽然我们使用的数学方法非常普通，但是模型出奇地准确”。截至目前，这套数学模型已经对700多对夫妇的婚姻持久性进行了检验。这些夫妇最初登记结婚时，研究人员利用这套数学模型判断他们今后是否会离婚，判断正确率高达94%。

这个模型使用的数据来自戈特曼教授在他实验室里拍摄的数百位夫妻谈话的录影带。生理学上的数据，例如谈话停顿的时间，也被收集起来进行分析。

# 数学公式测婚姻 据称准确率高达94%

据新华社电 一份感情，一段婚姻，能否天长地久？美国科学家最近别出心裁地设计出一套数学公式来对此进行推测。公式的设计者说，这套公式的准确率高达94%。

来自华盛顿大学的科学家在美国科学促进会年会上公布了这一研究成果。该校应用数学系教授詹姆斯·穆雷说，这套公式是数学与心理学结合的产物。他说，当一对夫妇接受测试时，需要就某个话题如性爱、金钱等进行15分钟的对话。研究人员根据对话的细节进行打分，比如说，当一方将另一方逗乐时，可以得正2分，而一方表示出对另一方的轻蔑时，将得到负4分。研究人员将每个细节的分数输入表格中，然后“像计算道琼斯工业指数一样”算出婚姻指数，推测婚姻的持久性。

《楚天都市报》2004年2月16日

# 补充实例： 美国《物理评论通讯》

## 剑桥大学： 数学定律预测眼角皱纹

人的皮肤与苹果皮之间有什么相同的地方？答案是：它们都会产生褶皱或者说皱纹。两位英国研究人员新总结出一条数学定律，能定量地对人眼角鱼尾纹等的尺寸、形状和结构进行预测。

在最新一期美国《物理评论通讯》杂志上，剑桥大学应用数学和理论物理系的塞尔达和马哈德万认为，皮肤、果皮等很多不同薄膜表面的皱纹，实际上有共同的产生机制。具体说，它们都是处于表面的薄膜与它所附着的更厚材料之间“妥协”的产物。

以苹果为例，当苹果果肉因为水分减少而出现收缩、直径变小时，它上面的果皮在保持面积不变的条件下就会起皱。皱纹的形状由果皮和果肉共同决定，其最终结构和形状取决于果皮与果肉因水分减少而收缩的程度。

塞尔达和马哈德万利用几何学和物理学工具，对上述机制进行分析后总结出了一条简单的定律。在这个定律中，表面皱纹的长度和它所附着的“基底”材料刚度的四次方根成反比。就人而言，皮肤皱纹的长度，是与皮肤下面组织的刚度相关的。

一些专家认为，定量地对人体皮肤皱纹进行比较详细的预测，将有助于寻找消除鱼尾纹的新方法。比如说，在产生鱼尾纹的区域附近选择特定的作用点，通过药物等使皮肤绷紧，也许可以“拉平”鱼尾纹。

类似的研究成果，据认为也可应用于制造更为平整的纸张，以及质量更高的玻璃、金属甚至硅芯片薄膜。（科技日报）





# 你会笑吗？英心理学家破解微笑的“完美公式”

2011-11-21 新华网

- 淡笑显矜持，皮笑肉不笑显得虚假，真诚美丽的笑脸则让人情不自禁回以一笑。
- 如果“深度”微笑，多会取得积极反馈；“淡淡”笑意多得不到回应。
- **相关因素**：包括**脸型**、**牙齿健康**、**微笑技巧和自信心**等。
- 曼彻斯特都市大学心理学系的戴维·霍姆斯博士发现，真诚的微笑需要动用脸部尤其是眼睛周围多块肌肉，这有助加强感染力。美丽笑容与脸型有较大关系，高颧骨、尖下巴更能突显笑容，显现笑意。
- **微笑公式**： $A \times ((S + T + F) - (W + L))$ 。
- A 指弓形，也就是微笑时嘴角上翘呈弓形，弧度一直延伸至眼角。
- S 指**充满自信的微笑**，这种笑容更明显，让人感觉这是一个发自内心的微笑。
- T 是**牙齿状况**，牙齿整齐，大小呈黄金分割比例，即从中间到两边依次减小，门牙两边的牙齿是门牙尺寸的 61.8%，下一颗紧挨的牙齿又是前一颗的 61.8%；露出的牙齿不能有残缺。
- F 为**脸型**，高颧骨、大眼睛和尖下巴让笑容更美。
- W 是**牙齿颜色**。
- L 是**嘴唇的状况和颜色**，最好是健康的粉色。
- **每一个元素最高为 5 分。总分最高为 65 分。**
- **完美微笑最重要的元素是弓形**，这可能在一定程度上受限于先天条件，“可以用其他元素弥补，**多呵护牙齿和嘴唇健康，微笑时充满自信**”。





1. ↵



①把手举到脸前：↵



②双手按箭头方向做“拉”的动作，一边想象笑的形象，一边使嘴笑起来。↵

2. ↵

## 如何微笑？



①把手指放在嘴角并向脸的上方轻轻上提：↵



②一边上提，一边使嘴充满笑意。↵





补充实例：

## 5、《中国人口科学》 2002年第3期

y: 夫妻争吵频率

x1: 结婚年数

x2: 受教育年数

x3: 个人收入

七年之痒 三年之痛

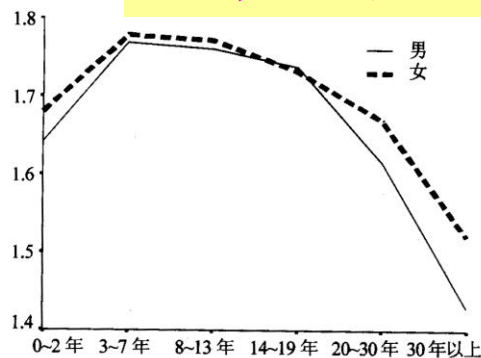


图1 夫妻各自述说的争吵频率

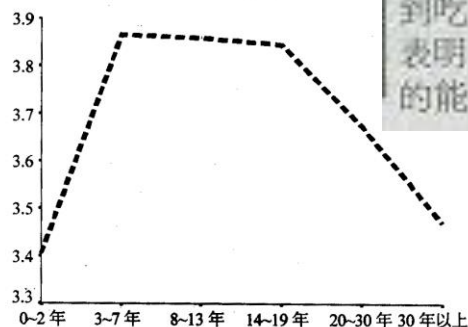


图2 双方述说合成后的争吵频率

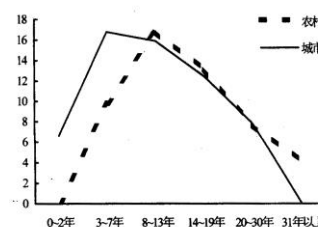


图3 控制其他变量后冲突频率的城乡差异

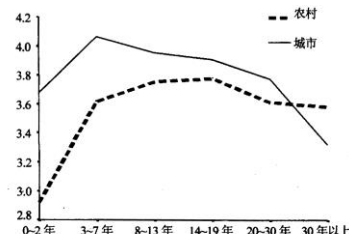


图4 控制其他变量前冲突频率的城乡差异

## 美国密歇根大学 社会调查研究所

研究  
表明

## 争吵越多,婚姻越牢固

清吵架的真相,从而更好地维持婚姻关系。

有800多名美国人接受了这项分别在1992年和2005年进行的调查。受调查者分20-39岁、40-59岁以及60岁以上3个不同年龄组。结果显示,各年龄段的人对配偶的负面感觉都会随着时间而增加。专家也对这一发现感到吃惊,因为大量的老年学研究

配偶则无法这么简单,离婚并不是件容易的事。共同生活在一起,彼此很难回避,负面感觉自然越积越多。

专家同时指出,应正确看待夫妻争吵。吵架,互不容忍是正常的家庭生活不可分割的组成部分,但通常不会导致关系的破裂。因此,只要不压抑自己的情绪,并在争吵后去理解和包容对方,婚姻关系就会愈加牢固。

刘青云、刘嘉玲主演都市爱情影片《心痒痒》(港名《七年之痒》)



常言道：贫贱夫妻百事哀

结婚 3—13 年争吵多  
个人收入少则争吵多

表 5 城乡夫妻冲突频率的多元回归分析结果 %

	城 市		农 村	
	Beta	Sig.	Beta	Sig.
结婚年数:0~2 年	0.066	0.001	0.000	0.000
3~7 年	0.168	0.000	0.098	0.006
8~13 年	0.159	0.000	0.161	0.000
14~19 年	0.124	0.000	0.131	0.000
20~30 年	0.078	0.001	0.076	0.074
31 年及以上	0.000	0.000	0.041	0.263
受教育年数	0.007	0.668	-0.006	0.775
个人收入	-0.005	0.768	-0.034	0.055
夫妻相对资源差(复合)	-0.043	0.008	-0.019	0.275
婚前性关系(1 为有)	0.059	0.000	0.031	0.089
婚前基础(复合)	0.022	0.225	-0.086	0.000
家庭类型(1 为核心小家庭)	0.001	0.942	-0.014	0.436
双方同质性(复合)	-0.197	0.000	-0.219	0.000
妻子承担更多家务	-0.011	0.480	0.021	0.225
家务劳动公平感	-0.047	0.004	-0.034	0.060
本人经济支配不自由	-0.104	0.000	-0.067	0.000
配偶家庭角色较合格	-0.160	0.000	-0.113	0.000
配偶对自己较信任	-0.167	0.000	-0.037	0.062
双方各不相让	0.034	0.029	0.067	0.000
烦恼、压力	0.123	0.000	0.169	0.000
R <sup>2</sup>	28.2		20.9	
F	61.873*		37.583*	

\* P<0.001

表 4 夫妻争吵频率的回归分析结果(全体样本) %

	模型 1		模型 2	
	Beta	Sig.	Beta	Sig.
结婚年数:0~2 年	0.028	0.043	0.017	0.217
3~7 年	0.122	0.000	0.104	0.000
8~13 年	0.140	0.000	0.123	0.000
14~19 年	0.107	0.000	0.094	0.000
20~30 年	0.060	0.000	0.045	0.008
31 年及以上	0.000	0.000	0.000	0.000
地区(1 为城市)	0.105	0.000		
受教育年数			0.069	0.000
个人收入	-0.021	0.082	-0.025	0.037
夫妻相对资源差(复合)	-0.039	0.001	-0.047	0.000
婚前性关系(1 为有)	0.052	0.000	0.051	0.000
婚姻基础(复合)	-0.053	0.000	-0.044	0.002
家庭类型(1 为核心小家庭)	-0.011	0.342	-0.010	0.631
双方同质性(复合)	-0.198	0.000	-0.202	0.000
妻子承担更多家务	0.002	0.865	-0.006	0.631
家务劳动公平感	-0.038	0.002	-0.038	0.002
本人经济支配不自由	-0.089	0.000	-0.092	0.000
配偶家庭角色较合格	-0.135	0.000	-0.133	0.000
配偶对自己较信任	-0.115	0.000	-0.110	0.000
双方各不相让	0.040	0.001	0.048	0.000
烦恼、压力	0.143	0.000	0.144	0.000
R <sup>2</sup>	24.1		23.6	
F	95.699*		92.884*	

\* P<0.001

# 挪威调查发现丈夫干家务越多离婚率越高

<http://msn.huanqiu.com> 2012-09-28

- 英国《每日电讯报》9月27日报道，挪威研究人员近日进行的一项调查结果显示，很多年轻夫妻都协议分工干家务，但是这类夫妻比妻子将家务全包的夫妻离婚率高得多。
- 报道称，这项调查结果看似在打性别平等的嘴巴。调查结果显示分工干家务的夫妻比妻子全包家务的夫妻的离婚率高出约50%。
- 挪威研究人员发现在生活中达成的公平对双方的亲密关系有害无益。家庭生活中的公平和生活质量几乎没有关系，甚至成反比。
- 研究人员托马斯·汉森（Thomas Hansen）称：“人们可能认为对于一些所谓男女不公平的家庭，他们的离婚率更高，但是结果截然相反。而且数据清晰显示，男性在家干的家务越多，双方离婚率越高。”

# 长篇小说：《同居时代》

## 校园情景剧：《大学生同居的事儿》

- 同居时间越长**结婚机率越低** 2011年12月21日，[人民网—人民电视](#)
  - 美国调查结果，90%的同居关系结局是分手。同居时间越长，越不容易结婚。  
2009-5-7 10:12:48 花花网
  - 美国另一项对2150名男性的调查也表明，有过同居经历的人，只有1/3后来和该女性结婚。
- 
- 同居时间越长**离婚率越高**
  - 多项研究证明，同居时间越长，对感情和当事人的**负面影响越大**。美国新泽西州罗杰斯大学的教授，大卫·波彭诺和巴巴拉·D·怀特赫德，研究同居关系历时10多年。结果显示，**经同居结成的婚姻，比未经同居结成的，离婚率高出46%**。同居时间越长，双方将更追求独立自主，更不愿受婚姻的约束，**永不结婚的可能性也越大**，同居关系的破裂率也比婚姻关系的破裂率更高。

# 沿用20多年的中国人帕金森患病率作出重大修正

## 2005-06-03

- 经过长达6年的跟踪调查，我国医学专家日前在国际权威杂志《柳叶刀》指出，65岁以上中国人帕金森患病率男性为1.7%、女性为1.6%。
- 由北京协和医院流行病学教研室主任张振馨教授牵头，北京、上海、西安三地专家参与的帕金森病研究组自1996年起，在137个城乡居民区进行了入户调查和定期随访，样本人群接近3万人。据张振馨解释，尽管新报告发病率为既往数据的13倍，但并不能简单理解为我国帕金森病患者成倍增加，因为过去的帕金森病调查存在方法不科学、参与调查的人员不够专业等问题，既往数据并不准确。
- 由于世界各国的老龄化水平存在差异，帕金森发病率很难简单地进行横向比较。按照国际通用做法，专家们以美国2000年人口数为标准，换算出我国帕金森发病率为2.1%，这个数据与国际水平相比并无差异。而在此之前，医学界普遍认为，黄种人患帕金森病的概率要低于白种人。

补充实例:

6、水稻研究 ---- 本科生毕业论文

水稻苗期根系性状的数量遗传分析

陈明明 董 伟 胡中立 章志宏

(武汉大学生命科学学院 武汉 430072)

QUANTITATIVE GENETIC ANALYSIS OF SEEDLING  
ROOT CHARACTERS IN RICE

Chen Mingming Dong Wei Hu Zhongli Zhang Zhibong

表 2 新发根率(Y)与其它性状的多元回归分析

Table 2 Multiple regression analysis between percentage of newly-developed  
root dry weight (Y) and other characters

自变量 X Independent variable X	Y <sub>i</sub>	Y <sub>r</sub> (分析中包括 X <sub>i</sub> ) (Including X <sub>i</sub> )		Y <sub>r</sub> (分析中不包括 X <sub>i</sub> ) (Not including X <sub>i</sub> )	
		一般回归 General regression	逐步回归 Stepwise regression	一般回归 General regression	逐步回归 Stepwise regression
常量 Invariant	0.033 4	0.024 8	0.043 7	0.018 7	0.013 3
X <sub>1</sub>	3.584 7(7.8403 <sup>*</sup> )	0.558 1(5.73 <sup>*</sup> )	0.552 1(989.85 <sup>*</sup> )	—	—
X <sub>2</sub>	-0.208 7(2.8375 <sup>*</sup> )	-0.012 7(2.64 <sup>*</sup> )	-0.013 3(258.65 <sup>*</sup> )	-0.007 1(2.06 <sup>*</sup> )	-0.008 4(15.88 <sup>*</sup> )
X <sub>3</sub>		0.000 0(0.06 <sup>*</sup> )		0.001 2(7.02 <sup>*</sup> )	0.001 2(75.66 <sup>*</sup> )
X <sub>4</sub>		0.000 0(0.08 <sup>*</sup> )		-0.000 1(0.37 <sup>*</sup> )	
X <sub>5</sub>		-0.001 0(0.32 <sup>*</sup> )		-0.001 8(0.80 <sup>*</sup> )	
F	933.33	194.37	190.11	18.73	18.42
R <sup>2</sup>	0.962 0	0.931 0		0.506 8	

## 9-2 多元相关分析

### 一、多元相关

多元相关系数

决定系数

调整决定系数

总相关

$m$  个自变量和依变量的总相关

### 二、偏相关

偏相关系数

简单相关系数

变量间的相关

# 一、多元相关

## 1、多元相关或复相关：

为  $m$  个自变量和依变量的总相关

多元相关系数：第四版P224(12.32)

$$R = R_{y/x_1, x_2, \dots, x_m} = \sqrt{[U_{y/x_1, x_2, \dots, x_m} / SS_y]}$$

## 2、多元相关系数的假设检验：

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

$$F = df_2 \times R^2 / [df_1 \times (1 - R^2)]$$

$$df_1 = m$$

$$df_2 = n - m - 1$$

方差分析

例：第四版P225

### 3、决定系数 $R^2$ :

多元回归平方和 占  $y$  的总变异平方和 的比率

### 4、调整决定系数 :

消除变量个数  $m$  的影响

$$\text{adj-}R^2 = 1 - (1 - R^2) (n-1) / (n-m-1)$$

例：大学生体重

大学生身高

回归统计	
Mult R	0.8507
R Square	0.72369
Adj R Squ	0.66281
标准误差	6.22611
观测值	73

回归统计	
Mult R	0.85498
R Square	0.731
Adj R Squ	0.6772
标准误差	4.39158
观测值	73



## 二、偏相关

排除其它变量影响下的两变量间的相关分析，称偏相关分析。

在其它变量都保持一定时，表示指定的两个变量之间相关密切程度的量值称为偏相关系数。记成：

$r_{ij.k}$   
 $r_{ij.kl}$   
 $r_{12.34\dots m}$

$r_{ij}$

计算方法：P226

- 1、计算简单相关系数并构成相关矩阵  $R$
- 2、求其逆矩阵  $R^{-1}$
- 3、计算偏相关系数：  $-c_{ij}/\sqrt{(c_{ii} \times c_{jj})}$

偏相关与简单相关的区别：P228

偏相关系数与简单相关系数不仅数据有异，而且符号也可能不同。表12—6

实例：  
水稻产量 Y 与产量构成性状  $X_i$  的多元回归与相关分析

$$Y = -69.3719 - 0.0342 X_1 + 0.0177 X_2 + 2.0767 X_3 + 0.2518 X_4 + 0.1254 X_5 + 0.2919 X_6 + 1.3251 X_7 + 0.8671 X_8$$

	RIL		ZSBC		MHBC	
	简单相关	偏相关	简单相关	偏相关	简单相关	偏相关
HD	-0.10712	-0.13555	0.361809	0.040498	0.012528	0.021242
PH	-0.00824	0.088338	0.129485	0.052516	-0.14532	0.027784
TP	0.366243	0.864656	0.417378	0.908205	0.631683	0.927276
PL	0.048484	0.173384	0.28453	0.03949	0.296942	0.151226
GPP	0.071832	0.353326	0.446495	0.450516	0.322444	0.466447
SS	0.498109	0.861387	0.328039	0.818881	0.36475	0.880578
GD	0.06921	0.171506	0.340058	0.030931	0.170961	0.193578
KGW	0.084975	0.782014	0.154815	0.757131	0.056976	0.725689

## 9-3 多项式回归分析

双变量资料的三种回归：

直线回归

$$y = a + bx$$

可化为直线的非线性回归

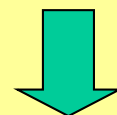
多项式回归

$$y = f(x)$$

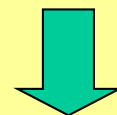
直线  
简单曲线  
波浪曲线

- 一、多项式回归的数学模型
- 二、多项式回归方程的建立
- 三、多项式回归方程的假设检验
- 四、相关指数

一元



多元



一元

# 一、多项式回归的数学模型

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \varepsilon_i$$

式中： $\beta_1, \beta_2, \dots, \beta_k$ 为  $x$  的一次项、二次项、...， $k$  次项的  
总体回归系数；

$\varepsilon_i$ 为随机误差，服从正态分布 $N(0, \sigma^2_{y/x, x_2, \dots, x_k})$

由样本估计时

$$y_i = a + b_1 x_i + b_2 x_i^2 + \dots + b_k x_i^k + e_i$$

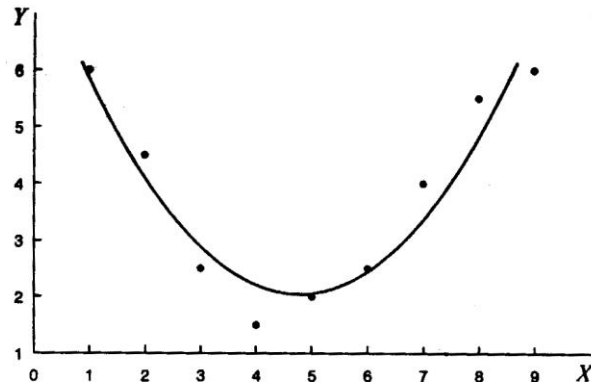


图 11-1 例 11.2 的散点图

波浪曲线

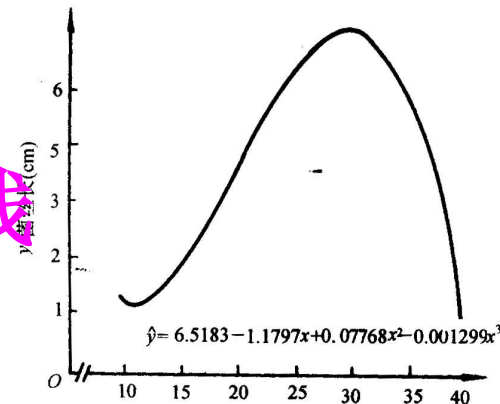


图 13.1 黑木耳菌丝长度与温度的关系

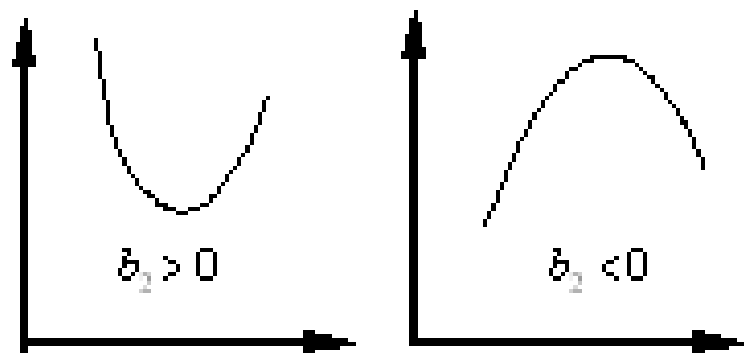


图 11.6 方程  $\hat{y}_2 = a + b_1x + b_2x^2$  的图象

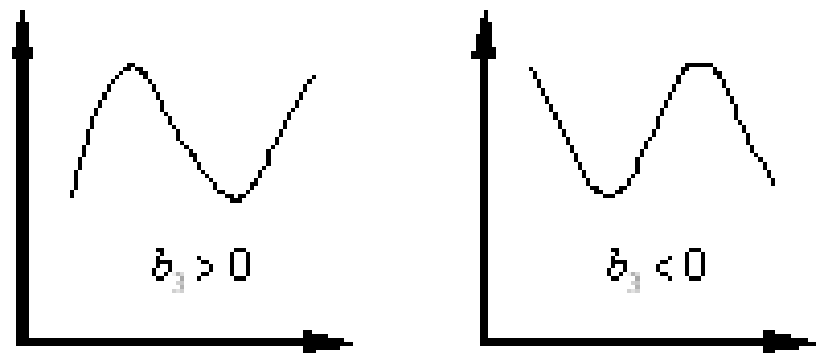


图 11.7 方程  $\hat{y}_3 = a + b_1x + b_2x^2 + b_3x^3$  的图象

波浪曲线

## 二、多项式回归方程的建立

根据数据资料的二维散点所表现的曲线趋势

谷+峰+1 = 多项式回归方程的次数  $k$

令： $x = x_1$   
 $x^2 = x_2$   
.....  
 $x^k = x_k$

上标改成下标  
下标改成上标

然后就进行多元线性回归分析：  
多项式回归方程的建立，采用最小二乘估计。  
 $Q = \sum (y - \hat{y})^2 = \text{最小值}$

再后就直接改写成：

$$\hat{y} = 6.5183 - 1.1797x + 0.0777x^2 + 0.0013x^3$$

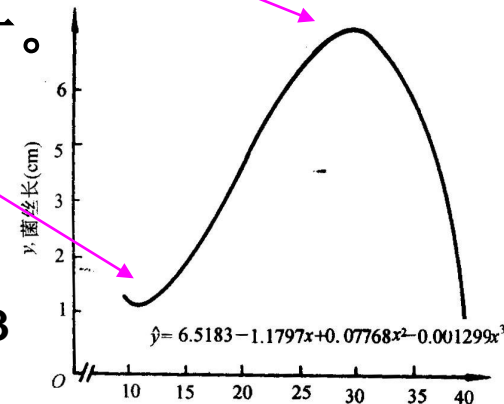


图 13.1 黑木耳菌丝长度与温度的关系

### 三、多项式回归方程的假设检验

在多项式回归中，经 **F 检验** 略去不显著项，即为建立的多项式回归方程。

教材第四版P252 表14-3

表 13-3 黑木耳菌丝长度依温度多项式回归的假设检验

变异来源	$df$	$SS$	$s^2$	$F$	$F_{0.05}$	$F_{0.01}$
回 归	3	27.3540	9.1180	61.82 **	9.28	29.46
1 次分量	1	1.1807	1.1807	8.00	10.13	34.12
2 次分量	1	20.4783	20.4783	138.84 **	10.13	34.12
3 次分量	1	5.6950	5.6950	38.61 **	10.13	34.12
离回归	3	0.4426	0.1475			
总变异	6	27.7966				

## 四、相关指数

**相关指数**：**k**次多项式的**回归平方和**占**总变异平方和**的比率的平方根值。

$$R_k = \sqrt{[U_k/SSy]}$$

**决定系数**：在 **y** 的总变异平方和中，可由 **x** 的 **k** 次多项式说明的部分所占的比率。

$$R_k^2 = U_k/SSy$$

例：第四版**P253**例**14.3**

$$R_k = 0.9920$$

$$R_k^2 = 0.9841$$

黑木耳生长的总变异**98.41%**可由温度的三次多项式说明



例1: 第四版P251 黑木耳生长与温度的关系

$$\hat{y} = 6.5183 - 1.1797x + 0.0777x^2 + 0.0013x^3$$

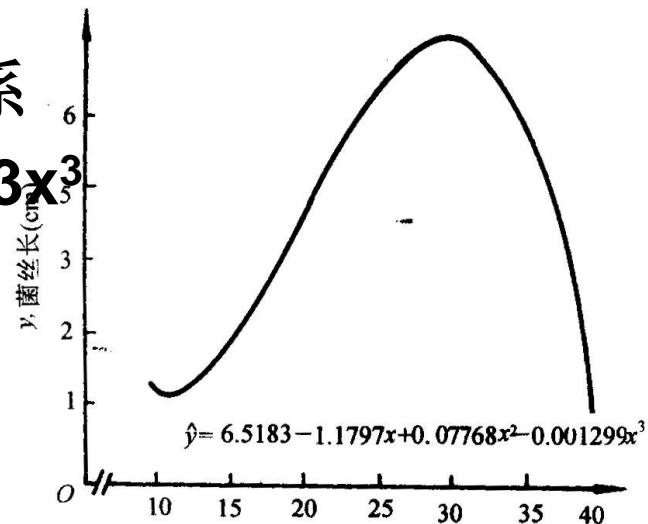


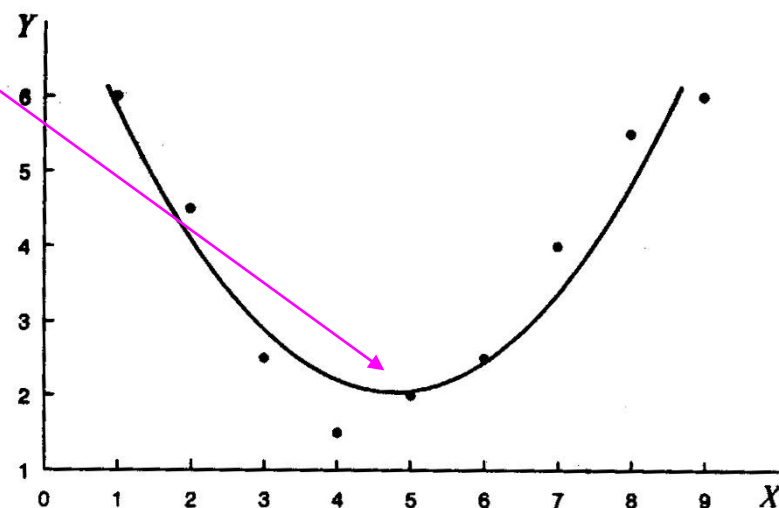
图 13.1 黑木耳菌丝长度与温度的关系

例2: 杜荣骞: 生物统计学

人被射线照射后白细胞含量  $y$  与受照后周数  $x$  的关系

$$\hat{y} = a + b_1x + b_2x^2$$

$$= 8.0239 - 2.4857x + 0.2619x^2$$



例3: 《中国社会科学》 2000年

y : 我赞成超前消费  
x : 消费阶层(富裕程度)

$y = a + b_1x$   
 $y = a + b_1x + b_2x^2$   
 $y = a + b_1x + b_2x^2 + b_3x^3$

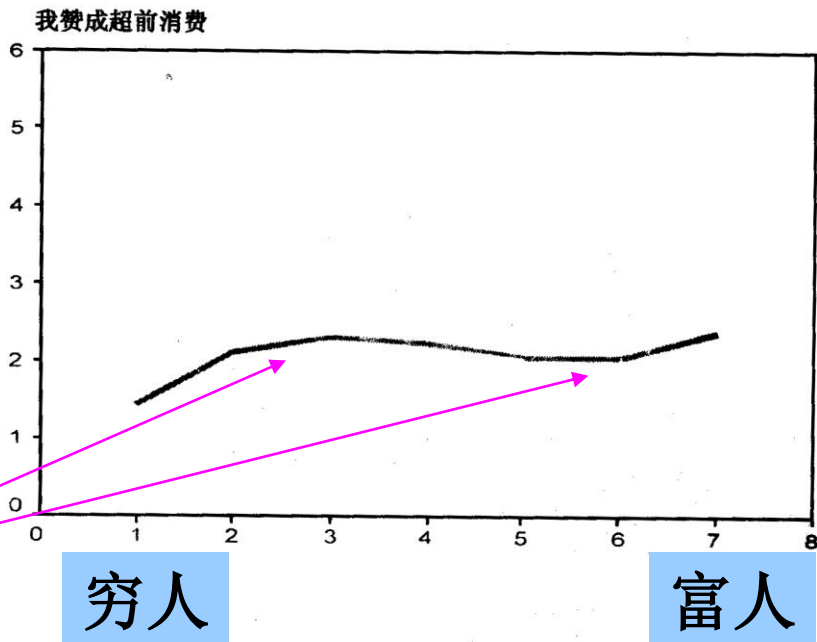


图 1. 消费态度随消费阶层的提高而改变的曲线估计①

表 5. 以“消费阶层”为自变量对“我赞成超前消费”的估计

因变量	预测方法	R <sup>2</sup>	自由度	F 值	显著性	b1 (x)	b2 (x <sup>2</sup> )	b3 (x <sup>3</sup> )
我赞成超前消费	直线	.694	1 246	2 829.14	.000	.4694		
	二次	.779	1 245	2 189.96	.000	1.0675	-.1171	
	三次	.798	1 244	1 640.07	.000	1.8293	-.4597	.0352

资料来源: 1999 年重庆调查

决定系数越大越好

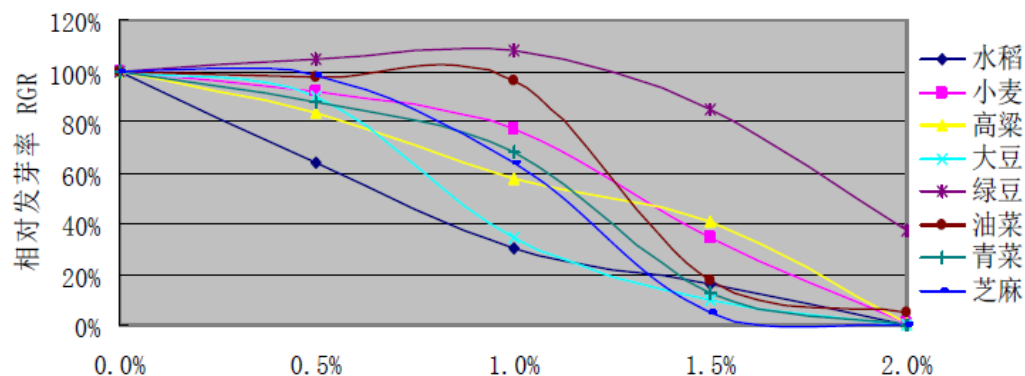
图形  
1  
2  
3

# 例4：主要农作物芽期耐盐性比较研究

*Hans Journal of Agricultural Sciences* 农业科学, 2012, 2, 59-65

- 表1. 各参试作物和品种在四种盐浓度下的相对发芽率
- 表2. 不同盐浓度对各作物相对发芽率影响
- 表3. 不同作物的相对发芽率曲线方程

作物Crop	盐浓度NaCl concentration			
	0.5%	1.0%	1.5%	2.0%
绿豆Mung bean	104.87 a	107.93 a	84.72 a	36.91 a
油菜Rape oil seed	97.80 ab	96.22 ab	17.51 bc	5.00 b
小麦Wheat	92.12 ab	76.98 bc	34.43 bc	0.70 b
青菜Green vegetables	87.50 b	67.73 bc	12.45 bc	0.00 b
芝麻Sesame	98.20 ab	63.77 c	5.01 c	0.00 b
高粱Sorghum	83.37 b	57.35 cd	40.89 b	0.72 b
大豆Soybean	89.55 ab	34.62 d	9.92 c	0.00 b
水稻Rice	64.15 c	30.26 d	16.06 bc	0.00 b



作物 Crop	曲线方程 Equation	决定系数 Determination coefficient	半致死盐浓度(%)Semi-lethal NaCl concentration
水稻 Rice	$y = -4961.8x + 91.712$	0.9573	0.84
小麦 Wheat	$y = -5125.8x + 112.1$	0.9269	1.21
高粱 Sorghum	$y = -4820.8x + 104.67$	0.9732	1.13
大豆 Soybean	$y = -5592.6x + 102.74$	0.9357	0.94
绿豆 Mung bean	$y = -376,086x^2 + 4595.1x + 97.348$	0.9821	1.89
油菜 Rape oil seed	$y = -107,200x^2 - 2676.8x + 99.314$	0.9900	1.23
青菜 Green vegetables	$y = -5501x + 108.55$	0.9377	1.06
芝麻 Sesame	$y = -5863.8x + 112.03$	0.9068	1.06

y 表示相对发芽率；x 表示盐浓度(%)。

## 例5：多项式拟合线的错用示例

- 李春喜译《Mathematics and Statistics for Life Scientists》  
《生命科学中的数学与统计学》 P42

多项式拟合线的错用示例

有时有一些杂乱无章的数据,如图 E2. 4 所示,不能很好地拟合一条简单的曲线或直线,用多项式进行拟合具有较大的诱惑力,这种拟合具有较好的适合度,似乎应该得到支持。要避免这种诱惑。

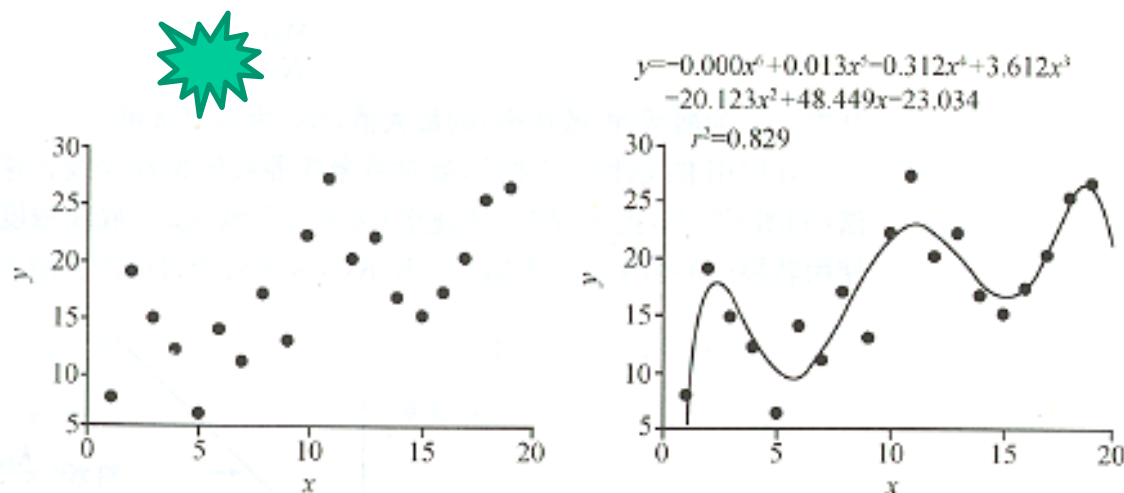


图 E2. 4 杂乱生物学数据模拟举例

对这些数据,用六次多项式显示了相当好的适合度。的确用来解释这条曲线变异比例的  $r^2$  达到了 83%。但是,这种模型没有生物学意义,所拟合的曲线也没有预测价值。面对这样的数据,线性回归是最好的方法。

# 例6：联合应用

G. R. Norman: 《biostatistics:The Bare Essentials》

女性更年期

- 男人情感综合症: **PMS**综合症 (**pathetic male syndrome**)
- **PMS**综合症在急性期常被称为: 中年危机、**男性更年期**
- 行为表现: 皮带逐步的上下移动

中年男士购买艳丽的帽子遮盖少头发

- **痛苦指数 PQ**: 皮带上没用过的槽口数  
郁闷叹息的次数







张敬和作品 www.ayj888.cn



从有了到十个月的感觉~~



表 16-1 16 个 PMS 患者数据

患者	年龄	腰围	治疗	PQ
1	24	46	睾丸激素	12
2	26	36	睾丸激素	14
3	27	40	其他	27
4	88	44	睾丸激素	35
5	32	36	睾丸激素	26
6	29	30	其他	21
7	70	42	睾丸激素	48
8	75	35	其他	51
9	37	42	其他	62
10	65	50	睾丸激素	64
11	72	45	睾丸激素	60
12	55	53	其他	77
13	45	48	其他	91
14	41	38	其他	84
15	63	43	其他	55
16	40	58	睾丸激素	74

赋值：虚拟变量

0 为对照  
1 为治疗

1	PQ值	治疗	腰围	年龄	年龄*年龄
2	12	1	46	24	576
3	14	1	36	26	676
4	27	0	40	27	729
5	35	1	44	88	7744
6	26	1	36	32	1024
7	21	0	30	29	841
8	48	1	42	70	4900
9	51	0	35	75	5625
10	62	0	42	37	1369
11	64	1	50	65	4225
12	60	1	45	72	5184
13	77	0	53	55	3025
14	91	0	48	45	2025
15	84	0	38	41	1681
16	55	0	43	63	3969
17	74	1	58	40	1600

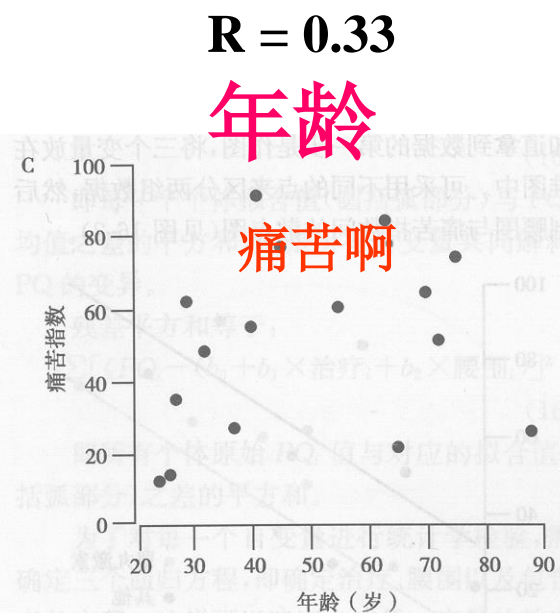
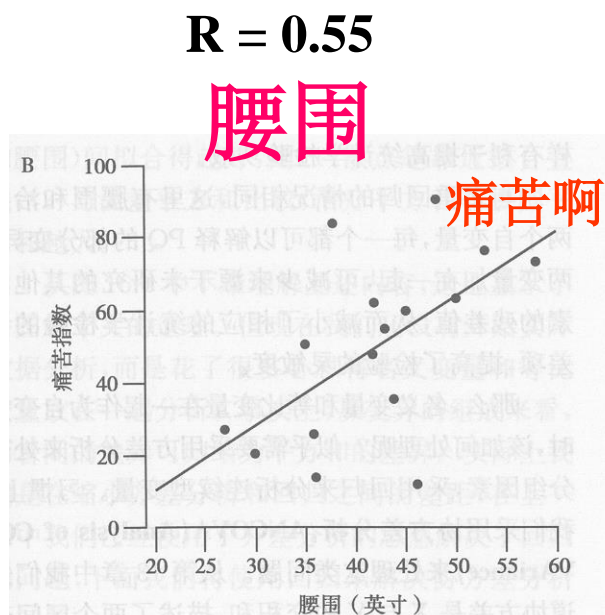
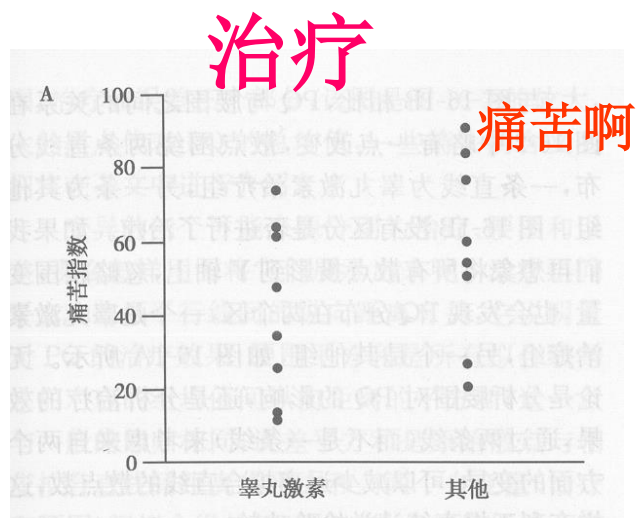


图 16-1 PQ 分别与治疗(A)、腰围(B)和年龄(C)间的关系

# 各种可能的回归分析（仅列 6 种）

$$PQ = b_0 + b_1 (\text{治疗})$$

$$PQ = b_0 + b_2 (\text{腰围})$$

$$PQ = b_0 + b_1 (\text{治疗}) + b_2 (\text{腰围})$$

$$PQ = b_0 + b_3 (\text{年龄})$$

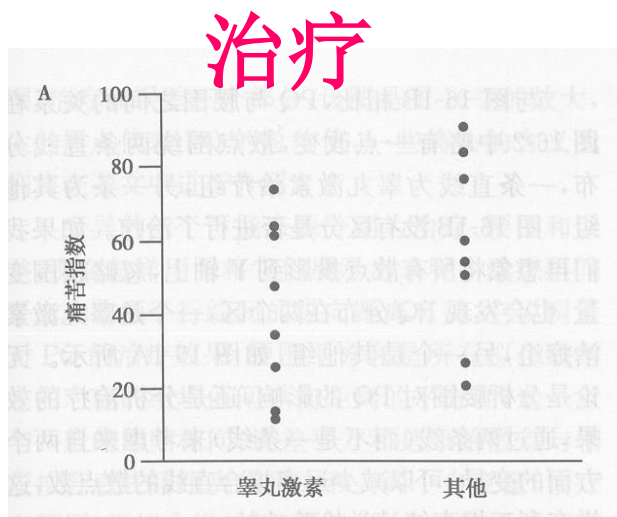
$$PQ = b_0 + b_3 (\text{年龄}) + b_4 (\text{年龄})^2$$

$$PQ = b_0 + b_1 (\text{治疗}) + b_2 (\text{腰围}) + b_3 (\text{年龄}) + b_4 (\text{年龄})^2$$

治疗: t 检验:  $t=1.38$   
 方差分析:  $F=1.90$   
 回归分析:  $PQ = b_0 + b_1 \times \text{治疗}$

	Coefficient	标准误差	t Stat	P-value
Intercept	58.5	8.65194	6.76149	9.1E-06
治疗	-16.875	12.2357	-1.3792	0.18948

回归统计	
Multiple R	0.34585
R Square	0.11961
Adjusted R Square	0.05673
标准误差	24.4714
观测值	16



t-检验: 双样本等方差假设		
	对照	治疗
平均	58.5	41.625
方差	646.857	550.839
观测值	8	8
合并方差	598.848	
假设平均	0	
df	14	
t Stat	1.37916	
P(T<=t)	0.09474	
t 单尾临	1.76131	
P(T<=t)	0.18948	
t 双尾临	2.14479	

$$1.38 * 1.38 = 1.904$$

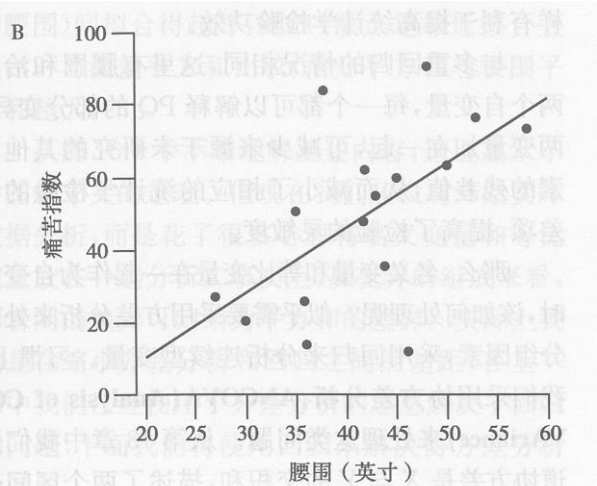
方差分析					
差异源	SS	df	MS	F	P-value
组间	1139.06	1	1139.06	1.90209	0.18948
组内	8383.88	14	598.848		
总计	9522.94	15			

# 腰围：回归分析

$PQ = b_0 + b_2 \times \text{腰围}$

方差分析:  $F=6.11$

腰围



SUMMARY OUTPUT						
回归统计						
Multiple R	0.55124					
R Square	0.30387					
Adjusted R Square	0.25414					
标准误差	21.7604					
观测值	16					
方差分析						
	df	SS	MS	F	Significance F	
回归分析	1	2893.72	2893.72	6.11115	0.02688	
残差	14	6629.22	473.515			
总计	15	9522.94				
Coefficients		标准误差	t Stat	P-value	lower 95%	upper 95%
Intercept	-32.533	33.8513	-0.961	0.35283	-105.14	40.0711
腰围	1.92642	0.77927	2.47207	0.02688	0.25505	3.59779



# 合并的回归分析

$$PQ = b_0 + b_1 \times \text{治疗} + b_2 \times \text{腰围}$$

对照:  $PQ = b_0 + b_1 \times 0 + b_2 \times \text{腰围}$

治疗:  $PQ = b_0 + b_1 \times 1 + b_2 \times \text{腰围}$

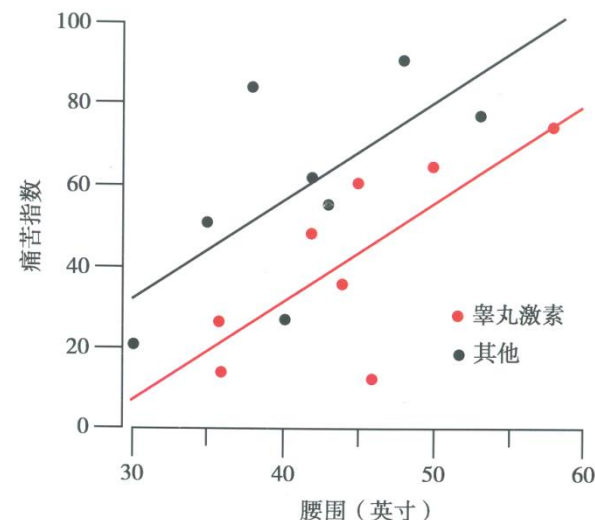
## 回归统计

Multiple R	0.74422
R Square	0.55387
Adjusted R Square	0.48523
标准误差	18.0778
观测值	16

$b_1$  是图中两条  
直线间的纵向距离

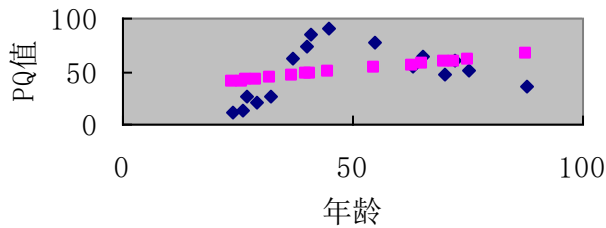
## 方差分析

	df	SS	MS	F	Significance F	
回归分析	2	5274.47	2637.23	8.06974	0.00527	
残差	13	4248.47	326.805			
总计	15	9522.94				
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-39.332	28.235	-1.393	0.18698	-100.33	21.6662
治疗	-25.201	9.33701	-2.6991	0.01823	-45.373	-5.0297
腰围	2.37889	0.66874	3.55725	0.00351	0.93416	3.82363



彩图 16-2 PQ 与治疗 and 腰围之间的关系

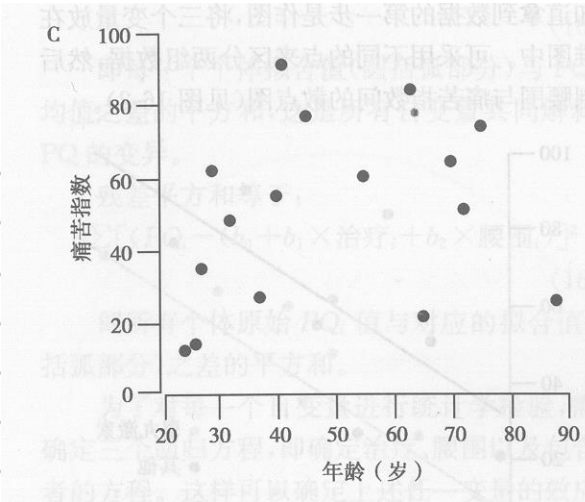
年龄



$PQ = b_0 + b_3 \text{ (年龄)}$

$PQ = b_0 + b_3 \text{ (年龄)} + b_4 \text{ (年龄)}^2$

SUMMARY OUTPUT						
回归统计						
Multiple R	0.84694					
R Square	0.71731					
Adjusted R Square	0.67382					
标准误差	14.3903					
观测值	16					
方差分析						
	df	SS	MS	F	Significance F	
回归分析	2	6830.91	3415.45	16.4935	0.00027	
残差	13	2692.03	207.079			
总计	15	9522.94				
Coefficients		标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-110.59	28.284	-3.91	0.00179	-171.69	-49.487
年龄	6.62277	1.18959	5.56726	9.1E-05	4.05281	9.19273
年龄*年龄	-0.0587	0.01111	-5.2867	0.00015	-0.0828	-0.0347



# 各种可能的回归分析及其决定系数

R <sup>2</sup>	adj-R <sup>2</sup>
----------------	--------------------

$PQ = b_0 + b_1 \text{ (治疗)}$

0.12    0.06

$PQ = b_0 + b_2 \text{ (腰围)}$

0.30    0.25

$PQ = b_0 + b_1 \text{ (治疗)} + b_2 \text{ (腰围)}$

0.55    0.49

$PQ = b_0 + b_3 \text{ (年龄)}$

0.11    0.05

$PQ = b_0 + b_3 \text{ (年龄)} + b_4 \text{ (年龄)}^2$

0.72    0.67

$PQ = b_0 + b_1 \text{ (治疗)} + b_2 \text{ (腰围)} + b_3 \text{ (年龄)} + b_4 \text{ (年龄)}^2$

0.81    0.75

回归统计	
Multiple R	0.90242
R Square	0.81437
Adjusted R Square	0.74686
标准误差	12.677
观测值	16

方差分析						
	df	SS	MS	F	Significance F	
回归分析	4	7755.17	1938.79	12.0642	0.00052	
残差	11	1767.77	160.706			
总计	15	9522.94				
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-112.61	28.0371	-4.0166	0.00203	-174.32	-50.905
治疗	-13.506	7.649	-1.7657	0.10514	-30.341	3.32916
腰围	1.24076	0.55717	2.22691	0.04778	0.01444	2.46708
年龄	4.70792	1.3213	3.5631	0.00445	1.79976	7.61608
年龄*年龄	-0.041	0.01233	-3.3298	0.00671	-0.0682	-0.0139

# 再次提醒

- 回归分析和相关分析研究的是变量之间的关系
- 这其中的假设检验（t 检验、F 检验-方差分析）仅仅只是为回归分析和相关分析服务的
- 注意：这里的方差分析与前面的检验多个平均数大小差异的方差分析：原理基本相同  
目的完全不同

# 作业

**第四版： P228**

**12.1**

**12.2**

**12.3**