

博學而不自滿 勤學而不自懈

# 第四章 回归分析

## 管理决策建模与可视化

彭世喆 数字经济系

[pengshizhe@csust.edu.cn](mailto:pengshizhe@csust.edu.cn)

## 第四章 回归分析



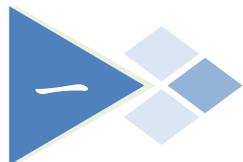
### 回归分析概述

回归分析的概念、  
回归预测的检验与  
步骤



### 回归预测模型

一元线性回归预测、  
多元线性回归预测、  
一元非线性回归预测

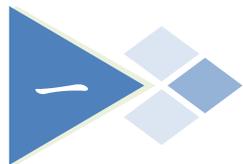


# 回归分析概述

## 回归分析的概念

回归分析：确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。

回归分析预测：通过一个变量或一些变量（自变量）的变化来解释另一个变量（因变量）的变化情况，从而由自变量的取值来预测因变量的可能值。



## 回归分析概述

### 回归模型的检验

判定系数  $> 0.9$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$Ra^2 = 1 - (1 - R^2) \frac{n-1}{n-k} \quad \begin{array}{l} \text{消除自变量} \\ \text{数目的影响} \end{array}$$

### t统计量

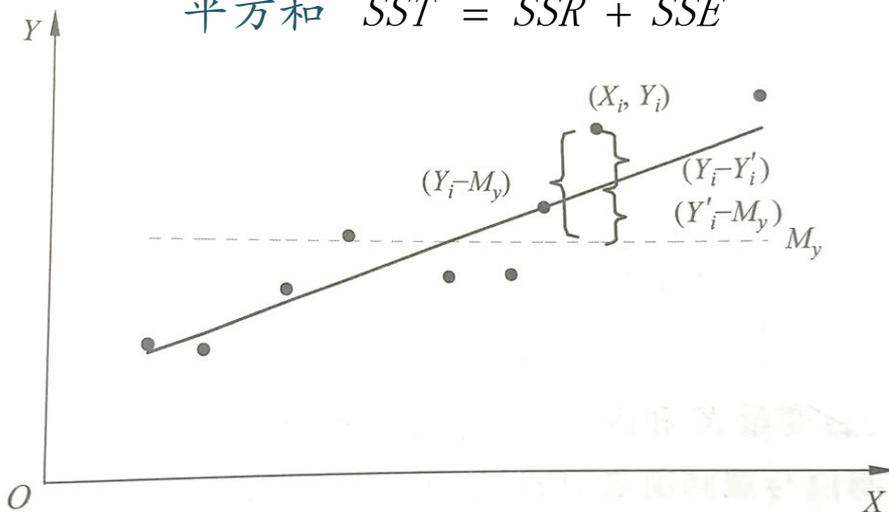
t统计量可以用来确定因变量和每个自变量之间是否存在显著的关系，即该自变量与因变量是否相关。

### F统计量

F-检验可以用来确定自变量的全体与因变量之间是否存在显著的关系，即回归方程的效果是否显著。

总离差=回归离差+残差

平方和  $SST = SSR + SSE$





## 回归预测的步骤

第一步

获取自变量和因变量的观测值

第二步

绘制XY散点图

第三步

写出带未知参数的回归预测方程

第四步

确定回归方程的最优参数值

第五步

判断回归方程的拟合优度

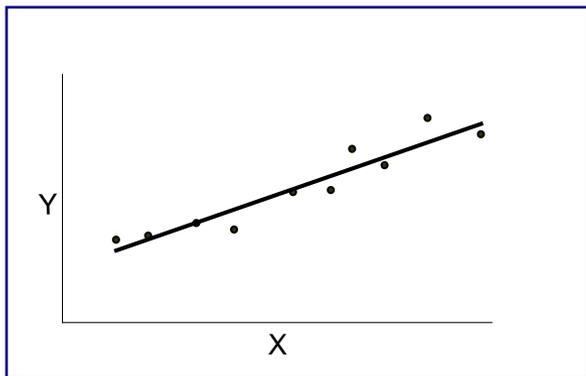
第六步

进行预测

## 一元线性回归分析

$$F_i = a + bX_i$$

原理：预测值与观测值之间均方误差最小



$$MSE = \frac{1}{n} \sum_{i=1}^n (F_i - Y_i)^2 = \frac{1}{n} \sum_{i=1}^n (a + bX_i - Y_i)^2 \quad \text{最小}$$

$$\Rightarrow \begin{cases} a = M_y - bM_x \\ b = \frac{\sum_{i=1}^n (X_i - M_x)(Y_i - M_y)}{\sum_{i=1}^n (X_i - M_x)^2} \end{cases}$$

例4-1

例4-2

确定拟合方程系数：

最小二乘法

回归方程的拟合优度：

判定系数  $R^2$

在Excel中可方便得到系数a和b：

添加趋势线并显示公式

SLOPE () 和 INTERCEPT () 函数、LINEST () 函数

规划求解

回归分析报告



## 例4-1 一元线性回归模型

【例4-1】“阿曼德匹萨”是一个制作和外卖意大利匹萨的餐饮连锁店，其主要客户群是在校大学生。为了研究各店铺销售额与店铺附近地区大学生人数之间的关系，随机抽取了十个分店的样本，得到的数据如下：

店铺编号	区内大学生数 (万人)	季度销售额 (万元)
1	0.2	5.8
2	0.6	10.5
3	0.8	8.8
4	0.8	11.8
5	1.2	11.7
6	1.6	13.7
7	2	15.7
8	2	16.9
9	2.2	14.9
10	2.6	20.2

建立回归模型并预测一个区内大学生人数为1.8万的店铺的季度销售额。

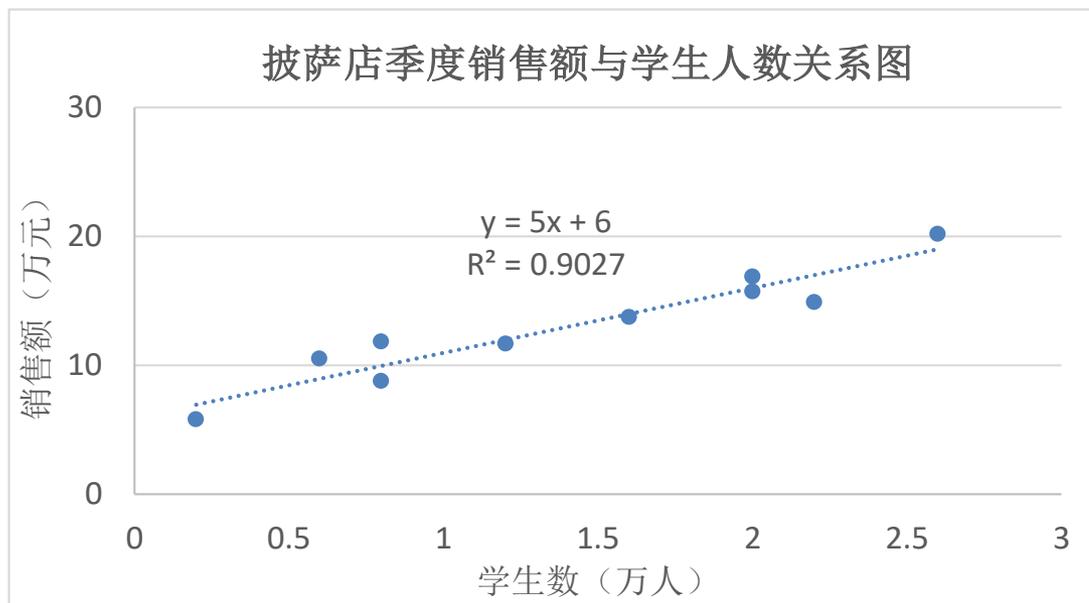


# 例4-1 一元线性回归模型

## 方法一 添加趋势线

第一步，绘制观测值散点图；  
第二步，添加一元线性趋势线，显示公式和R<sup>2</sup>值；  
第三步，根据一元线性趋势线公式计算预测值。

	A	B	C
1	店铺编号	区内大学生数(X)	季度销售额(Y)
2	1	0.2	5.8
3	2	0.6	10.5
4	3	0.8	8.8
5	4	0.8	11.8
6	5	1.2	11.7
7	6	1.6	13.7
8	7	2	15.7
9	8	2	16.9
10	9	2.2	14.9
11	10	2.6	20.2
12			
13		预测:	
14		大学生数	1.8
15		销售额	36.4





# 例4-1 一元线性回归模型

## 方法二 一元线性回归函数

	A	B	C	D	E	F
1	店铺编号	区内大学生数(X)	季度销售额(Y)	INTERCEPT/SLOPE函数		
2	1	0.2	5.8	a (截距)	6	
3	2	0.6	10.5	b (斜率)	5	
4	3	0.8	8.8	LINEST函数		
5	4	0.8	11.8	5		6
6	5	1.2	11.7	RSQ函数		
7	6	1.6	13.7	R平方	0.903	
8	7	2	15.7	预测:		
9	8	2	16.9	大学生数	1.8	
10	9	2.2	14.9	销售额	15.0	
11	10	2.6	20.2			
12						
13						

F9=RSQ (C2:C11, B2:B11)



# 例4-1 一元线性回归模型

## 方法三 规划求解工具

$$F_i = a + bX_i$$

	A	B	C	D	E	F	G
1	店铺编号	区内大学生数(X)	季度销售额(Y)	销售额预测值(F)			
2	1	0.2	5.8	7.0		a (截距)	6.0
3	2	0.6	10.5	9.0		b (斜率)	5.0
4	3	0.8	8.8	10.0			
5	4	0.8	11.8	10.0		MSE	1.53
6	5	1.2	11.7	12.0			
7	6	1.6	13.7	14.0		预测:	
8	7	2	15.7	16.0		大学生数	1.8
9	8	2	16.9	16.0		销售额	15.0
10	9	2.2	14.9	17.0			
11	10	2.6	20.2	19.0			

规划求解参数

设置目标:(I)

到:  最大值(M)  最小值(N)  目标值(V)

通过更改可变单元格:(B)

遵守约束:(L)

添加(A)

更改(C)

删除(D)

全部重置(R)

装入/保存(L)

使无约束变量为非负数(K)

选择求解方法:(E)

求解方法  
 为光滑非线性规划求解问题选择 GRG 非线性引擎。为线性规划求解问题选择单纯线性规划引擎，并为非光滑规划求解问题选择演化引擎。



# 例4-1 一元线性回归模型

## 方法四 回归分析报告

	A	B	C	D	E	F
1	店铺编号	区内大学生数 (X)	季度销售额 (Y)			
2	1	0.2	5.8	由回归分析报告得:		
3	2	0.6	10.5	a (截距)	6.0	
4	3	0.8	8.8	b (斜率)	5.0	
5	4	0.8	11.8	R平方	0.903	
6	5	1.2	11.7			
7	6	1.6	13.7	大学生数	1.8	
8	7	2	15.7	销售额预测值	15.0	
9	8	2	16.9			
10	9	2.2	14.9			
11	10	2.6	20.2			

回归

输入

Y 值输入区域(Y): \$C\$2:\$C\$11

X 值输入区域(X): \$B\$2:\$B\$11

标志(L)  常数为零(Z)

置信度(E) 95 %

输出选项

输出区域(O): \$H\$1

新工作表组(P):

新工作簿(W)

残差

残差(R)  残差图(D)

标准残差(I)  线性拟合图(L)

正态分布

正态概率图(N)

确定 取消 帮助(H)

选择“数据”菜单下的“数据分析”，选择“回归”。数据不包括列名就不用勾选“标志”



# 例4-1 一元线性回归模型

## 方法四 回归分析报告

SUMMARY OUTPUT

回归统计	
Multiple R	0.950122955
R Square	0.90273363
Adjusted F	0.890575334
标准误差	1.382931669
观测值	10

方差分析

	df	SS	MS	F	Significance F
回归分析	1	142	142	74.2484	2.54887E-05
残差	8	15.3	1.9125		
总计	9	157.3			

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	6	0.9226	6.50334	0.00019	3.872472558	8.127527442	3.872472558	8.127527442
X Variable	5	0.58027	8.61675	2.5E-05	3.661905962	6.338094038	3.661905962	6.338094038

报告里的结果可以通过单元格引用。标准误差为预测值与观测值之间的标准误差。t 统计量=8.62说明该自变量与因变量是相关的。F统计量=74.25说明回归方程有效。



## 例4-1 一元线性回归模型

### 方法步骤：

规划求解法的应用范围较广，但无法进行深入统计分析

- 第一步，取任意参数，根据一元线性回归方程计算预测值：  
$$= \$G\$3 * B5 + \$G\$2$$
- 第二步，用规划求解求出最小的MSE和一元线性回归预测模型的最优参数：  
$$\{=AVERAGE((D2:D11-C2:C11)^2)\}$$
- 第三步，根据一元线性回归方程计算预测值。

### 回归报告法主要用于线性模型

- 第一步，用数据分析工具产生一元线性回归分析报告；
- 第二步，用报告中提供的一元线性回归方程参数计算预测值。



## 例4-2 Northwind公司销售额线性回归模型

【例4-2】试根据Northwind Trader公司在1996年7月4日至1998年5月8日期间各种商品的销售额数据建立线性回归模型，然后再进一步根据回归方程预测该公司1998年5月和6月的月销售额。

建立数据透视表，将“订购日期”放在“行”区域，“销售额”放在“Σ值”区域，采用“经典数据透视表布局”，通过“日期筛选”，去除5月的数据。将汇总数据复制到一个新区域，在最左侧添加一列“月序号”。

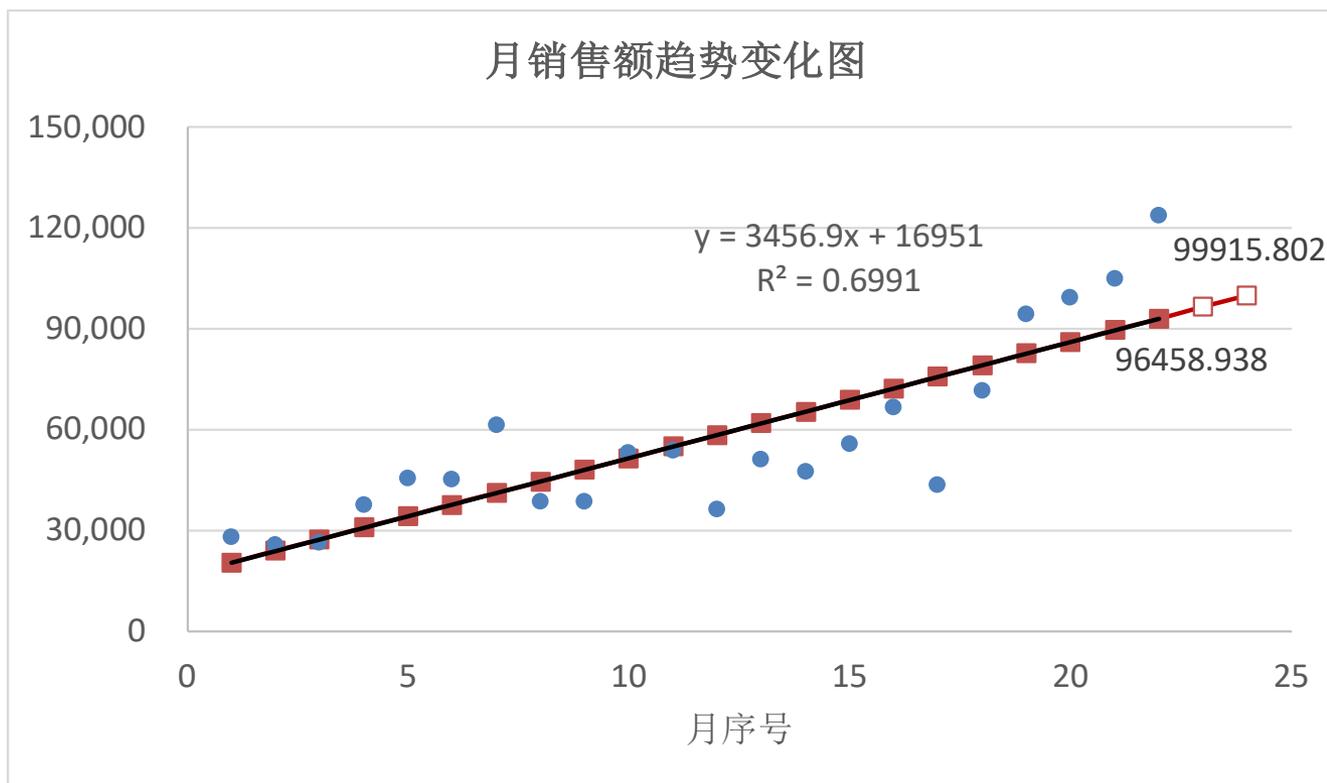


## 例4-2 Northwind公司销售额线性回归模型

	A	B	C	D	E	F	G	H	I
1	月序号	年	月	月销售额	月销售额估计值				
2	1	1996年	7月	27861.89	20407.935		回归直线方程截距 (a)	16951.072	
3	2		8月	25485.27	23864.799		回归直线方程斜率 (b)	3456.864	
4	3		9月	26381.40	27321.663		判定系数 (R <sup>2</sup> )	0.699	
5	4		10月	37515.72	30778.527		1998年5月预测值	96458.93811	
6	5		11月	45600.04	34235.390		1998年6月预测值	99915.80186	
7	6		12月	45239.63	37692.254				
8	7	1997年	1月	61258.07	41149.118				
9	8		2月	38483.63	44605.982				
10	9		3月	38547.22	48062.846				
11	10		4月	53032.95	51519.709				
12	11		5月	53781.29	54976.573				
13	12		6月	36362.80	58433.437				
14	13		7月	51020.86	61890.301				
15	14		8月	47287.67	65347.164				
16	15		9月	55629.24	68804.028				
17	16		10月	66749.23	72260.892				
18	17		11月	43533.81	75717.756				
19	18		12月	71398.43	79174.619				
20	19	1998年	1月	94225.31	82631.483				
21	20		2月	99415.29	86088.347				
22	21		3月	104901.65	89545.211				
23	22		4月	123798.68	93002.074				
24	23				96458.938				
25	24				99915.802				
26									



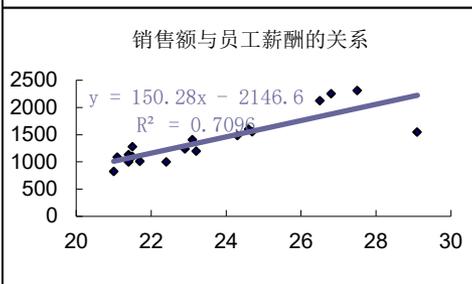
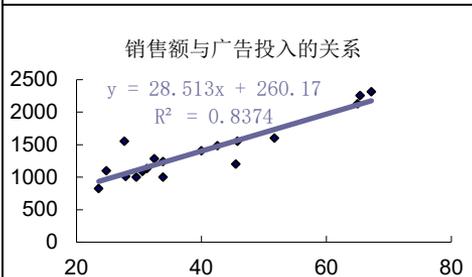
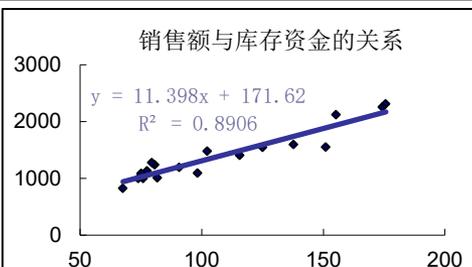
## 例4-2 Northwind公司销售额线性回归模型



三

## 多元线性回归分析

$$F_i = a + b_1X_{1i} + b_2X_{2i} + \dots + b_kX_{ki}$$



### 多元线性回归预测步骤：

1

获得候选自变量和因变量的观测值

2

从候选自变量中选择合适的自变量

3

判断回归方程的拟合优度

4

根据回归方程进行预测

例4-3

例4-4

逐步回归法、向前增选法（选出效果最好的变量继续）、  
向后删减法、最优子集法（受自变量数目影响很大）



## 例4-3 二元线性回归模型

【例4-3】某一生产空调的企业将其连续15年的销量和员工的薪酬及当地的平均户总收入情况的数据作了一个汇总，这些数据显示在工作表单元格A1：D16中，该企业的管理人员试图根据这些数据找到销量与其他两个变量之间的关系，以便进行销量的预测并为未来的预算工作提供参考。试根据这些数据分析一下，建立何种模型比较合适，并假设未来某月员工的薪酬为25万元，平均户总收入为33.4的情况下，预测该年的销量。



# 例4-3 二元线性回归模型

## 方法一 回归分析报告

SUMMARY OUTPUT

回归统计	
Multiple R	0.999
R Square	0.998
Adjusted R Square	0.998
标准误差	31.69
观测值	15

回归方程截距	105.44
斜率1 (对应员工薪酬)	5.92
斜率2 (对应平均户总收入)	8.65
拟合优度R <sup>2</sup>	0.998
调整后的R <sup>2</sup>	0.998
员工薪酬	250
平均户总收入	33.4
销量预测值 (千台)	1874.5

方差分析

	df	SS	MS	F	Significance F
回归分析	2	7546071	3773036	3758.2	1.64017E-17
残差	12	12047.5	1003.954		
总计	14	7558119			

年	员工薪酬(千元)	平均户月收入(千元)	销量(千台)
1	275	24.5	1924
2	182	32.5	1402
3	376	38	2666
4	204	28.4	1572
5	85	23.5	802
6	267	37.8	2026
7	96	30.1	970
8	331	24.5	2305
9	196	21.4	1393
10	54	25.6	658
11	432	40.2	3021
12	373	44.3	2684
13	235	26.6	1738
14	156	20.9	1246
15	372	26.1	2534

	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	105.4	35.4313	2.975941	0.0116	28.24330877	182.63979	28.243309	182.63979
员工薪酬(千元)	5.921	0.08761	67.57784	7E-17	5.729878086	6.1116676	5.7298781	6.1116676
平均户月收入(千元)	8.648	1.4119	6.125277	5E-05	5.57200312	11.724523	5.5720031	11.724523



## 例4-3 二元线性回归模型

	A	B	C	D	E	F	G	H
1	年	员工薪酬(千元)	平均户月收入(千元)	销量(千台)	二元线性回归			
2	1	275	24.5	1924	300.5		销量预测值:	
3	2	182	32.5	1402	215.5		回归方程截距	1.00
4	3	376	38	2666	415		斜率1 (对应员工薪酬)	1.00
5	4	204	28.4	1572	233.4		斜率2 (对应平均户总收入)	1.00
6	5	85	23.5	802	109.5		MSE	2670757.36
7	6	267	37.8	2026	305.8		员工薪酬	250
8	7	96	30.1	970	127.1		平均户总收入	33.4
9	8	331	24.5	2305	356.5		销量预测值 (千台)	284.4
10	9	196	21.4	1393	218.4			
11	10	54	25.6	658	80.6			
12	11	432	40.2	3021	473.2			
13	12	373	44.3	2684	418.3			
14	13	235	26.6	1738	262.6			
15	14	156	20.9	1246	177.9			
16	15	372	26.1	2534	399.1			

方法二  
规划求解



## 例4-4 多元线性回归模型

【例4-4】一家皮鞋零售店将其连续18个月的库存占用资金情况、广告投入的费用、员工薪酬以及销售额等方面的数据作了一个汇总，这些数据显示在工作表A1:E20范围。该皮鞋店的管理人员试图根据这些数据找到销售额与其他三个变量之间的关系，以便进行销售额预测并为未来的预算工作提供参考。试根据这些数据建立回归模型。如果未来某月库存资金额为150万元，广告投入预算为45万元，员工薪酬总额为27万元，试根据建立的回归模型预测该月的销售额。

# 例4-4 多元线性回归模型

## 方法一 回归分析报告

=库存资金、广告回归分析报告!B17  
=库存资金、广告回归分析报告!B18  
=库存资金、广告回归分析报告!B19

月份	库存资金 $X_1$ (万元)	广告投入 $X_2$ (万元)	员工薪酬总额 $X_3$ (万元)	销售额Y (万元)
1	75.2	30.6	21.1	1090.4
2	77.6	31.3	21.4	1133
3	80.7	33.9	22.9	1242.1
4	76	29.6	21.4	1003.2
5	79.5	32.5	21.5	1283.2
6	81.8	27.9	21.7	1012.2
7	98.3	24.8	21.5	1098.8
8	67.7	23.6	21	826.3
9	74	33.9	22.4	1003.3
10	151	27.7	24.7	1554.6
11	90.8	45.5	23.2	1199
12	102.3	42.6	24.3	1483.1
13	115.6	40	23.1	1407.1
14	125	45.8	29.1	1551.3
15	137.8	51.7	24.6	1601.2
16	175.6	67.2	27.5	2311.7
17	155.2	65	26.5	2126.7
18	174.3	65.4	26.8	2256.5

自变量集		$R^2$	调整后 $R^2$
库存资金	$X_1$	0.891	0.884
广告	$X_2$	0.837	0.827
薪酬	$X_3$	0.710	0.691
库存资金、广告	$X_1, X_2$	0.957	0.952
广告、薪酬	$X_2, X_3$	0.870	0.852
库存资金、薪酬	$X_1, X_3$	0.898	0.885
库存资金、广告、薪酬	$X_1, X_2, X_3$	0.957	0.948

销售额预测:

回归方程截距	86.95
斜率1 (对应库存资金)	7.11
斜率2 (对应广告)	13.68
斜率3 (对应员工薪酬)	0.00
库存资金 (万元)	150
广告 (万元)	45
员工薪酬 (万元)	27
销售额预测值 (万元)	1769.1

参与多元回归的自变量列必须相邻 (如 $X_1$ 、 $X_3$ )

之前生成的回归分析报告不会自动重算



# 例4-4 多元线性回归模型

## 方法二 规划求解工具

$$F_i = a + b_1X_{1i} + b_2X_{2i} + \dots + b_kX_{ki}$$

	A	B	C	D	E	F	G	H	I
1	月份	库存资金额 $X_1$	广告投入 $X_2$	员工薪酬总额 $X_3$	销售额Y	预测值			
2		(万元)	(万元)	(万元)	(万元)	(万元)			
3	1	75.2	30.6	21.1	1090.4	1040.27			
4	2	77.6	31.3	21.4	1133	1066.91			
5	3	80.7	33.9	22.9	1242.1	1124.525			
6	4	76	29.6	21.4	1003.2	1032.273			
7	5	79.5	32.5	21.5	1283.2	1096.837			
8	6	81.8	27.9	21.7	1012.2	1050.241			
9	7	98.3	24.8	21.5	1098.8	1125.116			
10	8	67.7	23.6	21	826.3	891.1662			
11	9	74	33.9	22.4	1003.3	1076.896			
12	10	151	27.7	24.7	1554.6	1539.432			
13	11	90.8	45.5	23.2	1199	1355.057			
14	12	102.3	42.6	24.3	1483.1	1397.125			
15	13	115.6	40	23.1	1407.1	1456.094			
16	14	125	45.8	29.1	1551.3	1602.283			
17	15	137.8	51.7	24.6	1601.2	1774.011			
18	16	175.6	67.2	27.5	2311.7	2254.824			
19	17	155.2	65	26.5	2126.7	2079.7			
20	18	174.3	65.4	26.8	2256.5	2220.951			

回归方程截距	86.96
斜率1 (对应库存资金)	7.11
斜率2 (对应广告)	13.68
斜率3 (对应员工薪酬)	0.00
MSE	7866.19

预测销售额:	
库存资金 (万元)	150
广告 (万元)	45
员工薪酬 (万元)	27
销售额预测值 (万元)	1769.1

用规划求解求出最小的MSE和最优回归模型参数a和 $b_1$ 、 $b_2$ 、 $b_3$



## 例4-4 多元线性回归模型

### 方法步骤：

#### 回归报告法

- 第一步，分别产生因变量关于各自变量的一元线性回归分析报告，找出 $R^2$ 最大所对应的变量作为第一影响因素；
- 第二步，分别产生因变量关于自变量的多份二元线性回归分析报告，找出调整后的 $R^2$ 最大所对应的第二自变量；
- 第三步，分别产生因变量关于自变量的多份三元线性回归分析报告，再次找出调整后的 $R^2$ 最大所对应的第三自变量；
- ……
- 第四步，根据调整后的 $R^2$ 最大所对应的那份回归分析报告所提供的多元线性回归方程参数计算预测值。



## 例4-4 多元线性回归模型

### 方法步骤：

#### 规划求解法

- 第一步，取任意参数，根据多元线性回归方程计算预测值；

$$= \$1 \$2 + \$1 \$3 * B3 + \$1 \$4 * C3 + \$1 \$5 * D3$$

- 第二步，用规划求解求出最小的MSE和多元线性回归预测模型的最优参数；

$$\{=AVERAGE((E3:E20-F3:F20)^2)\}$$

- 第三步，根据多元线性回归方程计算预测值。

## 四 一元非线性回归分析

### 一元非线性回归的方法

规划求解法：用规划求解  
工具找到回归方程的参数。

$$F_i = f(X_i; a, b, \dots)$$

$$\hookrightarrow MSE = \frac{1}{n} \sum_{i=1}^n (F_i - Y_i)^2 \text{ 最小}$$

$$\hookrightarrow \text{最优 } a, b, \dots$$

变量替换+回归分析报告法：通过变量替换，把非线性问题转化为线性回归问题后，用线性回归分析的方法找到回归方程的参数。

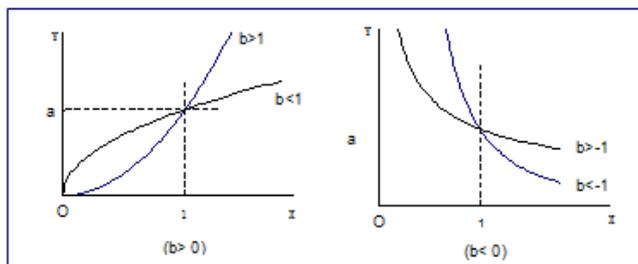


# 常用回归拟合曲线及变量替换方法

◎ 幂函数  $Y = aX^b$

设  $U = \ln X$      $V = \ln Y$

则  $V = \ln a + bU$

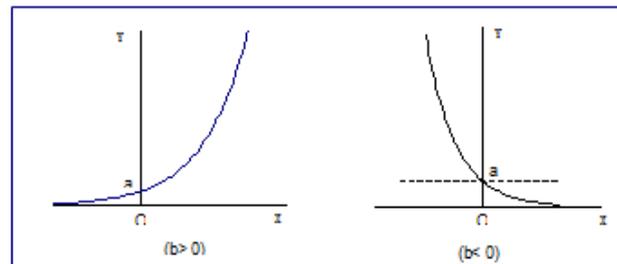


◎ 指数函数  $Y = ae^{bX}$

设  $V = \ln Y$

则  $V = \ln a + (b \ln e)X$

例4-5

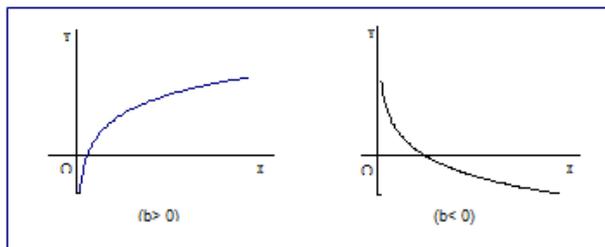


◎ 对数函数  $Y = a + b \ln X$

设  $U = \ln X$

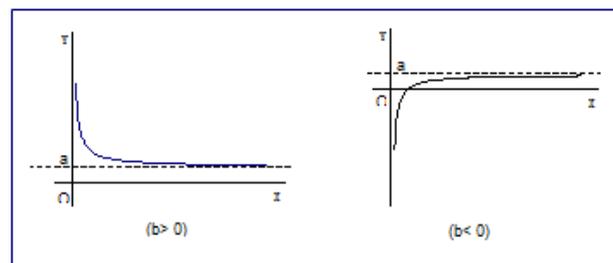
则  $Y = a + bU$

例4-6



◎ 双曲线函数  $Y = a + b \frac{1}{X}$

设  $U = \frac{1}{X}$  则  $Y = a + bU$





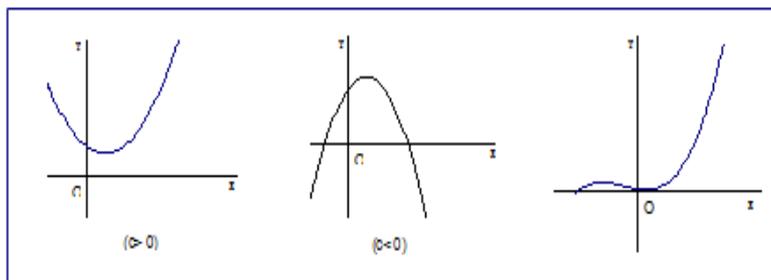
## 常用回归拟合曲线及变量替换方法

### 二次多项式及三次多项式

$$Y = a + bX + cX^2$$

$$Y = a + bX + cX^2 + dX^3$$

例4-7



.....

设  $X_1 = X$ ,  $X_2 = X^2$ ,  $\dots$ ,  $X_k = X^k$ , 变换为  $Y = a + bX_1 + cX_2 + \dots$



## 例4-5 指数回归模型

【例4-5】表4-4列出了连续十三年对某消费品年销售额的统计数据。试根据这些资料建立适当的模型，并预测第14年的销售额预测值。

年序号t	年销售额Y	年序号t	年销售额Y
1	3	8	36
2	8	9	32
3	12	10	57
4	10	11	70
5	25	12	115
6	14	13	150
7	18		



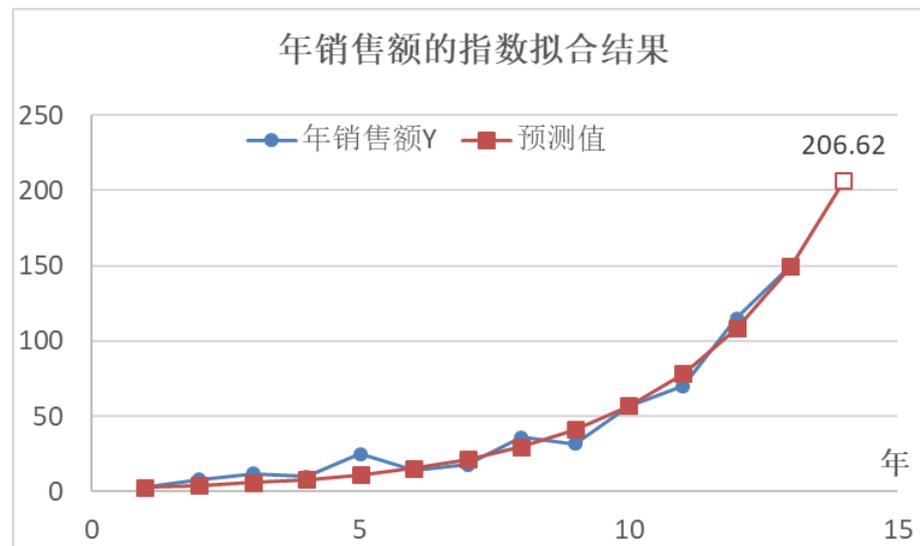
## 例4-5 指数回归模型

### 方法一 规划求解法

$$F_i = ae^{bX_i}$$

	A	B	C	D	E	F
1	年序号t	年销售额Y	预测值			
2	1	3	3.13	a		2.269
3	2	8	4.32	b		0.322
4	3	12	5.97	MSE		38.044
5	4	10	8.24			
6	5	25	11.37	年序号		14
7	6	14	15.69	年销售额预测值		206.62
8	7	18	21.65			
9	8	36	29.89			
10	9	32	41.25			
11	10	57	56.94			
12	11	70	78.58			
13	12	115	108.46			
14	13	150	149.70			
15	14		206.62			

用规划求解求出最小的MSE和最优回归模型参数a和b





# 例4-5 指数回归模型

指数函数  $Y = ae^{bX}$

设  $V = \ln Y$

则  $V = \ln a + (b \ln e)X$

## 方法二 变量替换+回归分析报告法

	A	B	C	D	E	F	G
1	年序号t	年销售额Y	V=lnY	预测值			
2	1	3	1.0986	4.72	a		3.573
3	2	8	2.0794	6.22	ln(a)		1.273
4	3	12	2.4849	8.22	b		0.278
5	4	10	2.3026	10.84	R <sup>2</sup>		0.929
6	5	25	3.2189	14.31			
7	6	14	2.6391	18.89	年序号		14
8	7	18	2.8904	24.93	年销售额预测值		173.99
9	8	36	3.5835	32.91			
10	9	32	3.4657	43.43			
11	10	57	4.0431	57.33			
12	11	70	4.2485	75.67			
13	12	115	4.7449	99.87			
14	13	150	5.0106	131.82			
15	14			173.99			

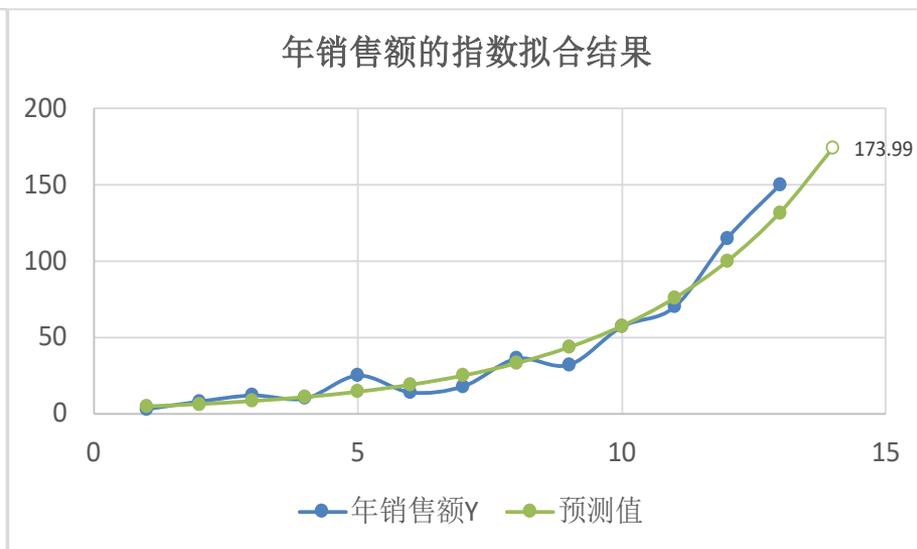
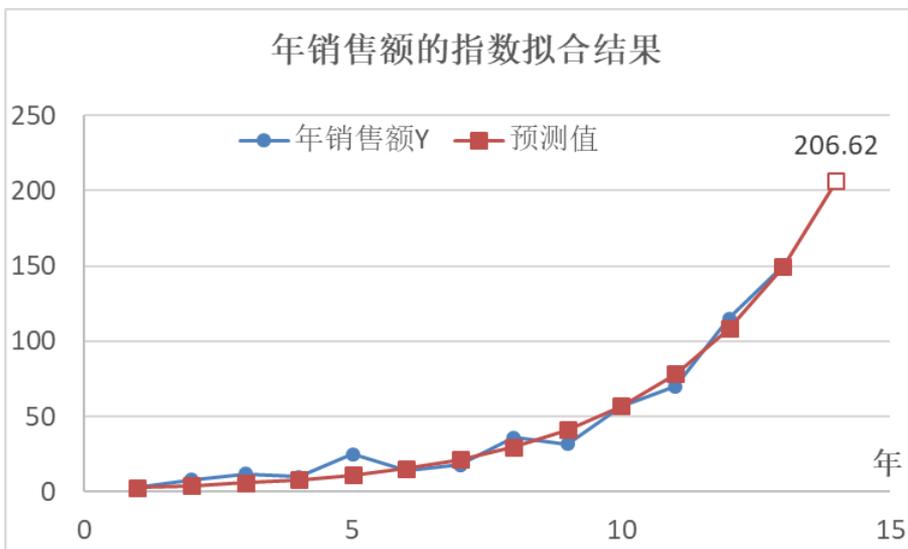
	I	J	K	L
1	SUMMARY OUTPUT			
2				
3	回归统计			
4	Multiple R	0.963932		
5	R Square	0.929165		
6	Adjusted R	0.922726		
7	标准误差	0.311711		
8	观测值	13		
15				
16	Coefficients	标准误差	t Stat	
17	Intercept	1.273343	0.183395	6.943178
18	X Variable	0.277547	0.023106	12.01212

C2=LN(A2) 和 G2=EXP(G3)



# 例4-5 指数回归模型

## 比较方法一和方法二





## 例4-6 一元非线性回归模型

【例4-6】某企业想了解公司某种产品产量与收益之间的关系，为此收集整理了历年的产量收益数据资料，如本工作表所列。根据这些资料建立适当模型说明产量与收益之间的关系。



## 例4-6 一元非线性回归模型

产量(X)	收益(Y)	收益预测值
473	1.47	7.166
639	7.94	10.410
741	7.28	12.404
824	6.55	14.026
874	9.18	15.003
914	16.67	15.785
939	14.9	16.274
956	14.91	16.606
972	15.81	16.919
1024	19.63	17.935
1055	17.41	18.541
1056	17.42	18.560
1132	22.51	20.046
1136	15.39	20.124
1075	22.53	18.932
1240	24.58	22.157
1253	27.93	22.411
1281	24.51	22.958
1285	20.55	23.036
1319	23.31	23.701
1366	27.01	24.619
1403	30.3	25.343
1407	29.52	25.421
1443	29.39	26.124
1457	32.36	26.398
1467	30.25	26.502

a	-2.079
b	0.0195
R <sup>2</sup>	0.848

### SUMMARY OUTPUT

#### 回归统计

Multiple R	0.9209778
R Square	0.8482001
Adjusted F	0.84503761
标准误差	3.74297356
观测值	50

#### 方差分析

	df	SS	MS	F	Significance F
回归分析	1	3757.523	3757.523	268.2057	2.78E-21
残差	48	672.4729	14.00985		
总计	49	4429.995			

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限	95.0%上限	95.0%
Intercept	-2.0860942	1.80035	-1.15872	0.252304	-5.70594	1.533752	-5.70594	1.533752	
X Variable	0.01955038	0.001194	16.37699	2.78E-21	0.01715	0.021951	0.01715	0.021951	



## 例4-6 一元非线性回归模型

回归

输入

Y 值输入区域(Y): \$B\$2:\$B\$51

X 值输入区域(X): \$A\$2:\$A\$51

标志(L)       常数为零(Z)

置信度(E)      95 %

输出选项

输出区域(O): \$H\$1

新工作表组(P):

新工作簿(W)

残差

残差(R)       残差图(D)

标准残差(I)       线性拟合图(U)

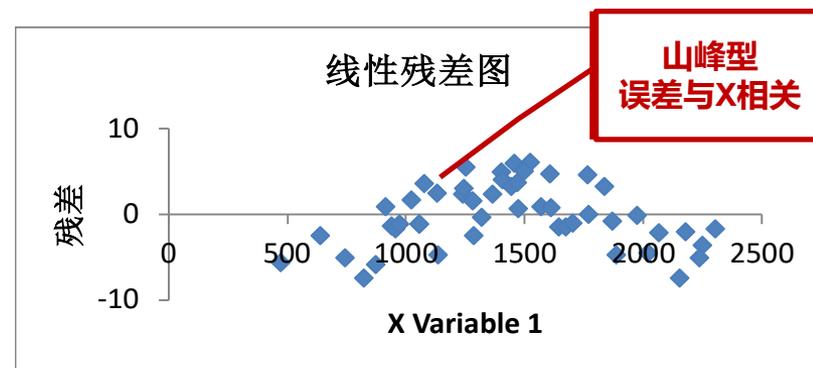
正态分布

正态概率图(N)

确定 取消 帮助(H)

RESIDUAL OUTPUT

观测值	预测 Y	残差
1	7.16123666	-5.69124
2	10.4066001	-2.4666
3	12.4007391	-5.12074
4	14.0234209	-7.47342
5	15.00094	-5.82094
6	15.7829553	0.887045
7	16.2717149	-1.37171
8	16.6040714	-1.69407
9	16.9168775	-1.10688
10	17.9334974	1.696503
11	18.5395592	-1.12956
12	18.5591096	-1.13911
13	20.0449387	2.465061
14	20.1231402	-4.73314
15	18.9305669	3.599433
16	22.15638	2.42362
17	22.4105349	5.519465
18	22.9579456	1.552054
19	23.0361472	-2.48615
20	23.7008602	-0.39086
21	24.6107281	2.300272



残差图反映了残差和自变量之间的散点图。如果拟合结果理想，残差图中的各点应该较为均匀地分布在横轴的上下方



# 例4-6 一元非线性回归模型

对数函数  $Y = a + b \ln X$

设  $U = \ln X$

则  $Y = a + bU$

产量(X)	$U = \ln(X)$	收益(Y)	收益预测值
473	6.1591	1.47	-1.878
639	6.4599	7.94	6.051
741	6.608	7.28	9.955
824	6.71417	6.55	12.754
874	6.77308	9.18	14.307
914	6.81783	16.67	15.487
939	6.84482	14.9	16.198
956	6.86276	14.91	16.671
972	6.87936	15.81	17.109
1024	6.93147	19.63	18.482
1055	6.9613	17.41	19.269
1056	6.96224	17.42	19.294
1132	7.03174	22.51	21.126
1136	7.03527	15.39	21.219
1075	6.98008	22.53	19.764
1240	7.12287	24.58	23.528
1253	7.1333	27.93	23.803
1281	7.1554	24.51	24.385
1285	7.15851	20.55	24.467
1319	7.18463	23.31	25.156
1366	7.21964	27.01	26.079
1403	7.24637	30.3	26.783
1407	7.24922	29.52	26.858
1443	7.27448	29.39	27.524
1457	7.28413	32.36	27.779
1467	7.29097	30.35	27.959

SUMMARY OUTPUT

回归统计	
a	-164.240
b	26.3613
R <sup>2</sup>	0.905
Multiple R	0.95138387
R Square	0.90513126
Adjusted R Square	0.90315483
标准误差	2.95898378
观测值	50

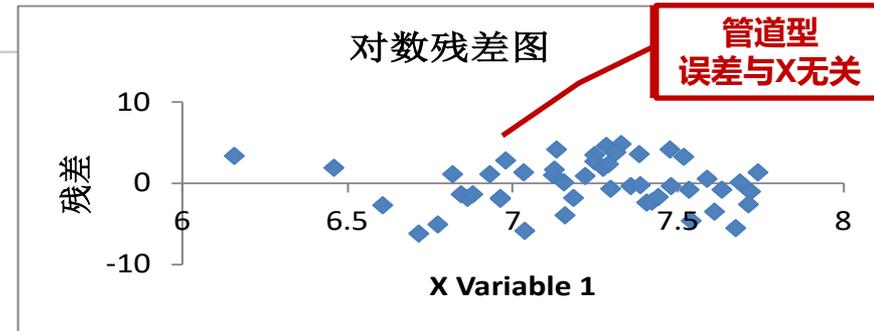
方差分析

	df	SS	MS	F	Significance F
回归分析	1	4009.727	4009.727	457.9622	3.39E-26
残差	48	420.2681	8.755585		
总计	49	4429.995			

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限	95.0%上限	95.0%
Intercept	-164.24025	8.903987	-18.4457	1.99E-23	-182.143	-146.338	-182.143	-146.338	
X Variable 1	26.3612999	1.231833	21.40005	3.39E-26	23.88453	28.83807	23.88453	28.83807	

RESIDUAL OUTPUT

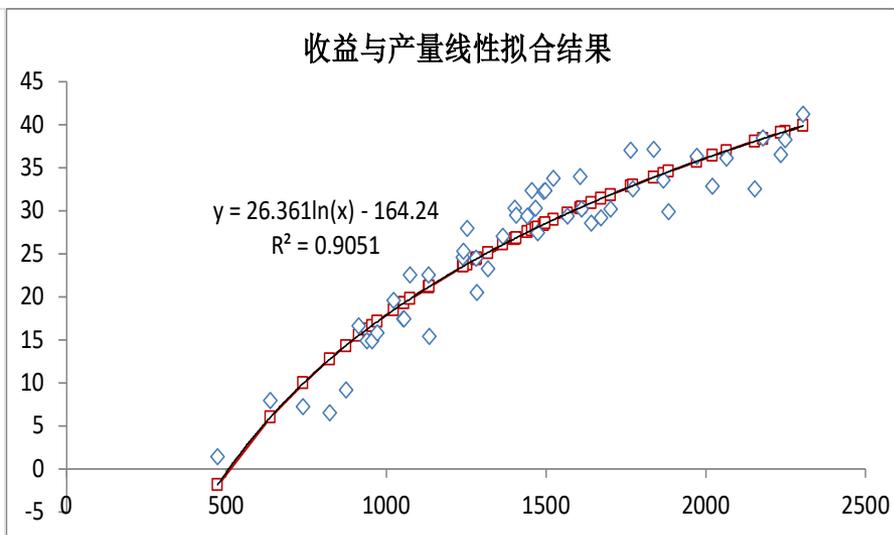
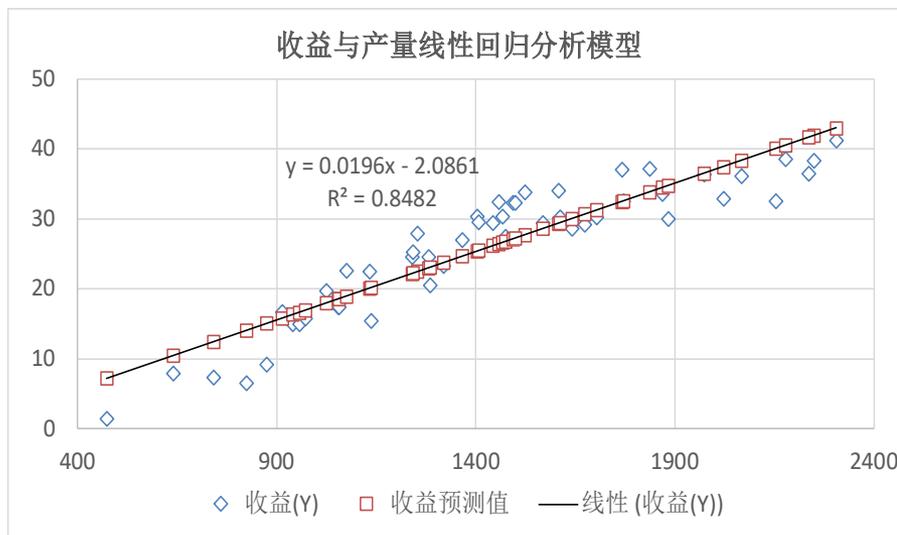
观测值	预测 Y	残差
1	-1.8784866	3.348487
2	6.05123141	1.888769
3	9.95523899	-2.67524





# 例4-6 一元非线性回归模型

## 比较线性和非线性回归模型





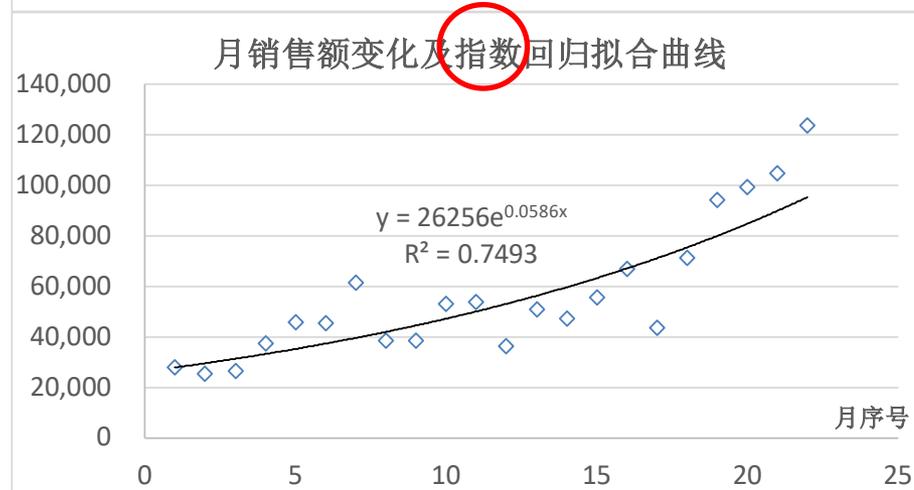
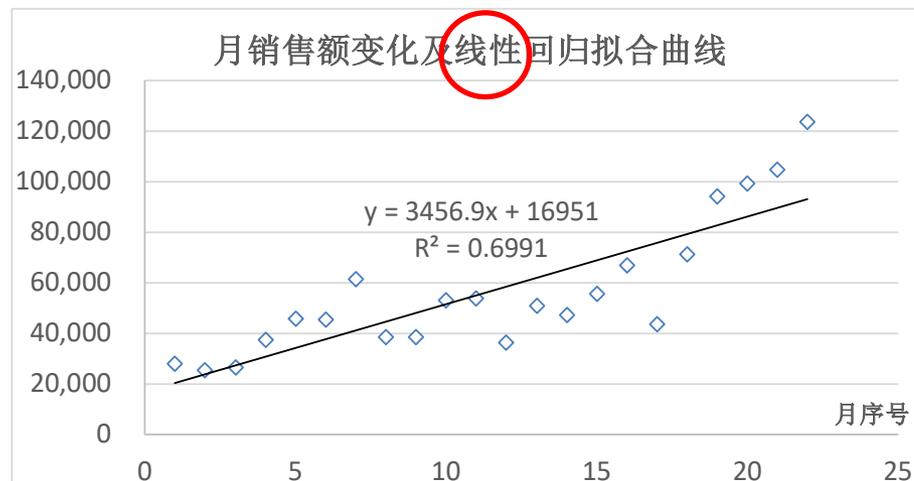
## 例4-7 非线性回归模型

【例4-7】 就【例4-2】Northwind Trader公司的销售额数据，进行非线性回归分析，并预测98年5月和6月的销售额。



# 例4-7 非线性回归模型

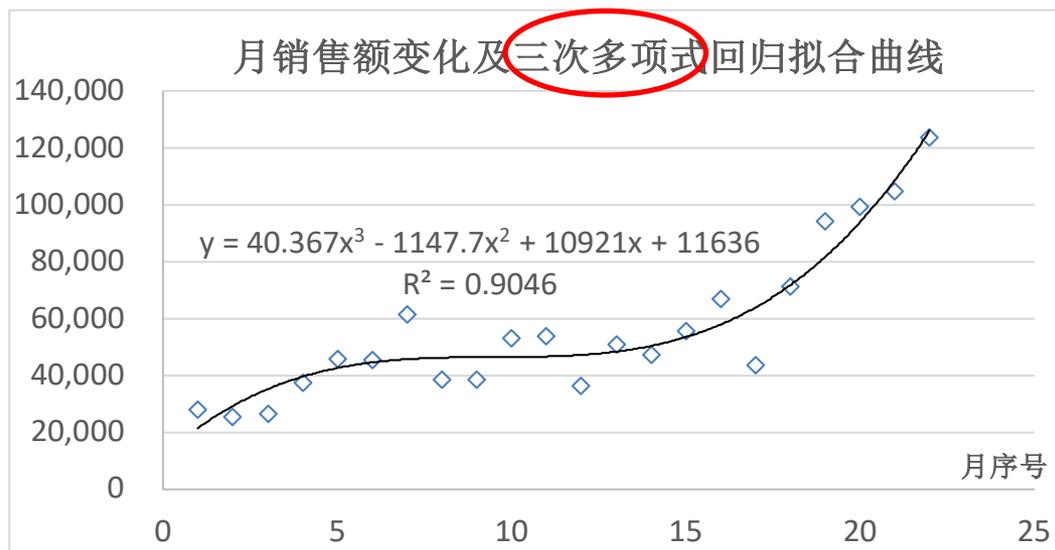
月序号X	X <sup>2</sup> (U)	X <sup>3</sup> (V)	年	月	月销售额	月预测值
1	1	1	1996年	7月	27861.89	21450.02
2	4	8		8月	25485.27	29210.97
3	9	27		9月	26381.40	35160.96
4	16	64		10月	37515.72	39542.20
5	25	125	1997年	11月	45600.04	42596.89
6	36	216		12月	45239.63	44567.22
7	49	343		1月	61258.07	45695.41
8	64	512		2月	38483.63	46223.64
9	81	729	1998年	3月	38547.22	46394.13
10	100	1000		4月	53032.95	46449.06
11	121	1331		5月	53781.29	46630.66
12	144	1728		6月	36362.80	47181.11
13	169	2197	1998年	7月	51020.86	48342.61
14	196	2744		8月	47287.67	50357.38
15	225	3375		9月	55629.24	53467.60
16	256	4096		10月	66749.23	57915.48
17	289	4913	1998年	11月	43533.81	63943.23
18	324	5832		12月	71398.43	71793.03
19	361	6859		1月	94225.31	81707.11
20	400	8000		2月	99415.29	93927.65
21	441	9261	1998年	3月	104901.65	108696.85
22	484	10648		4月	123798.68	126256.92
23	529	12167		5月		146850.06
24	576	13824		6月		170718.48





## 例4-7 非线性回归模型

月序号X	X <sup>2</sup> (U)	X <sup>3</sup> (V)	年	月	月销售额	月预测值
1	1	1	1996年	7月	27861.89	21450.02
2	4	8		8月	25485.27	29210.97
3	9	27		9月	26381.40	35160.96
4	16	64	1997年	10月	37515.72	39542.20
5	25	125		11月	45600.04	42596.89
6	36	216		12月	45239.63	44567.22
7	49	343		1月	61258.07	45695.41
8	64	512	1998年	2月	38483.63	46223.64
9	81	729		3月	38547.22	46394.13
10	100	1000		4月	53032.95	46449.06
11	121	1331		5月	53781.29	46630.66
12	144	1728	1998年	6月	36362.80	47181.11
13	169	2197		7月	51020.86	48342.61
14	196	2744		8月	47287.67	50357.38
15	225	3375		9月	55629.24	53467.60
16	256	4096	1998年	10月	66749.23	57915.48
17	289	4913		11月	43533.81	63943.23
18	324	5832		12月	71398.43	71793.03
19	361	6859		1月	94225.31	81707.11
20	400	8000	1998年	2月	99415.29	93927.65
21	441	9261		3月	104901.65	108696.85
22	484	10648		4月	123798.68	126256.92
23	529	12167		5月		146850.06
24	576	13824	6月		170718.48	



二次多项式及三次多项式

$$Y = a + bX + cX^2$$

$$Y = a + bX + cX^2 + dX^3$$

“添加趋势线”，选择“多项式”，将“顺序”调为“3”



## 例4-7 非线性回归模型

### SUMMARY OUTPUT

回归统计	
Multiple R	0.9511006
R Square	0.9045923
Adjusted R	0.888691
标准误差	8957.0402
观测值	22

### 方差分析

	df	SS	MS	F	gnificance F
回归分析	3	1.369E+10	4.564E+09	56.888025	2.207E-09
残差	18	1.444E+09	80228570		
总计	21	1.514E+10			

截距 (a)	11635.911
X <sup>1</sup> 项系数 (b <sub>1</sub> )	10921.419
X <sup>2</sup> 项系数 (b <sub>2</sub> )	-1147.679
X <sup>3</sup> 项系数 (b <sub>3</sub> )	40.367
判定系数 (R <sup>2</sup> )	0.905
调整后的 (R <sup>2</sup> )	0.889
1998年5月预测值	146850.06
1998年6月预测值	170718.48

	Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	11635.911	9150.8594	1.2715649	0.2197129	-7589.346	30861.168	-7589.346	30861.1684
月序号X	10921.419	3369.2755	3.2414739	0.0045297	3842.8281	18000.01	3842.8281	18000.0095
X <sup>2</sup> (U)	-1147.679	336.44256	-3.411219	0.0031132	-1854.519	-440.839	-1854.519	-440.83896
X <sup>3</sup> (V)	40.366875	9.6292223	4.192122	0.0005477	20.136614	60.597136	20.136614	60.5971356



## 例4-5、6、7 非线性回归模型

### 方法步骤：

#### 规划求解法

- 第一步，取任意参数，根据非线性回归方程计算预测值；  
指数方程、对数方程、多项式方程、……
- 第二步，用规划求解求出最小的MSE和非线性回归预测模型的最优参数；  
MSE计算公式
- 第三步，根据非线性回归方程计算预测值。

非线性模型的解  
可能是局部最优

回归分析报告产生  
的是线性方程参数

#### 变量替换+回归分析报告法

- 第一步，添加各种趋势线，找出 $R^2$ 最大的非线性回归拟合方程；
- 第二步，将非线性回归拟合方程变换为线性方程；
- 第三步，利用线性回归分析报告所提供的方程参数计算预测值。

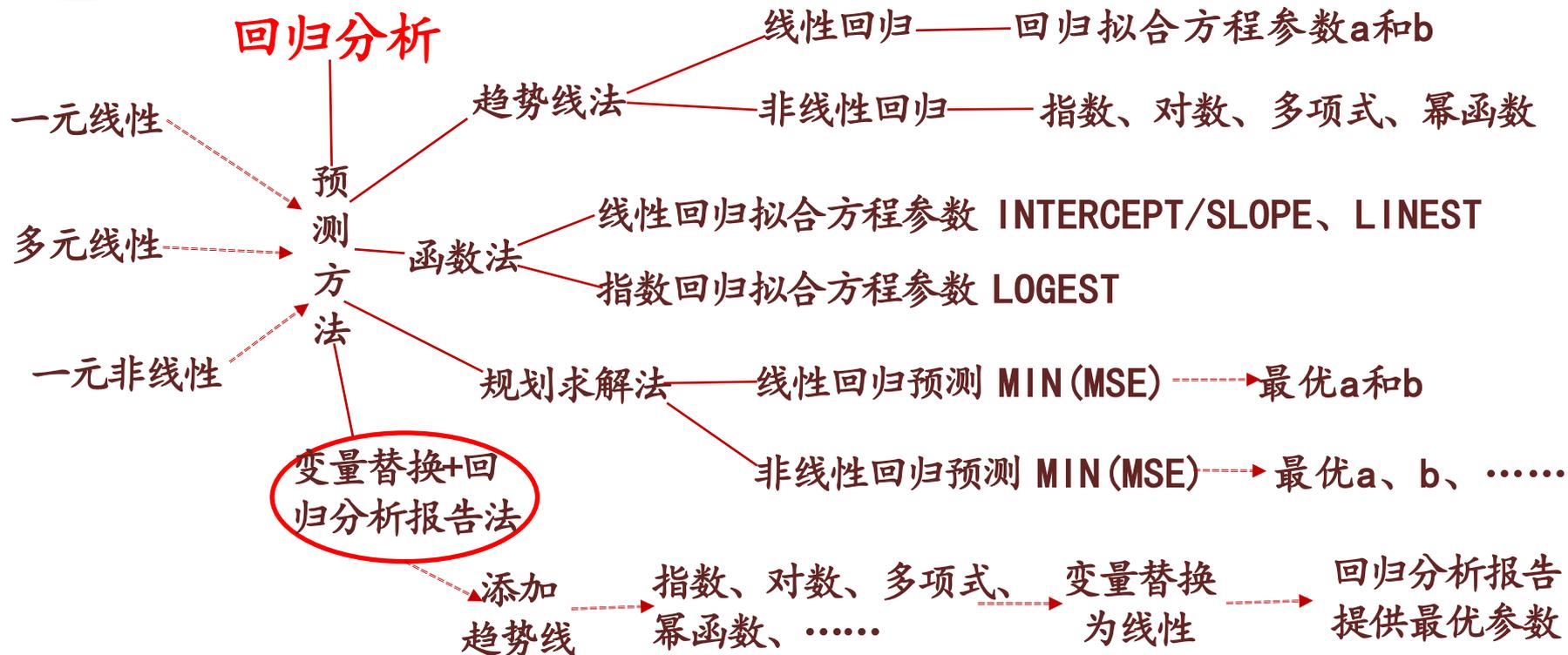
## 本章学习的Excel函数有：

INTERCEPT ()、SLOPE ()、LINEST ()、  
SUMXMY2 ()、AVERAGE () 等。

## 本章用到的Excel工具主要有：

“规划求解”工具、“回归分析  
报告”数据分析工具。

## 第四章 总结



Q&A?



**经济与管理学院**  
SCHOOL OF ECONOMICS & MANAGEMENT

**谢 谢**

博學而後身正 篤新