

《传播统计学》

回归分析（二）

教师：林志良

邮箱：linzhl@nfu.edu.cn

个人网站：www.zhilianglin.com

目录

- 多元回归模型与方程
- 案例
- 自变量为分类型变量

多元回归模型与方程

多元回归模型

检验

1个因变量 (Y) 与 2个或以上自变量 (X_i) 是否存在线性关系

有 k 个自变量的多元回归模型

截距项

总体的斜率

残差项

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i$$

多元回归模型与方程

多元回归方程

用样本数据估计多元回归方程的回归系数

有k个自变量的多元回归方程

$$\hat{Y}_i = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Y的预测值
(估计值)

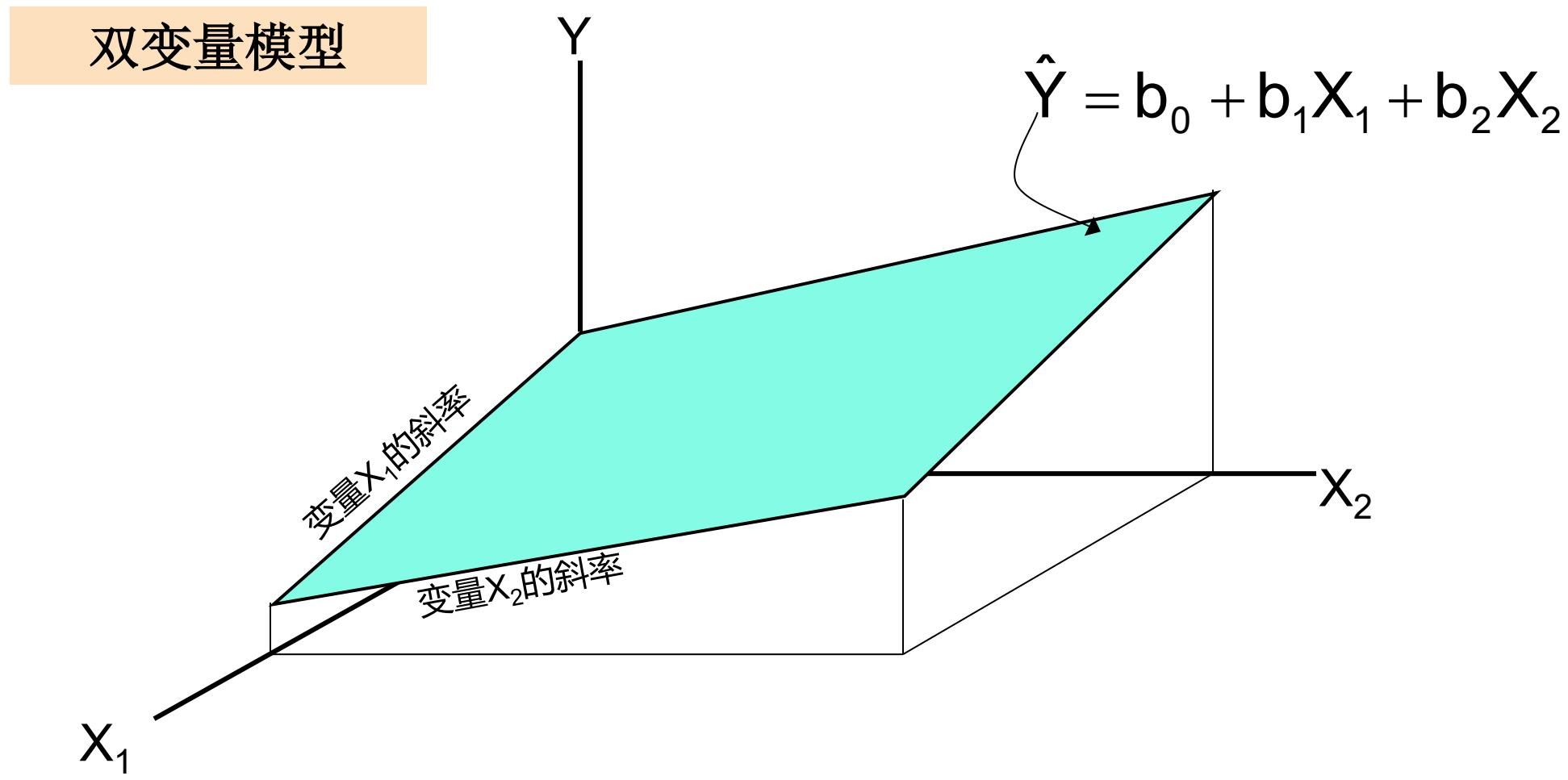
估计的截距

估计的回归系数

```
graph TD; A[Y的预测值<br>(估计值)] --> D["b0"]; B[估计的截距] --> D["b1X1"]; C[估计的回归系数] --> D["b2X2"]; C --> D["..."]; C --> D["bkXk"]
```

多元回归模型与方程

多元回归方程



案例

周数	销量	价格 (\$)	广告费 (\$100s)
1	350	5.50	3.3
2	460	7.50	3.3
3	350	8.00	3.0
4	430	8.00	4.5
5	350	6.80	3.0
6	380	7.50	4.0
7	430	4.50	3.0
8	470	6.40	3.7
9	450	7.00	3.5
10	490	5.00	4.0
11	340	7.20	3.5
12	300	7.90	3.2
13	440	5.90	4.0
14	450	5.00	3.5
15	300	7.00	2.7

馅饼销量

多元回归方程：

$$\hat{\text{销量}} = b_0 + b_1 (\text{价格}) + b_2 (\text{广告费})$$



案例

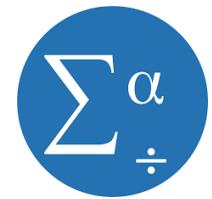


SUMMARY OUTPUT								
回归统计								
Multiple R	0.722134292							
R Square	0.521477936							
Adjusted R Squa	0.441724259							
标准误差	47.46341263							
观测值	15							
方差分析								
	df	SS	MS	F	ignificance F			
回归分析	2	29460.03	14730.01	6.538607	0.012006			
残差	12	27033.31	2252.776					
总计	14	56493.33						
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	306.5261933	114.2539	2.682851	0.019932	57.58834	555.464	57.58834	555.464
价格(\$)	-24.97508952	10.83213	-2.30565	0.039788	-48.5763	-1.37392	-48.5763	-1.37392
广告费(\$100s)	74.13095749	25.96732	2.854779	0.014494	17.55303	130.7089	17.55303	130.7089

回归方程：

$$\widehat{\text{销量}} = 306.53 - 24.98 \text{ (价格)} + 74.13 \text{ (广告费)}$$

案例



输入/除去的变量^a

模型	输入的变量	除去的变量	方法
1	广告费, 价格 ^b	.	输入

- a. 因变量: 销量
 b. 已输入所请求的所有变量。

模型摘要

模型	R	R 方	调整后 R 方	标准估算的错误
1	.722 ^a	.521	.442	47.463

a. 预测变量: (常量), 广告费, 价格

ANOVA^a

模型		平方和	自由度	均方	F	显著性
1	回归	29460.027	2	14730.013	6.539	.012 ^b
	残差	27033.306	12	2252.776		
	总计	56493.333	14			

- a. 因变量: 销量
 b. 预测变量: (常量), 广告费, 价格

系数^a

模型		未标准化系数		Beta	t	显著性
		B	标准错误			
1	(常量)	306.526	114.254		2.683	.020
	价格	-24.975	10.832	-.461	-2.306	.040
	广告费	74.131	25.967	.570	2.855	.014

a. 因变量: 销量

回归方程:

$$\widehat{\text{销量}} = 306.53 - 24.98 \text{ (价格)} + 74.13 \text{ (广告费)}$$

案例

$$\widehat{\text{销量}} = 306.53 - 24.98 \text{ (价格)} + 74.13 \text{ (广告费)}$$



$b_1 = -24.97$:
在广告费不变的前提下，馅饼的价格每提升1美元，馅饼销量就下降24.98个。

$b_2 = 74.13$:
在价格不变的前提下，广告费每提升1美元，馅饼销量就上升74.13个。



案例

用回归方程预测

价格为\$5.50，广告费为\$350，预测馅饼的销量：

$$\begin{aligned}\hat{\text{销量}} &= 306.53 - 24.98(\text{价格}) + 74.13(\text{广告费}) \\ &= 306.53 - 24.98(5.50) + 74.13(3.5) \\ &= 428.62\end{aligned}$$

预测销量为
428.62个。

注意：广告费的单位是
\$100s, 因此\$350的表达
应该是 $X_2 = 3.5$

案例

判定系数：R² & 调整后的R²

模型摘要

模型	R	R 方	调整后 R 方	标准估算的错误
1	.722 ^a	.521	.442	47.463

a. 预测变量: (常量), 广告费, 价格

回归统计

Multiple R	0.722134
R Square	0.521478
Adjusted R Square	0.441724
标准误差	47.46341
观测值	15

结论:

- 馅饼价格、广告费可以解释销量52.1%的变异。
- 在考虑了样本量和自变量个数的前提下，馅饼价格、广告费可以解释销量52.1%的变异。

案例

线性关系检验：F检验

ANOVA^a

模型		平方和	自由度	均方	F	显著性
1	回归	29460.027	2	14730.013	6.539	.012 ^b
	残差	27033.306	12	2252.776		
	总计	56493.333	14			

a. 因变量: 销量

b. 预测变量: (常量), 广告费, 价格

结论:

回归模型成立 (具有统计学显著意义)。/至少有一个自变量显著影响因变量。

方差分析

	df	SS	MS	F	Significance F
回归分析	2	29460.03	14730.01	6.538607	0.012006372
残差	12	27033.31	2252.776		
总计	14	56493.33			

案例

回归系数检验：t检验

系数^a

模型	未标准化系数		标准化系数		显著性
	B	标准错误	Beta	t	
1	(常量)	306.526	114.254		.020
	价格	-24.975	10.832	-.461	-2.306
	广告费	74.131	25.967	.570	2.855

a. 因变量: 销量

结论:

在 $\alpha=0.05$ 的水平下，
价格和广告费都能显著影响馅饼的销量。

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	306.5261933	114.2539	2.682851	0.019932	57.58834426	555.464	57.58834	555.464
价格(\$)	-24.97508952	10.83213	-2.30565	0.039788	-48.5762627	-1.37392	-48.5763	-1.37392
广告费(\$100s)	74.13095749	25.96732	2.854779	0.014494	17.55303206	130.7089	17.55303	130.7089

自变量为分类型变量

- 当自变量为二分类型变量时，可以将自变量直接放进回归模型里面进行分析。
- 当自变量为多分类型变量时，则需要将之处理成 $(n-1)$ 个二分类型变量

自变量为分类型变量

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

假设：

Y = 馅饼销量

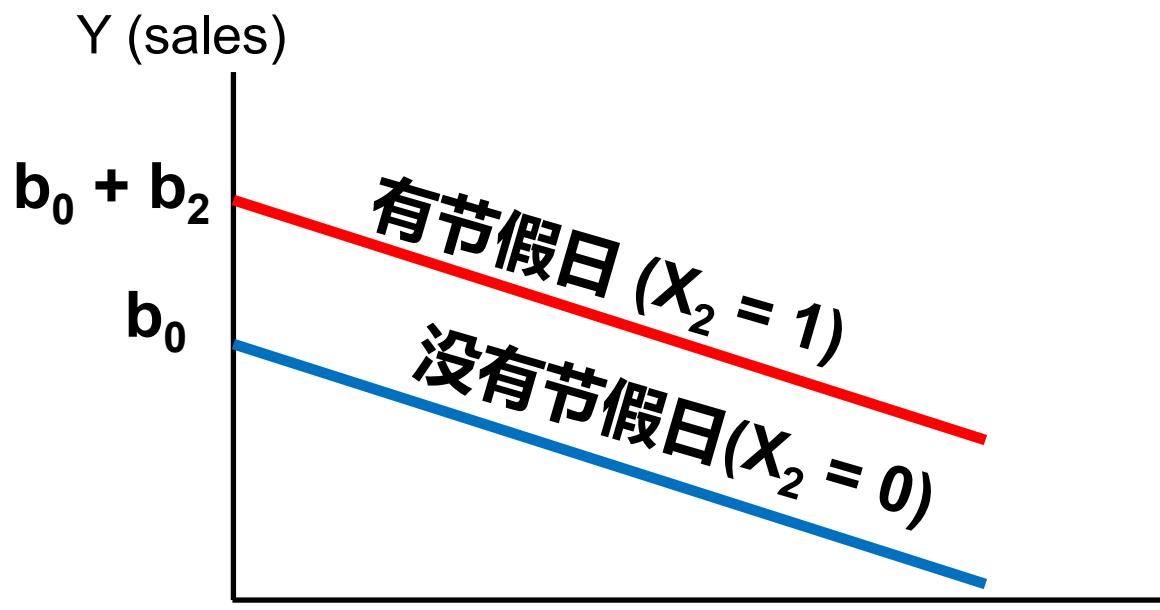
X_1 = 价格

X_2 = 假期 ($X_2 = 1$ 代表当周有节假日)
($X_2 = 0$ 代表当周没有节假日)

自变量为分类型变量

$\hat{Y} = b_0 + b_1 X_1 + b_2 (1) = (b_0 + b_2) + b_1 X_1$ 有节假日
$\hat{Y} = b_0 + b_1 X_1 + b_2 (0) = b_0 + b_1 X_1$ 没有节假日

截距不同 斜率相同



如果 $H_0: \beta_2 = 0$ 被拒绝，那么变量“节假日”对馅饼的销量有显著影响。



自变量为分类型变量

例子：

$$\widehat{\text{销量}} = 300 - 30(\text{价格}) + 15(\text{节假日})$$

销量：每周馅饼销量

价格：馅饼价格（单位：\$）

节假日： $\begin{cases} 1 & \text{当周出现节假日} \\ 0 & \text{当周没有出现节假日} \end{cases}$

$b_2 = 15$: 在价格不变的前提下，如果当周有节假日的馅饼销量比当周没有节假日的销量高15个。



谢谢！