浙江财经大学
Zhejiang University of Finance & Economics

双因素方差分析

朱宗元

浙江财经大学数据科学学院

2017.8.5

双因素方差分析问题提出

提出问题

1 引例1

饮料销售量与包装形式有关，也受价格水平因素的影响。

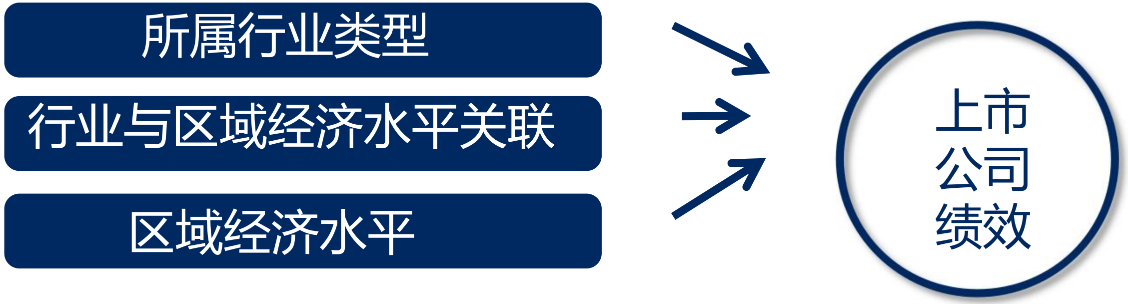
1 概念

方差分析中研究两种因素的影响时，称为双因素方差分析。

交互作用

提出问题

1 引例2



2 交互作用

两个因素的不同水平的搭配可能对试验（调查）观察指标产生新的影响，这种现象称为交互作用。

交互作用

模型分类

- 双因素方差分析
  - 无交互作用
  - 有交互作用

无交互作用的双因素方差分析

提出问题

因素A（r水平）

因素A和B无交互

因素B（s水平）

无交互作用双因素方差分析

无交互作用双因素方差分析

数据结构表

		因素B				A因素各水平之下的均值
		B_1	B_2	…	B_s	
因素A	A_1	x_{11}	x_{12}	…	x_{1s}	$\bar{x}_{1\cdot}$
	A_2	x_{21}	x_{22}	…	x_{2s}	$\bar{x}_{2\cdot}$
	…	…	…	…	…	…
	A_r	x_{r1}	…	…	x_{rs}	$\bar{x}_{r\cdot}$
B因素各水平之下的均值		$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$	…	$\bar{x}_{\cdot s}$	$\bar{x}$

模型形式

$$\left.\begin{array}{l}
X_{ij} = \overline{X} + \alpha_i + \beta_j + e_{ij} \\
e_{ij} \sim N(0, S^2), \text{各 } e_{ij} \text{独立} \\
i = 1,2,\cdots,r;\ j = 1,2,\cdots,s \\
\sum_{j=1}^{r} \alpha_i = 0, \sum_{j=1}^{s} \beta_j = 0
\end{array}\right\}$$

无交互作用双因素方差分析模型

检验形式

无交互作用双因素方差分析

对因素A的检验

$$\begin{cases} H_{01}: \alpha_1 = \cdots = \alpha_r = 0 \\ H_{11}: \alpha_1, \cdots \alpha_r \text{不全为 } 0 \end{cases}$$

对因素B的检验

$$\begin{cases} H_{02}: \beta_1 = \cdots = \beta_s = 0 \\ H_{12}: \beta_1, \cdots \beta_s \text{不全为 } 0 \end{cases}$$

A效应$\alpha_i = \overline{X}_{i\bullet} - \overline{X}$

B效应$\beta_j = \overline{X}_{\bullet j} - \overline{X}$

分析思路

1 计算中间量

总平均值

因素A水平i下均值

因素B水平j下均值

$$\bar{x} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s} x_{ij}, \quad \bar{x}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s} x_{ij}, \quad \bar{x}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} x_{ij}$$

无交互作用双因素方差分析

2平方和分解

恒等关系

$$x_{ij}-\bar{x}=(\bar{x}_{i\cdot}-\bar{x})+(\bar{x}_{\cdot j}-\bar{x})+(x_{ij}+\bar{x}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j})$$

平方求和

$$\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}-\bar{x})^2=\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{i\cdot}-\bar{x})^2+\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{\cdot j}-\bar{x})^2+\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}+\bar{x}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j})^2$$

2平方和分解

符号约定

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}-\bar{x})^2,SS_A=\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{i\cdot}-\bar{x})^2$$

$$SS_B=\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{\cdot j}-\bar{x})^2,SS_E=\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}+\bar{x}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j})^2$$

平方分解

$$SS_T=SS_A+SS_B+SS_E$$

总离差平方和

A偏差平方和

B偏差平方和

误差偏差平方和

无交互作用双因素方差分析

3 构造统计量

$$\frac{SS_T}{S^2} \sim \chi^2(rs-1), \frac{SS_A}{S^2} \sim \chi^2(r-1),$$

$$\frac{SS_B}{S^2} \sim \chi^2(s-1), \frac{SS_E}{S^2} \sim \chi^2((r-1)(s-1))$$

分布理论

$$\overline{S}_A^2 = SS_A/(r-1),$$

$$\overline{S}_B^2 = SS_B/(s-1),$$

$$\overline{S}_E^2 = SS_E/(rs-s-r+1)$$

$$F_A = \overline{S}_A^2 \Big/ \overline{S}_E^2$$

$$F_B = \overline{S}_B^2 \Big/ \overline{S}_E^2$$

统计量

4方差分析表

影响因素	偏差平方和	自由度	均方差	F值	拒绝原假设判断
因素A	SS_A	r-1	$\overline{S}_A^2$	$F_A=\overline{S}_A^2/\overline{S}_E^2$	$F_A>F_\alpha(r-1,rs-r-s+1)$
因素B	SS_B	s-1	$\overline{S}_B^2$	$F_B=\overline{S}_B^2/\overline{S}_E^2$	$F_B>F_\alpha(s-1,rs-r-s+1)$
误差	SS_E	rs-r-s+1	$\overline{S}_E^2$		
总和	SS_T	rs-1			

例1

为了认识客户消费时段的特征，新开业的环山市旋门湾咖啡厅按消费时段统计消费额。每天分为上午、下午和晚上三个时段，每周七天全部营业。假设营业时段与营业日之间不存在交互作用，因此可只统计一周的数据。假设没有季节性的差异，也假设消费群体结构不会产生较大的变动，假设其他因素可忽略。现在的问题是：周内每天的消费额是否存在差异？一天之内三个不同时段之间是否存在显著性差异？分时段的平均消费额见下表。

例1数据

时段		B因素							平均
		周一	周二	周三	周四	周五	周六	周日	
A因素	上午	4152	4852	3546	5456	3426	6124	5846	4771.71
	下午	6852	5112	5786	6105	3998	10124	9789	6823.71
	晚上	9852	8912	9978	9105	15918	16124	10100	11427.00
	平均	6952	6292	6436.67	6888.67	7780.67	10790.7	8578.33	7674.14

无交互作用双因素方差分析

计算均方误差

$$\overline{S_A^2} = SS_A/(r-1) = 80584204$$

天内营业时段（因素A）

$$\overline{S_B^2} = SS_B/(s-1) = 7560880$$

周内营业日（因素B）

$$\overline{S_E^2} = SS_E/(rs-s-r+1) = 4643544$$

无交互作用双因素方差分析

方差分析表

影响因素	偏差平方和	自由度	均方差	F值	0.05临界值
因素A	161168407.12	2	80584203.56	17.354	3.89
因素B	45365282.16	6	7560880	1.6283	3.00
误差	55722525.29	12	4643544		
总和	262256214.57	20	13112811		

无交互作用双因素方差分析

分析结论

5%显著性水平下，因素A是显著的，而因素B并不显著。

一天内不同时段咖啡消费量有显著差异，但一周不同天无明显差异。

有交互作用的双因素方差分析

提出问题

因素A（r水平）

因素A和B有交互

因素B（s水平）

有交互作用双因素方差分析

有交互作用的双因素方差分析

检验形式

$$H_{01}: \alpha_{1\cdot} = \alpha_{2\cdot} = \cdots = \alpha_{r} = 0$$

A因素检验

$$H_{02}: \beta_{1\cdot} = \beta_{2\cdot} = \cdots = \beta_{s} = 0$$

B因素检验

$$H_{03}: \gamma_{ij} = 0 (i = 1,2,\cdots,r; j = 1,2,\cdots,s)$$

交互检验

有交互作用的双因素方差分析

3构造统计量

$$\frac{SS_T}{S^2} \sim \chi^2(rsn-1), \frac{SS_A}{S^2} \sim \chi^2(r-1), \frac{SS_B}{S^2} \sim \chi^2(s-1),$$

$$\frac{SS_{A\times B}}{S^2} \sim \chi^2((r-1)(s-1)), \frac{SS_E}{S^2} \sim \chi^2(rsn-rs)$$

分布理论

$$\overline{S}_A^2 = SS_A/(r-1), \overline{S}_B^2 = SS_B/(s-1),$$

$$\overline{S}_{A\times B}^2 = SS_{A\times B}/(rs-s-r+1)$$

$$\overline{S}_E^2 = SS_E/(rsn-rs).$$

统计检验

$$H_{01}, F_A = \overline{S}_A^2 / \overline{S}_E^2,$$

$$H_{02}, F_B = \overline{S}_B^2 / \overline{S}_E^2,$$

$$H_{03}, F_{A\times B} = \overline{S}_{A\times B}^2 / \overline{S}_E^2.$$

有交互作用的双因素方差分析

方差分析表

影响因素	偏差平方和	自由度	均方差	F值	拒绝原假设的判断
因素A	SS_A	r-1	$\overline{S}_A^2$	$F_A = \overline{S}_A^2 / \overline{S}_E^2$	$F_A > F_\alpha(r-1, rsn-rs)$
因素B	SS_B	s-1	$\overline{S}_B^2$	$F_B = \overline{S}_B^2 / \overline{S}_E^2$	$F_B > F_\alpha(s-1, rsn-rs)$
交互作用	$SS_{A\times B}$	(r-1)(s-1)	$\overline{S}_{A\times B}^2$	$F_{A\times B} = \overline{S}_{A\times B}^2 / \overline{S}_E^2$	$F_{A\times B} > F_\alpha(rs-r-s-1, rsn-rs)$
误差	SS_E	rs(n-1)	$\overline{S}_E^2$		
总和	SS_T	rsn-1			

例2

在 例1中假设了因素A和因素B不存在交互作用。如果取消该假设，即不确定是否存在交互效应。现按消费时段，对开业8周的咖啡消费额进行了统计，数据见下表。请做有交互作用的双因素方差分析。

有交互作用的双因素方差分析

周次		日期						
		周一	周二	周三	周四	周五	周六	周日
上午	1	4151	4852	3546	5456	3426	6124	5846
	…	…	…	…	…	…	…	…
	8	3968	4568	5541	4879	3895	6123	5680
下午	1	6852	5112	5786	6105	3998	10124	9789
	…	…	…	…	…	…	…	…
	8	7581	6124	7001	5261	5097	14121	14589
晚上	1	9852	8912	9978	9105	15918	16124	10100
	…	…	…	…	…	…	…	…
	8	12781	11029	9989	9123	16444	16879	11589

有交互作用的双因素方差分析

（1）计算中间量

时段		B因素							时段总平均
		周一	周二	周三	周四	周五	周六	周日	
A因素	上午	4876.25	5306.88	4079.38	4636.75	5297.50	7911.25	7265.75	5624.82
	下午	6309.38	6785.88	4626.38	6031.50	6180.00	12734.88	12741.75	7915.68
	晚上	10365.50	9975.38	15490.88	10262.1	9454.50	15570.50	10941.13	11722.9
日总平均		7183.71	7356.04	8065.54	6976.79	6977.33	12072.21	10316.21	8421.12

有交互作用的双因素方差分析

(2)方差分析表

方差来源	平方和	自由度	均方差	F值	F检验5%显著性水平临界点
因素A（日时段）	1062668594.333	2	531334297.167	503.787	$F_{0.05}(2,147)=19.5$
因素B（周时段）	573226212.536	6	95537702.089	90.585	$F_{0.05}(6,147)=3.67$
交互效应A×B	399549054.500	12	33295754.542	31.570	$F_{0.05}(12,147)=2.3$
误差	155037898.250	147	1054679.580		
总和	2190481759.619	167			

- 经检验，因素A、因素B及两者的交互作用A×B，对实验（观察）指标都有显著影响。
- 即一天之内的不同时段（上午、下午、晚上）咖啡消费量存在显著差异；周一至周日各天消费也存在显著差异，且一周七天不同时段咖啡消费量差异的规律也不完全相同。

谢 谢

日期：17/08/5