

市场调研与数据分析

“正大杯”第十二届全国市场调查与分析大赛公益培训

2.2 数据建模与分析（二）

第十二届
(Credamo)

数据建模与分析（二）

主要内容

1

交叉表分析

2

卡方分析

3

方差分析



交叉表分析

第十二届市调大赛公益培训课件
(Credamo 见数版权所有)

交叉表分析

交叉表分析又称列联表分析，即由两个以上的变量交叉分类的频数分布表。

- 人们依据两个或两个以上变量的值，通过交叉表将变量分组，计算分组频率，然后比较各组的分布状况，以寻找变量间的相关关系。
- 如性别与手机品牌的关系，年龄与旅游地选择的关系等。

A \ B	B_1	B_2	合计
A_1	n_{11}	n_{12}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$n_{2\cdot}$
A_3	n_{31}	n_{32}	$n_{3\cdot}$
合计	$n_{\cdot 1}$	$n_{\cdot 2}$	n

交叉表常规模式

交叉表分析

- **交叉表的用途：**一是根据收集的样本数据，产生二维或多维交叉列联表；二是在交叉列联表的基础上，对两个变量间是否存在相关性进行检验。交叉表只能表明变量间的**相关关系**，并不能说明变量的**因果关系**。
- 交叉表分析常用于**分类变量**的相关分析之中。



交叉表分析

案例演示

某手机大厂想通过调研了解其顾客的**年龄差异**与**手机产品推荐意愿**之间的是否存在相关关系，以便针对性地进行广告营销。

 年龄

年龄:

- 20岁以下
- 21-30岁
- 31-40岁
- 41-50岁
- 51-60岁
- 60岁以上

 推荐意愿

我将会给我周围的朋友推荐该产品

- 否
- 是

交叉表分析

年龄与推荐意愿

- 通过Credamo的建模分析，生成被试年龄——推荐意愿交叉表；
- 为了探究被试是否受到了不同实验图片刺激而产生差异，接下来对数据进行筛选后分析。

交叉表1

推荐意愿	频次				行总计
	年龄：0-20岁	年龄：21-30岁	年龄：31-40岁	年龄：41-50岁	
是	10	96	41	8	155
否	8	29	5	3	45
	18	125	46	11	200

交叉表分析

年龄与推荐意愿

- 仅通过变量频次较难分析判断年龄同是否推荐之间的关系，将频次替换为行占比后，将能更加直观地看出变量之间的关系。
- 由表可知，只有在31-40岁年龄组的被试推荐的人数比例超过不推荐的人数比例。

交叉表1

推荐意愿	行占比				
	年龄: 0-20岁	年龄: 21-30岁	年龄: 31-40岁	年龄: 41-50岁	行总计
是	6.5%	61.9%	26.5%	5.2%	100%
否	17.8%	64.4%	11.1%	6.7%	100%

2

卡方分析

第十二届市场调查大赛公益培训课件
(Credamo 见数 版权所有)

卡方分析

卡方分析即是通过通过对分类变量之间进行频数分析，利用卡方检验方法（Chi-Squared Test），从而得到变量间相关关系的统计分析方法，常用皮尔森卡方检验法。

- 实质上，卡方分析是表示理论值和实际实际观测值的**吻合程度**或**拟合优度**，卡方值 χ^2 越大，则表明理论值与实际值差异越大；卡方值越小，则二者差异越小；卡方值为0时，表明二者完全一致。
- 在市场运营中，卡方分析通常用于检验某个变量在不同类别中是否存在显著差异，如检验性别差异对保险品类的影响、收入差异对网购平台偏好的影响等。

卡方分析

卡方分析使用要求

- 变量要求：适用于**离散型**的变量。
- 样本要求：卡方分析更适用于**大样本数据**。卡方分析的理论频数不能太小，通常需要有超过五分之四的频数值超过**5**以上的数据。



卡方分析

卡方分析操作步骤

① 计算卡方值

② 确定样本的自由度。自由度即计算某一统计量时，取值不受限制的变量个数。在列联表中，通常自由度 $df = (\text{行变量} - 1) \times (\text{列变量} - 1)$ 。

③ 根据卡方分布的临界值表确定这个自由度对应的不同边界值以及P值。若卡方值小于临界值，则说明没有差异，若卡方值大于临界值，则说明变量之间的影响存在显著差异。

A \ B	B_1	B_2	合计
A_1	n_{11}	n_{12}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$n_{2\cdot}$
A_3	n_{31}	n_{32}	$n_{3\cdot}$
合计	$n_{\cdot 1}$	$n_{\cdot 2}$	n

图中自由度 $df = (3-1) \times (2-1) = 2$

卡方分析

案例演示（一）

目的：研究不同职业消费者在收入水平上是否存在差异。

方法：利用Credamo的卡方分析工具，列分类选择月收入水平，行分类选择职业。



卡方分析

案例演示

- 通过Credamo卡方分析工具直接得出以下分析结果：由图看出， $\chi^2(df=25, n=200)=170.7953$ ， $p=0 \leq 0.05$ 显著。
- 因此可以认为，被试的**职业差异**在收入水平上存在显著差异。

卡方统计值表（括号内为频数百分比）

卡方统计值表 下载数据

变量	水平	月收入						总计	χ^2	p
		0-2000元	2000-4000元	4000-6000元	6000-8000元	8000-10000元	10000元以上			
职业	学生	41(97.62)	14(43.75)	4(11.11)	0(0.0)	0(0.0)	1(3.85)	60(30.0)	170.7953	0
	国有企业	0(0.0)	1(3.12)	4(11.11)	16(39.02)	3(13.04)	5(19.23)	29(14.5)		
	事业单位	1(2.38)	3(9.38)	6(16.67)	5(12.2)	3(13.04)	1(3.85)	19(9.5)		
	公务员	0(0.0)	2(6.25)	4(11.11)	2(4.88)	0(0.0)	1(3.85)	9(4.5)		
	民营企业	0(0.0)	12(37.5)	18(50.0)	16(39.02)	14(60.87)	15(57.69)	75(37.5)		
	外资企业	0(0.0)	0(0.0)	0(0.0)	2(4.88)	3(13.04)	3(11.54)	8(4.0)		
总计		42	32	36	41	23	26	200		

3

方差分析

第十二届市统计大赛公益培训课件
(Credamo 见数 版权所有)

What is ANOVA?

方差分析(Analysis of Variance, 简称ANOVA), 又名“**F检验**”, 是用于两个及两个以上样本均数差别的**显著性检验**。

方差分析

目的

现实中研究目标受到众多因素**相互制约相互影响**，通过方差分析可以梳理出对该事物有**显著影响**的因素，以及各因素之间的交互作用，例如研究不同性别在焦虑程度上是否存在差异、不同类型广告对消费者购买意愿影响是否有差异等。

原理

方差分析利用组内以及组间的偏差平方和与自由度，计算出组内组间的**均方值**从而估计出**F值**，依靠**F-分布**判断变量间的关系显著差异程度。

方差分析

方差分析使用要求

- 变量要求：用于分析**分类变量**与**数值变量**之间的关系（区别于卡方分析）。
- 样本要求：分析**两组或多组数据**之间的组间差异（t检验仅用于检验两组之间的差异）。

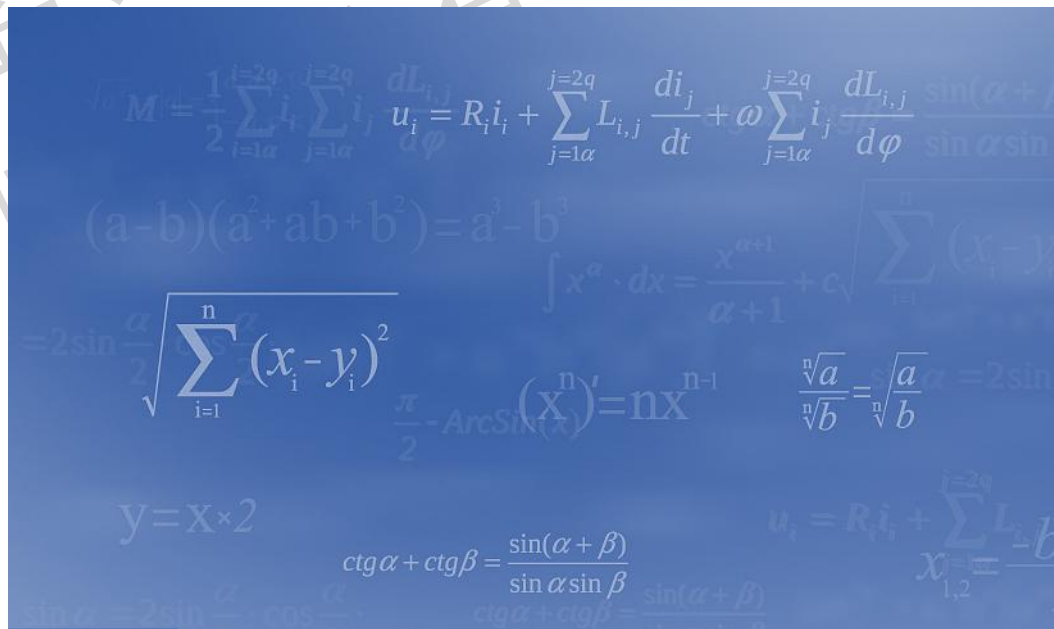
方差分析类型

依照**自变量数量**差异可分为：单因素方差分析、双因素方差分析、多因素方差分析

方差分析

方差分析操作步骤

- ① 计算组内方差和与组间方差（需要计算均值以及自由度）。
- ② 由组间方差与组内方差的比值得出**F值**，根据**F分布的临界值表**确定对应的临界值以及P值。若F值小于对应临界值，则说明没有差异，若F值大于临界值，则说明变量之间的影响存在显著差异。



方差分析

案例演示

目的：利用方差分析，研究在不同性别的消费者在**信息性感知**、**趣味性感知**以及**购买意愿**上是否存在差异。

变量说明：

自变量

性别

男
女

因变量

变量名称	测度项目
信息性	广告中所展示的产品信息是有用的
	广告提供了较为完整的产品信息
	广告中提供的信息是有价值的
趣味性	我认为广告的语言是有趣的
	我认为广告的语言是生动的
	我认为广告的语言是幽默的
购买意愿	我愿意购买该产品
	如果发了一笔奖金，我将购买该产品

方差分析

- 如果需要平衡性别、地域等被试特征，可选择多次发布

编号	发布渠道	计划发布量
1	数据集市	800
2	数据集市	400
3	数据集市	1200
4	数据集市	50

方差分析

购买意愿：男生 vs 女生

- 通过Credamo可以直接生成方差分析结果，可以看出，P值 >0.05 ，说明不同性别消费者的购买意愿无显著差异。

购买意愿均值信息

[下载数据](#)

女

5.3235

男

5.1939

方差分析结果

[下载数据](#)

因素	自由度	平方和	均方和	F值	p值
性别	1	0.8401	0.8401	0.4587	0.4990
Residual	198	362.6399	1.8315		

方差分析

购买意愿：不同年龄

- 由方差分析结果可知，不同年龄的消费者的购买意愿存在显著差异 ($F=3.0692$, $p=0.029 < 0.05$)

购买意愿均值信息

[下载数据](#)

0-20岁 ▾

31-40岁 ▾

21-30岁 ▾

41-50岁 ▾

4.5278

5.6304

5.2440

5.0909

方差分析结果

[下载数据](#)

因素 ▾	自由度 ▾	平方和 ▾	均方和 ▾	F值 ▾	p值 ▾
年龄	3	16.3094	5.4365	3.0692	0.0290
Residual	196	347.1706	1.7713		

谢谢大家

关注右边公众号



及时了解大赛资讯和进程

随时学习大赛公益培训

中国商业统计学会

官方网址：<http://www.china-cssc.org>

公众号：



Credamo见数

官方网址：www.credamo.com

公众号：

