

# 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

# 案例引入



想要知道：

- 哪个类型的广告更受欢迎？
- 不同用户对广告的感受是否存在差异？
- 广告的内容对购买意愿有什么影响？

.....



## 哪些因素会影响用户的购买意愿？

- 哪个类型的广告更受欢迎？
    - 信息性 vs. 情感性
  - 不同用户对广告的感受是否存在差异？
    - 人口统计特征
  - 广告的内容对购买意愿有什么影响？
    - 信息性、说服力、趣味性和对广告的信任度
- .....

# 案例引入

## • 广告效果对比



8200旗舰芯片  
低耗、高效、流畅

徕卡摄像头  
广角、夜景、三摄

4200mAh大电池  
快充、节能、续航

信息型



照亮生活  
让你更出彩!



情感型

# 数据类型

数值型

分类变量

文本

语音

图片

视频

# 分类变量

职业类型：

- 学生
- 国有企业
- 事业单位
- 公务员
- 民营企业
- 外资企业
- 自由职业
- 其他

你的手机里的常用APP包括：

[多选题]

- 微信
- 抖音
- 支付宝
- 淘宝
- 微博
- B站
- 腾讯视频
- 爱奇艺
- 其他

选择你所在城市

省级地区



市级地区



你是否使用过华为手机的鸿蒙系统 (Harmony OS) ?

- 是
- 否



条件：是 已选定 跳至：你觉得鸿蒙系统的使用体验如何？

# 数值变量

该广告是吸引眼球的

非常不同意

1 2

点击编写问题内容

变量名称	测度	项目				
	非常不同意	不同意	不确定	同意	非常同意	
该广告能引起我的注意	陈述 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
非常不同意	陈述 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1 2	陈述 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	陈述 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	陈述 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
我还能回忆起广告中的内容	陈述 6	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
非常不同意	陈述 7	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1 2						

如不及于 七天, 我们默认以 0 分

# 数据建模与分析

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \varepsilon$$

- $Y$ : 因变量、被解释变量
- $X_1 - X_3$ : 自变量、解释变量



## 2.1 数据建模与分析（一）

第十二届  
(Credamo)

# 数据建模与分析（一）

## 主要内容

1

描述性统计

2

相关分析

3

回归分析



## 描述性统计

第十二届市统计大赛公益培训课件  
(Credamo 版权所有)

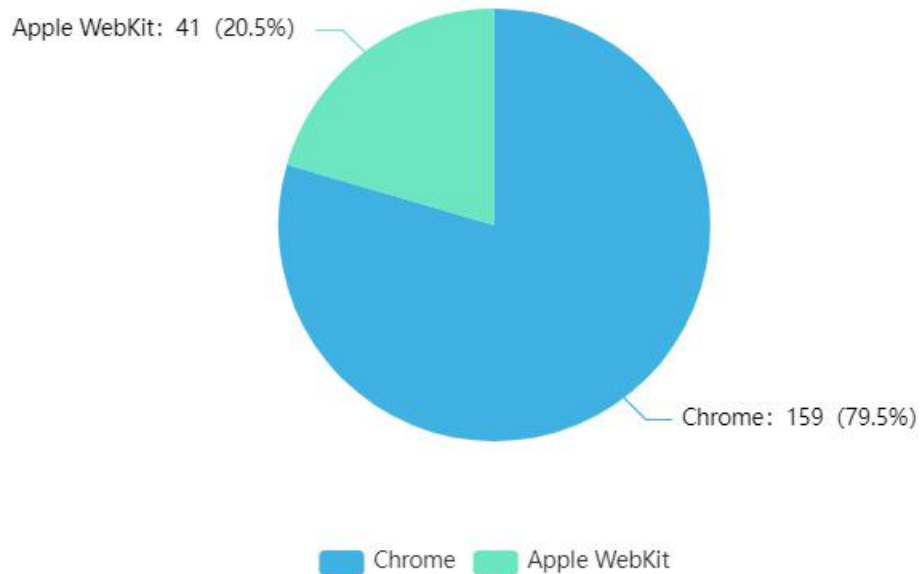
# 描述性统计

- **描述性统计**（descriptive study）指对调查总体的有关数据情况做**统计性描述**，通常用几个关键的指标（**平均值、众数、中位数、标准差、数据分布等**）描述整体的数据情况。
- **描述性统计**常用于频数、占比、集中趋势分析、离散程度分析等，通常以图形（**条形图、饼图、直方图等**）的形式表达分析的数据。



# 描述性统计

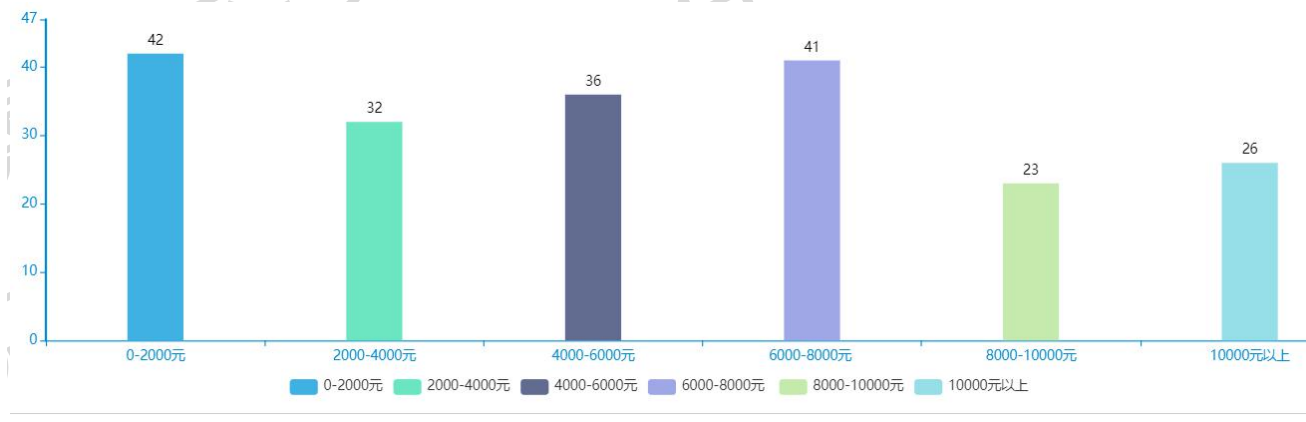
## • 占比分析



被试浏览器类型分布示意图

由图可知，使用Chrome浏览器参与调查的被试占比更多，占总人数的79.5%，而使用Apple Webkit浏览器的人较少。

## • 频数分析



被试收入水平分布示意图

由图可知，在0—2000元的区间内被试分布最多，共有42名被试分布在该区域，本次调研被试的收入水平主要集中在0—8000元的区间内。

# 描述性统计

The screenshot shows the Credamo software interface with the '直播行为调查' (Live Behavior Survey) project. The '建模分析' (Modeling Analysis) tab is active. A modal dialog titled '选择分析模型' (Select Analysis Model) is displayed, offering the following options:

- 描述统计 (Descriptive Statistics):**
  - 描述统计
  - 多选题描述分析
  - 排序题描述分析
  - 矩阵量表描述分析
  - 热力图分析
  - 交叉表
  - 作答统计
- 高级分析 (Advanced Analysis):**
  - 效度分析
  - 卡方分析
  - 相关分析
  - 信度分析
  - 回归分析
  - 方差分析
  - 逻辑回归
  - 中介分析
  - 因子分析
  - 聚类分析
  - 城市热力图
- 文本分析 (Text Analysis):**
  - 词云图
  - 情感分析

Buttons for '取消' (Cancel) and '确定' (Confirm) are located at the bottom right of the dialog.

# 描述性统计

- 在建模分析版块选中**描述统计**分析方法和**目标变量**，即可自动生成数据和图表，包括**数据名称、频次以及百分比**。

添加分析

分析方法: 描述统计 | 分析名称: 描述统计1

选择变量

- 作答ID
- 用户名
- 结束时间
- 作答渠道
- 问卷发布名称
- 经度
- 省份
- 设备类型
- 浏览器类型
- 随机元素
- 年龄
- 月收入
- 海报2
- 广告语\_情感型
- 趣味性\_问题
- 用户ID
- 开始时间
- 作答时长
- 发布ID
- IP
- 纬度
- 城市
- 操作系统类型
- 屏幕分辨率
- 性别
- 职业
- 海报1
- 广告语\_信息型
- 感受表达
- 趣味性1

类别	频次	百分数
1	42	21.00%
4	41	20.50%
3	36	18.00%
2	32	16.00%
6	26	13.00%
5	23	11.50%

月收入

类别	频次	百分数
1	42	21.00%
4	41	20.50%
3	36	18.00%
2	32	16.00%
6	26	13.00%
5	23	11.50%

## 2

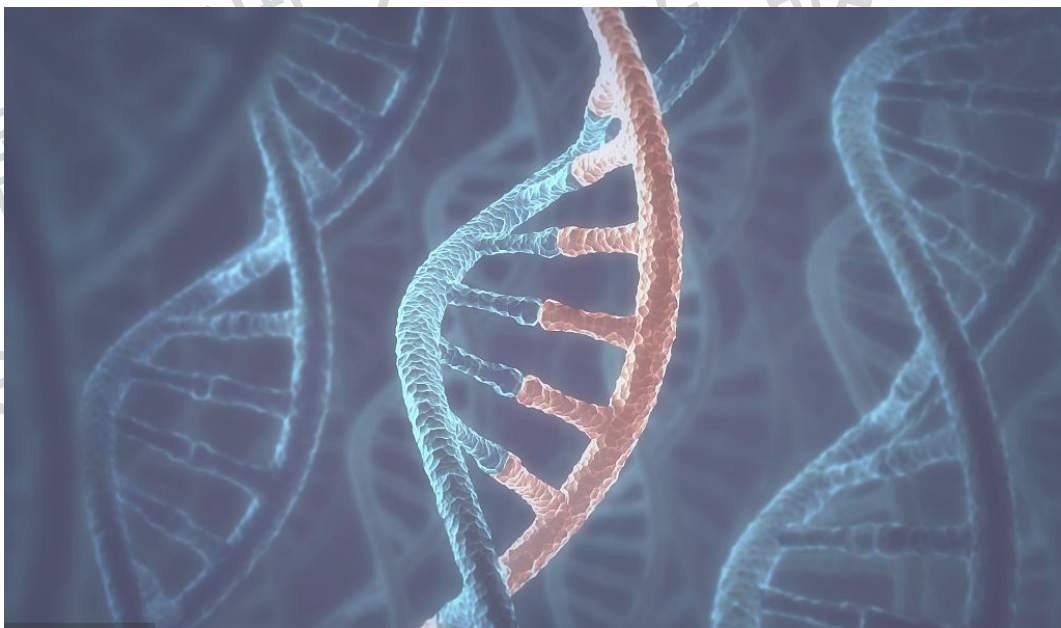
## 相关分析

第十二届市桥大赛公益培训课件  
(Credamo 见数 版权所有)



# 相关分析

- 相关分析（correlation analysis）是研究现象之间是否存在某种依存关系，并分析关系的强弱程度的一种统计方法。
- 相关分析只能反映数值型变量之间的关系，如身高与体重的相关性、客户满意度与产品销量的相关性等；



# 相关分析

- 通过相关分析能得到反映变量相关关系的**相关系数**。相关系数常见有三类，分别是：Pearson相关系数、Spearman等级相关系数和Kendall相关系数，常见为**Pearson（皮尔森）相关系数**。
- 皮尔森相关系数是反应变量之间**关系密切程度**的指标，它的取值区间在1到-1之间。1表示两个变量完全线性相关，-1表示两个变量完全负相关，0表示两个变量不相关。



# 相关分析

- 除了相关系数，相关分析中还需要关注判断相关关系的显著程度的指标——P值（P Value），P值能够检验假设关系的成果，P值越小，则说明变量间相关关系越可靠。
- 通常来说， $P \leq 0.05$  视为关系显著的指标，如果P值不显著，相关系数再高也无法证明其关系相关关系。



# 相关分析

## 案例演示

- 目的：探究手机广告的信息性、趣味性、信任度、说服力以及购买意愿之间的相关关系。
- 方法：采用七级量表题对变量进行测量，让被试在1—7分的区间内打分，1分为非常不同意，7分为非常同意。

变量名称	测度	项目
信息性	1	广告中所展示的产品信息是有用的
	2	广告提供了较为完整的产品信息
	3	广告中提供的信息是有价值的
趣味性	1	我认为广告的语言是有趣的
	2	我认为广告的语言是生动的
	3	我认为广告的语言是幽默的
信任度	1	我相信广告中的信息是准确的
	2	我相信广告中的信息是真实的
	3	我信任广告中的内容
说服力	1	这则广告的内容是真诚的
	2	该广告对我来说是有说服力的
	3	我被广告的内容所打动
购买意愿	1	我愿意购买该产品
	2	如果发了一笔奖金，我将购买该产品

# 相关分析

## 案例演示

### 变量说明

将相关测度项目进行以下处理得到变量：

$$\text{信息性} = (\text{信息性1} + \text{信息性2} + \text{信息性3}) / 3$$

$$\text{说服力} = (\text{说服力1} + \text{说服力2} + \text{说服力3}) / 3$$

$$\text{信任度} = (\text{信任度1} + \text{信任度2} + \text{信任度3}) / 3$$

$$\text{趣味性} = (\text{趣味性1} + \text{趣味性2} + \text{趣味性3}) / 3$$

$$\text{购买意愿} = (\text{购买意愿1} + \text{购买意愿2}) / 2$$

分析方法

相关分析

修改

分析名称

相关分析1

分析配置

相关分析仅支持数值变量

选择变量

作答时长

趣味性2

信息性1

信息性3

态度2

信任度1

信任度3

说服力2

购买意愿1

趣味性

态度

说服力

趣味性1

趣味性3

信息性2

态度1

态度3

信任度2

说服力1

说服力3

购买意愿2

信息性

信任度

购买意愿

# 相关分析

- 先判断变量之间的显著关系。由图可知，趣味性与信息性相关关系并不显著，趣味性和信任度相关关系较为显著，其余变量相关关系均非常显著。
- 再判断关系之间的正负影响。由图可知，信息性、说服力、信任度、趣味性以及购买意愿之间均为正相关关系。
- 最后判断关系之间的强弱程度。由图可知，信息性、说服力、信任度和购买意愿之间关系紧密，说服力与信任度的相关关系最紧密。

相关系数矩阵

	趣味性		信息性		信任度		说服力		购买意愿	
	相关系数	p值	相关系数	p值	相关系数	p值	相关系数	p值	相关系数	p值
趣味性	1									
信息性	0.0368	0.6053	1							
信任度	0.1363	0.0543	0.7247**	0	1					
说服力	0.3283**	0	0.6304**	0	0.8208**	0	1			
购买意愿	0.4022**	0	0.5556**	0	0.6559**	0	0.8129**	0	1	

\*  $p < 0.05$  \*\*  $p < 0.01$  在剔除两变量缺失值后计算Pearson相关系数

## 3

## 回归分析

第十二届市赛公益培训课件  
(Credamo 见数版权所有)

# 回归分析

- **回归分析 (Regression Analysis)** 是在分析自变量和因变量之间相关关系的基础上，建立变量之间的**回归方程**，并将回归方程用于预测的统计方法。
- **分类**：根据相关关系中变量的个数差异分类，可以分为一元回归分析与多元回归分析；根据自变量与因变量的关系类型分类，可以分为线性回归分析与非线性回归分析。
- **与相关分析的区别**：回归分析侧重于研究随机变量间的依赖关系，侧重于对**因变量Y**的解释，而相关分析主要关注变量间线性相关的程度；回归分析不仅可以解释变量X对变量Y的影响大小，还可以由回归方程进行**预测**和控制。
- **应用**：吸烟、饮食、锻炼与人体寿命之间的关系；产品质量、顾客体验、收入水平等与顾客购买意愿的关系。



# 回归分析

## 常用回归分析法

- **一元线性回归分析：**线性回归方程中只涉及到两个变量的，称一元线性回归分析。即仅用单个自变量X去预测单个因变量Y。

$$Y=b_0+b_1X_1+\varepsilon$$

- **多元线性回归分析：**线性回归方程中涉及到两个以上变量的，称多元线性回归分析。即用多个自变量X去预测单个因变量。

$$Y=b_0+b_1X_1+b_2X_2+b_3X_3+\varepsilon$$

# 案例：哪些因素会影响用户的购买意愿？



- 广告的内容对购买意愿有什么影响？
  - 信息性、说服力、趣味性和对广告的信任度
- 广告类型？
  - 信息性 vs. 情感性
- 不同用户对广告的感受是否存在差异？
  - 人口统计特征

.....

# 回归分析

## 案例演示

目的：预测在手机广告中，广告的信息性、说服力、信任度和趣味性对购买意愿的影响。

方法：利用Credamo的建模工具，在回归分析中因变量选中购买意愿，自变量选中信息性、说服力、信任度和趣味性，系统即可自动生成多元回归模型。



# 回归分析

## 案例演示—结果分析（一）

- 首先可以关注模型的解释效力，重点关注**R方**指标。R方为模型的拟合优度，R方数值越接近1（一般认为0.5以上即可），说明模型的拟合程度与解释力越强。
- 其次，为了避免模型过度拟合，可以关注模型的**AIC值**与**BIC值**两个指标。建构回归模型时往往需要多次的回归分析，选取拥有更低的AIC值与BIC值的模型，有助于防止模型复杂度过高，缺乏普适性。

### 数据描述

[下载数据](#)

样本数	R方	调整后的R方	F值	AIC值	BIC值
200	0.6903	0.684	108.6623	462.6242	479.1158

# 回归分析

## 案例演示一结果分析（二）

实训课件

参数摘要

下载数据

参数名称	系数	标准误	t值	P值	[0.025	0.975]
截距项	0.0471	0.3063	0.1537	0.8780	-0.5571	0.6513
趣味性	0.1621	0.0419	3.8705	0.0001	0.0795	0.2447
信息性	0.1384	0.0588	2.3525	0.0196	0.0224	0.2543
信任度	-0.0648	0.0983	-0.6585	0.5110	-0.2587	0.1292
说服力	0.7810	0.0834	9.3595	0.0000	0.6164	0.9456

- 对于模型中变量间的关系影响，可以关注参数P值与参数系数，P值说明变量关系的解释力度，系数说明变量关系的正负影响；P值小于0.05即可认为关系显著。

# 回归分析

## 案例演示—结果分析（三）

- 在Credamo见数上进行的建模分析可以自动生成结果解读；
- 由下图可知，该模型有一定的解释力度，说明手机广告的信息性、说服力和趣味性正面影响消费者的手机购买意愿，而广告的信任度对购买意愿的影响不显著。

### 指标解读

#### 💡 整体解读：

- 1.本次建模将**购买意愿**作为因变量，将[**"趣味性"**, **"信息性"**, **"信任度"**, **"说服力"**]作为自变量进行线性回归分析；
- 2.从数据概述表中可以看出，模型R方为0.6903，表明该模型可解释因变量变异的69.03%；
- 3.通过对模型整体的F检验 ( $F=108.6623$ ,  $p=0$ )，证明自变量中至少存在一项会对因变量产生影响。

#### 💡 系数解读：

在控制其他因素不变时，可以得出以下结论：

对于趣味性这一变量，趣味性每增加一单位，购买意愿平均 增加 0.1621单位

对于信息性这一变量，信息性每增加一单位，购买意愿平均 增加 0.1384单位

对于信任度这一变量，信任度每增加一单位，购买意愿平均 减少 0.0648单位

对于说服力这一变量，说服力每增加一单位，购买意愿平均 增加 0.781单位

# 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](http://www.credamo.com)

公众号：

