

用数据解读市场

——市场调查的数据分析与案例

黄京华 著

3.52

中国广播电视出版社

用数据解读市场

——市场调查的数据分析与案例

黄京华 著

中国广播电视出版社

图书在版编目(CIP)数据

用数据解读市场:市场调查的数据分析与案例/黄京华
著. —北京:中国广播电视出版社,1998. 6

ISBN 7-5043-3179-1

I. 用… I. 黄… III. 市场-调查-数据分析 IV. F713.52

中国版本图书馆 CIP 数据核字(98)第 15342 号

书名:用数据解读市场——市场调查的数据分析与案例

作 者:	黄京华
责任编辑:	张兆晋
出版发行:	中国广播电视出版社
社 址:	北京复外真武庙二条 9 号(邮政编码 100866)
经 销:	全国各地新华书店
印 刷:	河北省高碑店市印刷厂
开 本:	850×1168 32 开
字 数:	160 千字
印 张:	5.125
版 次:	1998 年 6 月第 1 版 1998 年 6 月第 1 次印刷
印 数:	2043 册
书 号:	ISBN 7-5043-3179-1/G · 1199
定 价:	10.00 元

(版权所有 翻印必究·印装有误 负责调换)

前 言

计算机的普及,网络大潮的兴起,使人类社会进入了全新的信息时代。信息将影响社会的发展,改变价值观念,决定人们的生存。

随着我国社会主义市场经济的建立和完善,数据信息对于市场营销的影响越来越显著,以至成为企业决策的核心问题。因此,对于当今在中国市场运作的企业而言,了解市场,占有信息,是经营者正确决策的关键。企业经营者只有获得了数据的支持,只有用获得的数据解读了市场,才能谈得上把握现在,预测未来,科学决策。

市场信息的来源是多方面的,用市场调查的方法可以获得最直接的有关市场信息的数据。如何用数据解读市场,如何从有限的数据中得到更多的市场信息,又是了解市场,占有市场的关键。本书系统地介绍了如何利用市场调查的数据,了解市场、分析市场、预测市场的方法。书中使用了北京广播学院广告系下属的IMI市场信息研究所,1997年在北京、上海、广州、重庆四城市所做的《IMI消费行为与生活形态抽样调查》得到的最新数据。通过案例的分析,介绍了数据分析的主要方法。案例分析包括市场分析和消费者分析。本书通过分析案例,深入浅出地介绍了描述性统计方法、交互分析、方差分析、因子分析等统计方法,详尽地叙述了数据的分析过程和结果,以启发众多的市场预测人员、企业管理人员、广告商和广大消费者,用多种方法分析研究数据,用数据解读市场,从中获取更多的市场信息。即使缺少统计知识的读者,也会从本书的市场实例分析的结果中受益。

作者

1998年4月

目 录

一、概述

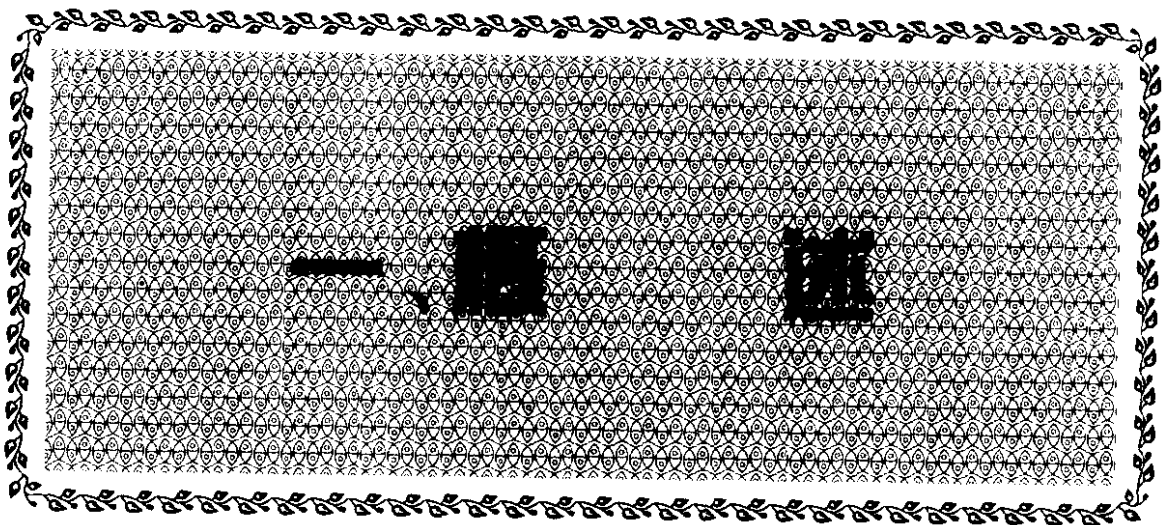
1. 数据分析 (3)
 - 1-1 市场调查的数据分析 (3)
 - 1-2 数据分析的工作流程 (3)
 - 1-3 数据分析工作流程图 (15)
2. 本书所使用的数据及软件 (16)

二、用数据描述市场

3. 各种类型的录/放像机在场所占份额是多少 (19)
 - 3-1 您家最近购买的录/放像机是哪一种类型的 (19)
 - 3-2 频数表 (19)
 - 3-3 图形表示 (21)
4. 真正的好产品不需要做广告吗 (23)
 - 4-1 真正的好产品不需要做广告吗 (23)
 - 4-2 累计百分比 (23)
 - 4-3 图形表示 (24)
5. 品牌排名 (25)
6. 北京地区家庭主要用浓缩洗衣粉 (27)
 - 6-1 您家洗衣粉的主要类型是什么 (27)
 - 6-2 多项选择频数表 (27)
 - 6-3 图表形式 (28)
7. 一次旅游大约花多少钱 (30)
 - 7-1 您最近一次旅游的大约花费是多少 (30)
 - 7-2 分组数据的频数表 (30)
 - 7-3 图表形式 (33)

8. 用统计量描述数据	(34)
8-1 集中趋势的测度	(34)
8-2 离散程度的测度	(34)
8-3 用 SPSS 程序计算统计量	(37)
8-4 分组数据的统计量	(42)
8-5 标准化数据	(43)
9. 女性比男性更爱逛百货公司/购物中心吗	(47)
9-1 您最近三个月有没有去百货公司/购物中心买东西	(47)
9-2 列联表	(47)
9-3 图表形式	(50)
10. 四城市家庭使用洗衣粉类型比较	(51)
10-1 多项选择列联表	(51)
10-2 图表形式	(53)
三、用数据分析市场	
11. 假设检验	(57)
11-1 假设检验的概念	(57)
11-2 假设检验的一般方法	(57)
11-3 关于假设检验的说明	(59)
12. 女性比男性更爱逛百货公司/购物中心	(61)
12-1 交互分析与卡方检验	(61)
12-2 性别与最近三个月去没去过百货公司/购物中心的 交互分析	(64)
13. 已婚男人最不爱逛百货公司/购物中心	(67)
13-1 带控制变量的交互分析	(67)
13-2 图表形式	(71)
14. 消费者认同广告吗	(74)
14-1 累加的李克量表	(74)
14-2 单项与总和相关效度分析	(81)
14-3 皮尔逊相关系数	(81)
14-4 总体相关系数的检验	(85)

14-5 消费者认同广告吗	(85)
15. 爱逛百货公司/购物中心的人更认同广告	(87)
15-1 两种人对广告认同度的比较	(87)
15-2 t 检验	(89)
15-3 用 t 检验的方法分析两种人对广告的认同度	(90)
16. 逛百货公司/购物中心的频度与对广告的认同 有没有关系	(94)
16-1 四种人对广告认同度的比较	(94)
16-2 单因素方差分析	(95)
16-3 用单因素方差分析的方法分析四种人对广告 的认同度	(96)
17. 四城市广告观念的比较	(100)
17-1 四城广告观念的均值比较	(100)
17-2 因子分析	(102)
17-3 广告观念的因子分析	(106)
17-4 用因子得分比较四城市的广告观念	(122)
18. 四城市消费行为与观念的比较	(130)
18-1 关于消费行为的一些观点	(130)
18-2 消费行为与观念的因子分析	(131)
18-3 用因子得分比较四城市的消费行为与观念	(137)
附录	(141)
参考文献	(153)



1. 数据分析

1-1 市场调查的数据分析

作为一个广告人,规划完整的广告活动的一项最基础的工作就是要做市场调查,研究市场、研究消费者。完成这项工作可以有很多途径和方法,比如对现有资料的研究,用小组访谈的方式,用问卷调查的方法,或者将多种方法综合使用,以达到目的。采用什么样的方法,要根据研究的问题和目的以及可能的人力、财力等条件来决定。其中问卷调查的方法是一种较常用的市场调查方法,是一种定性分析和定量分析结合使用的方法。这里的定量分析就是把问卷数字化,对问卷所对应的数据进行分析。如果在设计调查方案时使用随机抽样的方法获得调查样本,即调查的是随机样本,那么可以利用统计分析的方法对数据进行相关性分析,从而揭示事物间的相互影响和相互作用。

1-2 数据分析的工作流程

调查问卷回收之后,就要开始做数据分析的准备工作及数据分析工作,整个的作业主要包括以下五方面:

- 检查问卷;
- 建立数据库;
- 修改变量或新建变量;
- 对数据进行统计分析;

●撰写数据分析报告。

下面分别对这五个方面做简要说明。

检查问卷

检查所回收的问卷,要确认:(1)问卷是根据受访者的真实情况填写的;(2)题目间的选项结果不能相互矛盾;(3)漏答的题目不能太多,特别是主要的人口统计学项目,如性别、年龄等不能漏答,等等。对于有问题的问卷,可以对被访者再次访问,进行修正,否则就只能舍弃不用。

建立数据库

建立数据库分以下几个步骤:

(1)建立数据库结构

对问卷的每一问题,设定为相应的一个或几个变量,每个变量赋予一个变量名,每份问卷的问卷号也要做为一个变量赋予变量名。所有变量不能重名。考虑变量取值的位数及小数位等,建立数据库结构。

(2)确定变量的可能取值

一般在设计问卷时应该已经考虑到每一题目选项的量化问题,此时要再确认一下。比如年龄变量,变量值可以是实际年龄,如果受访者年龄范围是20岁至60岁,变量的可能取值是20至60间的任何数值;也可以将实际年龄分为几段,如果年龄分段如下:①20~29岁;②30~39岁;③40~49岁;④50~60岁,那么变量的可能取值就是1,2,3,4。特别是对有些开放性的问题,比如对这样的问题:“您最常用的牙膏品牌是_____。”这时要给问卷涉及到的所有品牌编号:

- | | |
|--------|---------|
| 1. 中华 | 16. 康齿灵 |
| 2. 洁银 | 17. 白玉 |
| 3. 高露洁 | 18. 奥琪 |

- | | |
|----------|-------------------|
| 4. 洁诺 | 19. 天然 |
| 5. 蓝天 | 20. 洁齿健 |
| 6. 蓝天六必治 | 21. 美晨 |
| 7. 两面针 | 22. 冷酸灵 |
| 8. 佳洁士 | 23. 芳草 |
| 9. 黑妹 | 24. 黑人 |
| 10. 洁灵 | 25. 黑白 |
| 11. 上海防酸 | 26. 富乐 |
| 12. 美加净 | 27. 生物素 |
| 13. 上海特效 | 28. 安利 |
| 14. 皓清 | 29. Sunrider 仙尼雷德 |
| 15. 草珊瑚 | 30. 小白兔 |

此时该问题对应的变量所有可能的取值就是 1, 2, 3, 4, ……

30. 每一个数值代表一个牙膏品牌。

(3) 输入数据

按照所建立的数据库结构,对每份问卷依次输入对应的变量值。所有问卷输入完毕即得到一个数据库文件。数据库的一行数据即是一份问卷的量化形式,一列数据或几列数据是所有问卷在某变量上的取值。

对于问卷中的某些问题,会有一些被访者拒答或无法填答。对于没有填答某变量的个案,我们称个案在该变量上的值是缺失值。对于缺失值,一般用一个不会与原变量的所有可能取值混淆的数值表示。例如“收入”变量常常会出现拒答的现象,如果所有个案的最高收入是 5000 元,则缺失值可用“9999”来表示。如果用“0”来表示,就会与“无收入”的情况混淆。

(4) 检查数据

数据输入完毕,要对数据做仔细的检查。由于输入时的疏忽或者误操作,可能会出现错误的数值。例如年龄是 29 岁却输成 92

问卷号: _____
城市号: _____

1. 被访对象性别记录 (由访员完成)	1 男	2 女
2. 请问您的年龄: _____ 岁		

1. 请问您阅读报纸的习惯:

1 从来不看 2 很少看 3 经常看 4 大约一天看一次 5 天天看

2. 请问您经常阅读什么报纸? (填写三个) 它们的主要来源如何?

报纸名:	家中 订阅 1	单位 订阅 2	报摊 购买 3	赠阅 4	借阅 5	街头 报栏 6
1 _____						
2 _____						
3 _____						

3. 请问您平时看报纸时比较关心哪方面的报道?... (最多选 5 项)

1 新闻报道	9 休闲旅游	17 散文
2 人物专访	10 文化信息	18 评论
3 热点追踪	11 健康医疗	19 广告
4 影视娱乐报道	12 生活常识	20 漫画
5 电视节目预告	13 读者来信	21 小说
6 海外见闻	14 体育	22 经济类
7 消费指南	15 科技	23 其它
8 服饰美容	16 军事	

1. 下面关于广告的一些观点您的态度如何? 请您按真实想法填答。

	比较		比较	
	同意	无所谓	不同意	不同意
1. 购买商品时, 还是以有广告的厂牌比较可靠.....	5	4	3	2
2. 广告是可以信赖的.....	5	4	3	2
3. 许多产品的广告都很落俗套.....	5	4	3	2
4. 专家推荐的产品应该不会出错.....	5	4	3	2
5. 当电视出现广告时, 我通常会换台.....	5	4	3	2
6. 真正的好产品不需要作广告.....	5	4	3	2
7. 广告格调低的产品, 我不会去买.....	5	4	3	2
8. 广告是生活中必不可少的.....	5	4	3	2

图 1.1

岁。如果样本年龄要求要在 20 岁至 60 岁之间,那么通过检查,这种错误很容易被发现并修改。但是如果把 42 岁输成了 24 岁,就不太好办了。所以在输入数据时要尽可能地认真仔细。

【例 1.1(a)】《1997—1998 IMI 消费行为与生活形态年鉴》部分调查问卷及对应数据。该问卷(部分)如图 1.1 所示。

问卷中“问卷号”的变量名是“NO”,共调查了 2400 份问卷,所以其取值是从 1 到 2400。“城市号”的变量名是“CITY”,共调查了广州、北京、上海、重庆四个城市,城市代号依次是 0,1,2,3,变量“CITY”的取值是 0 到 3。“性别”变量名是“I1”,取值如问卷所示,是 1 或 2。“年龄”变量名是“I2”,由于规定调查对象年龄范围是 16 到 60 岁,所以其取值是 16 到 60 岁,这里考虑到这样取值分析数据时会更灵活,所以没有用分段数据。以上四个变量不允许有缺失值。

阅读报纸习惯这道题要用两个变量表示,第一个变量 G31 表示阅读频次,取值是 1,2,3,4;第二个变量 G31F 是当阅读频次是经常看时,大约多少天看一次的天数,这个变量的取值没有限定。经常阅读的报纸,由于是限选三个,所以要用三个变量 G321、G322 和 G323 来表示。它们的取值是把问卷涉及到的所有报纸名称从 1 开始编号,这些编号就是它们的所有可能取值,比如这次调查问卷广州地区的报纸编号是从 1 到 72,这三个变量的取值范围就是 1 到 72。与这三个变量相对应的变量 G321F、G322F 和 G323F 分别代表三种报纸的来源,其取值如问卷中所列,是 1 到 6。平时看报纸所关心的报道要求最多选五项,所以要用五个变量 G33A、G33B、G33C、G33D 和 G33E 表示,它们的取值是 1 到 23。

对于广告的八种说法的态度,用八个变量 H101、H102、H103、H104、H105、H106、H107 和 H108 表示,取值范围都是 1 到 5。

例 1.1(b)就是这些变量按问卷号(NO)排序后前 20 个个案的数据。其中第一列数据是数据库的记录号(RECORD#)。

【例 1.1(b)】前 20 个个案的数据(见图 1.2)。

Record#	N0	CITY	I1	I2	G31	G31F	G321	G321F	G322	G322F	G323	G323F	G33A	G33B
1	1	0	1	41	1	0	0	0	0	0	0	0	0	0
2	2	0	1	18	4	0	2	3	52	3	36	3	4	5
3	3	0	2	38	3	3	2	2	1	2	46	2	1	2
4	4	0	2	24	4	0	1	1	2	3	23	4	1	10
5	5	0	1	45	4	0	1	3	0	0	0	0	1	2
6	6	0	2	18	2	0	2	1	1	1	0	0	1	9
7	7	0	1	25	4	0	2	3	1	3	16	3	1	9
8	8	0	2	28	4	0	2	1	3	3	29	3	1	5
9	9	0	1	41	4	0	1	3	0	0	0	0	1	12
10	10	0	1	44	4	0	1	1	0	0	0	0	1	3
11	11	0	1	20	4	0	2	1	1	1	0	0	9	10
12	12	0	2	30	2	0	1	3	2	3	0	0	11	12
13	13	0	1	18	4	0	1	3	2	3	26	3	1	3
14	14	0	1	52	4	0	1	1	2	1	0	0	1	11
15	15	0	2	17	3	7	2	5	29	1	49	3	1	3
16	16	0	2	36	3	0	2	2	1	1	52	3	1	2
17	17	0	1	49	3	0	2	3	1	3	56	3	1	11
18	18	0	1	23	4	0	2	3	1	3	29	3	1	11
19	19	0	1	38	4	0	2	3	1	3	0	0	1	12
20	20	0	2	26	4	0	2	3	1	3	0	0	1	8

Record#	G33C	G33D	G33E	H101	H102	H103	H104	H105	H106	H107	H108
1	0	0	0	4	4	3	3	4	4	3	2
2	6	15	20	3	1	1	1	5	3	1	3
3	7	11	0	4	4	4	3	4	4	3	4
4	12	0	0	3	2	1	3	2	4	4	1
5	6	11	14	1	1	1	1	3	0	5	4
6	0	0	0	3	2	3	3	3	1	4	4
7	20	14	22	3	3	4	2	5	3	3	4
8	12	21	0	5	3	5	3	3	3	0	3
9	13	22	0	4	2	3	1	1	2	5	4
10	0	0	0	4	4	1	3	1	5	1	3
11	11	12	21	3	3	3	4	5	5	5	5
12	20	21	0	4	3	3	3	4	4	2	4
13	6	7	10	3	3	3	3	3	3	3	5
14	14	15	22	3	1	3	3	3	5	4	4
15	4	13	17	3	3	4	4	2	2	3	4
16	3	8	9	2	3	2	3	3	4	1	5
17	14	23	0	3	3	4	4	3	4	4	5
18	22	0	0	3	3	4	3	3	3	3	5
19	14	16	0	4	3	2	1	3	5	5	5
20	22	0	0	4	4	3	3	3	2	3	5

图 1.2

修改变量或建立新变量

在建立数据库文件后,就可以考虑对数据进行统计分析了。在分析之前,有时要修改变量或新建变量。这主要有以下几种情况:

(1)定义缺失值,修改变量

SPSS 的统计程序对缺失值有固定的处理方法。比如对“婚姻状况”这个变量,“1”代表已婚,“2”代表“未婚”,“3”代表“离异”,“4”代表“丧偶”,“99”代表“未答此题”。如果直接对此变量做频数分析,分析结果见例 1.2(该例为《1997—1998IMI 消费行为与生活形态调查》中上海地区的数据)。

【例 1.2】“婚姻状况”变量 I5A 的频数表(见图 1.3)。

I5A				
Value	Frequency	Percent	Valid Percent	Cum Percent
1.00	399	66.5	66.5	66.5
2.00	179	29.8	29.8	96.3
3.00	8	1.3	1.3	97.7
4.00	6	1.0	1.0	98.7
99.00	8	1.3	1.3	100.0

Total	600	100.0	100.0	
Valid cases		600	Missing cases	0

图 1.3

此时有效个案是 600,8 个未答此题的个案参加有效百分比的计算。从表中数据可以看到有 399 人是已婚的,占了总人数(600 人)的 66.5%。

如果将“99”定义为缺失值,就会得到如例 1.3 所示的频数表。

【例 1.3】定义缺失值后的“婚姻状况”变量 NEWI5A 的频数表(见图 1.4)。