



第3章 显著性检验

本章主要内容:

- 认识统计显著性的实质
- 理解假设建立的概念及如何检验假设
- 理解第一类误差和第二类误差的区别
- 描述几种常见的统计检验方法
- 方差分析

3.1 评估差分和变化

测量方法是否有区别是营销管理人员关心的众多问题的核心。下面是几个具体例子:

- 我们对认知度所做的事后测试的结果高于事先测试的记录。是认知度真的提高了, 还是另有别的解释? 我们是该终止与代理商的关系还是继续委托它?
- 我们的整体顾客满意度得分从3个月前的92%上升到今天的93.5%, 顾客满意度真的提高了吗? 我们是否应该庆祝一番?
- 采用10点制量表, 达拉斯顾客对我们的有线电视系统提供的客户满意程度比辛辛那提的平均要高1.2。达拉斯的顾客顾客真的更为满意么? 辛辛那提的顾客服务负责人员是否该被撤换? 达拉斯的负责人是否该受到奖励?
- 在最近的一次产品创意调查中, 19.8%的被调查者说他们非常可能购买他们评估的新产品。这是否是好事? 是否比我们去年对类似产品所做的调查结果要好? 这种结果对我们是否推出新产品的决策有何意义?
- 在一项市场细分调查中, 我们发现年收入30000美圆以上的人去快餐店的次数平均每月为6.2次, 而年收入为30000美元或以下的人去的次数为6.7次, 这种差别是否真实? 是否有意义?
- 在一项认知度调查中, 28.3%的被调查者在无提示的情况下说知道我们的产品。这是一个好的结果么?

以上这些是在营销调研中永远存在的问题。尽管被一些人认为枯燥, 但它们却正是统计假设检验为什么重要的原因。假设检验有助于更接近这些问题的最终答案。我们说接近, 是因为在市场调研中很少能完全确定地回答这些问题。

3.2 统计显著性

统计推断最根本的目的是从抽样调查的结果中归纳出总体的特征。统计推断的基本信条是, 在数学意义上不同的数字在统计学意义上可能并没有显著的不同。例如, 调查人员要求喝可乐的人蒙上眼睛品尝两种不同的可乐并说出自己更喜欢哪一种。结果表明, 51%的人倾向于被试验的一种, 49%的人倾向于另一种。这里有一个数学上的差别, 但这种差别极小且并不重要。它可能在我们准确判断自己口味偏好能力的误差范围之内, 这种差别在统计意义上可能并不显著。关于差分有三个不同的概念:

- 数学差分 根据定义, 如果几个数字不完全相同, 它们就有差分。然而, 这并不能说明差分的重要性及在统计上的显著性。
- 统计显著性 如果一个特殊的差分大到不可能是由于偶然因素或抽样误差引起的地步, 那么这个差分在统计意义上是显著的
- 管理意义上的重要差分 如如果或数字的差异程度从管理角度看是有意义的, 那么我们可以说这个差分是重要的。例如, 在顾客对两种不同包装的反应的调查中, 其差分在统计上也许很显著, 但可能小到没有实际及管理上的重要性。



在本章，我们将介绍几种检验结果是否具有统计显著性的方法。

3.3 假设检验

假设可以定义为一个管理者或管理者对被调查者总体特征所做的一种假定或猜想。营销人员常常要面临这样的问题，即调查结果是否与标准有很大的差别，以便决定公司营销策略的某些方面是否需要改变。让我们看看下面几种情形：

- 一项跟踪调查的结果表明，顾客对产品的了解程度比 6 个月前所做的类似调查中显示的要低。结果是否明显降低？是否低到需要改变广告策略的程度？
- 一位经理认为其产品购买者平均年龄为 35 岁。为检验其假设，他进行了一项调查，调查表明购买者平均年龄为 36.5 岁。调查结果与其观点的差别是否足以说明此经理的观点是不正确的？
- 一家快餐店的营销部长认为她的顾客中 60% 为女性，40% 为男性。为此她进行了一项调查，通过调查她发现顾客 55% 为女性，45% 为男性。调查结果与她原来假设的差别是否足以让她得出“她原来的假设是错误的”这一结论？

所有这些问题都可以通过一定的统计检验来评估。在假设检验中，调查者测定一个关于总体特征的假设是否有根据。如果假设确实正确，统计假设检验便可以让我们计算出一个特殊结果的概率。

对于一项具体调查结果与假设值之间的差分有两种主要的解释：假设是正确的，差分很可能是因为抽样的错误造成的；或者假设很可能是错误的，真正的数值是另外的值。

假设检验的步骤

检验一个假设主要有五个步骤。首先必须被明确。其次，选择适当的统计方法来检验假设。第三，判断标准必须明确，并作为决定是否拒绝或不能拒绝（FTR）原假设 H_0 的基础。请注意，我们不说“拒绝 H_0 或接受 H_0 ”，尽管这看似很细微的区别，但却很重要。对这一区别我们在后文中进一步论述。第四，计算检验统计值并进行检验。第五，从初始问题或调查问题的角度陈述结论。

1 陈述假设

假设主要用两种形式：原假设 H_0 与备择假设 H_1 ，原假设 H_0 与备择假设 H_1 在检验时是相对的。例如：调研部经理认为他的操作程序将使顾客在汽车购物窗前的等待时间为 2 分钟，1000 个样本的平均时间为 2.4 分钟，这时，原假设与备择假设将表示如下：

原假设 H_0 ：标准等待时间=2 分钟

备择假设 H_1 ：标准等待时间 \neq 2 分钟

2 选择适当的检验统计量

本章将介绍几种检验方法，分析人员必须根据调查情况选择合适的统计检验方法。

3 确定判定规则

从前面关于样本均值的讨论中可以发现，抽样结果与总体参数完全相等的情况几乎是不可能发生的。关键的问题是，如果统计假设是正确的，实际样本平均数和假设平均数间的差分是否会在一个可以接受的范围。因此，需要一个判断规则或标准，来决定是否拒绝或不拒绝原假设。统计上用显著水平来说明判定规则。

显著水平 (α) 在选择原假设和备择假设的过程中是很关键的。显著水平是判定原假设的可接受性时的一种认为很低的概率，如 0.10, 0.05 或 0.01。

如果我们决定一项显著水平为 0.05 的假设，这表示如果由于巧合或抽样误差，检验表明观察结果发生的概率为 5%，那么我们将拒绝原假设。对原假设的拒绝相等于认同备择假设。

4 计算统计检验值

在这一步骤中，要求：



- 运用适当的公式计算统计检验值
- 比较计算值与根据判定规则所得的严格的统计值（从适当的统计表格中查得）
- 通过比较，得出是否拒绝原假设 H_0 的结论

5 表述结果

从初始研究问题的角度表述你的结论，以总结检验结果。

3.4 假设检验中的两类错误

假设检验中容易犯两种错误类型，一般称之为第一类错误和第二类错误。第一类错误发生在如下情形：调查者拒绝了原假设，而实际上它是正确的。调查得出的这种不正确的结论也许是由于样本与总体值的差分是由于抽样误差造成的。调查者必须决定在多大程度上认同第一类错误。认同第一类错误的可能性被认为是 α 水平。相反地，如果在它事实上正确的时候，我们没有拒绝它，那么 $1-\alpha$ 将是得出正确结论的概率。

第二类错误发生在如下情形：原假设错误，而调查者没有拒绝它。第二类错误被认为是 β 错误， $1-\beta$ 的值，反映了在原假设错误的时候，作出正确决定拒绝它的概率。这些不同的概率总结于表 3.1 中

表 3.1 第一类错误和第二类错误

原假设的实际状况	不拒绝 H_0	拒绝 H_0
H_0 正确	正确 ($1-\alpha$) 没有错误	第一类错误 α
H_0 不正确	第二类错误 β	正确 ($1-\beta$) 没有错误

然而，对 β 的估计更为复杂，且已超出我们讨论的范围。并且，第一类错误与第二类错误并不是互补的，即 $\alpha+\beta\neq 1$ 。

对于任何假设检验，我们都希望能控制 n （样本容量）、 α （第一类错误）及 β （第二类错误）。遗憾的是，这三个量中我们只能控制两个。对于一个规定了样本量的问题， n 是固定的，因此， α 与 β 只能控制一个。

3.5 假设检验方法

本章将介绍最常用的 χ^2 检验和 t 检验

3.5.1 t 检验

(1) t 分布的背景

从均数为 μ 、标准差为 σ 的正态分布总体中，随机抽取一个含量为 n 的样本，样本均数为 \bar{x} ，

理论证明： \bar{x} 服从均数为 μ 、标准差为 $\frac{\sigma}{\sqrt{n}}$ 的正态分布，对其进行标准化，表达式 $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ 服从

标准正态分布，由于 σ 未知，用样本标准差 S 代替，表达式 $\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$ 所服从的分布即为 t 分布。

(2) t 分布的特点

t 分布类似于标准正态分布，以 0 为中心，左右两侧对称，中间面积大，两侧面积小，象一个‘钟型’曲线，曲线下的面积为 1 ，理论证明，当样本含量趋向无穷时， t 分布即为标准正态分布。

对任意给定的概率 α ，总对应一个点，在这个点的左侧的面积为 $\alpha/2$ ，这个点，在统计学上称



为界点，常用 $t_{\alpha(n-1)}$ 表示，即从 $-t_{\alpha(n-1)}$ 至 $t_{\alpha(n-1)}$ ，包含了总体 $1-\alpha$ 的个体， α 常取值 0.05、0.01。

(3) t 检验的基本思想

有一个服从正态分布的总体，研究者想知道总体均数是否为 μ_0 ，于是从中随机抽取一个样本，若总体均数为 μ_0 ，不妨取 $\alpha=0.05$ ，则 $t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$ 服从 t 分布，t 值应位于 $-t_{\alpha(n-1)} \sim t_{\alpha(n-1)}$ ，

否则，有理由怀疑总体均数为 μ_0 。

(4) t 检验的基本步骤

第一步 建立假设并确定检验水准

假设有两种：1) 检验假设 H_0 (hypothesis to be tested)

$$H_0: \mu = \mu_0$$

2) 备择假设 H_1 (alternative hypothesis)

$$H_1: \mu \neq \mu_0$$

常取 $\alpha=0.05$ 或 $\alpha=0.01$

第二步 选定统计方法，计算统计量

$$t = \frac{|\bar{x} - \mu_0|}{\frac{S}{\sqrt{n}}}$$

第三步 确定 P 值和作出推断结论

如 $t > t_{\alpha(n-1)}$ ，则 $P < \alpha$ ，认为 H_0 不成立，否则，不拒绝 H_0 。

(5) 实例

某媒体报道：2000 年广州居民家庭平均总收入为 3.6 万元，报道是否正确？某调查公司随机调查了 20 个家庭，数据如下（单位：万元）：2.3、3.8、4.5、3.3、3.5、5.6、4.3、6.2、2.8、5.2、3.2、3.8、4.8、3.2、3.9、4.6、4.5、5.2、3.8、5.5

经计算： $\bar{x} = 4.2$ ， $S = 1.02$ ， $n = 20$

$$t = \frac{|\bar{x} - \mu_0|}{\frac{S}{\sqrt{n}}} = \frac{|3.6 - 4.2|}{\frac{1.02}{\sqrt{20}}} = 2.63$$

查 t 界值表 $t_{0.05(20-1)} = 2.093$ ， $t > t_{0.05(20-1)}$ ， $P < 0.05$ ，可以认为，广州市居民的年平均总收入不是 3.6 万元。

(6) t 检验的基本类型

t 检验常用于以下 3 种情况



(a) 样本均数与总体均数比较的 t 检验

(b) 两样本均数比较的 t 检验

检验假设 $H_0: \mu_1 = \mu_2$, 备择假设 $H_1: \mu_1 \neq \mu_2$

统计量为:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} \quad \text{自由度 } v = n_1 + n_2 - 2$$

$$\text{其中, } s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

(c) 配对设计的 t 检验

3.5.2 χ^2 检验

基本思想

举例说明, 调研部经理对广州、北京两城市居民进行了电话访问, 样本量分别为 110、105。广州 100 名被访者中, 60 名使用该公司洗头水, 40 名使用其他厂家产品; 北京 100 被访者总 52 名使用公司洗发水, 48 名使用其他厂家产品, 调研部经理需要判断, 该公司在两城市的洗发水市场的占有率是否相同。现将资料整理成如下形式:

表 3.2 甲、乙两地某品牌洗发水使用情况

地区	使用某公司产品	使用其他公司产品	合计
广州	$A_{11} (=65)$	$A_{12} (=45)$	$A_{1t} (=110)$
北京	$A_{21} (=56)$	$A_{22} (=49)$	$A_{2t} (=105)$
合计	$A_{t1} (=121)$	$A_{t2} (=94)$	$n (=215)$

表中, A_{11} 、 A_{12} 、 A_{21} 、 A_{22} 这四个格子的数据是整个表的基本数据, 其余数据都是从这四个基本数据推算而来, 因而, 称该表为四格表。我们想了解, 广州、北京该公司的占有率是否相同? 经计算, 广州样本率为 $P_1 = 59.1\%$; 北京为 $P_2 = 46.7\%$, 广州、北京两地的样本率不同, 这个不同是由两地总体率不同造成的? 还是仅由抽样误差造成的? 下面进行假设检验

检验步骤为:

$H_0: \pi_1 = \pi_2$, 即两地该公司占有率相同

$H_1: \pi_1 \neq \pi_2$, 即两地占有率不相同

显著水平 α 取 0.05, 由相应表格查得, χ^2 界值为 3.84

统计方法, 选用 χ^2 检验, 公式为:

$$\chi^2 = \sum \frac{(A - T)^2}{T} = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(A_{ij} - T_{ij})^2}{T_{ij}}$$

其中 A_{ij} 为实际数, T_{ij} 为理论数, 计算公式为 $T_{ij} = A_{it} \times A_{tj} / n$



本例计算如下： $T_{11}=121 \times 110 / 215 = 61.9$ ，同理可计算 $T_{12}=110-61.5=48.1$ ， $T_{21}=59.1$ ， $T_{22}=45.9$ ，代入公式，可得：

$$\chi^2 = \frac{(65-61.9)^2}{61.9} + \frac{(45-48.1)^2}{48.1} + \frac{(56-59.1)^2}{59.1} + \frac{(49-45.9)^2}{45.9} = 0.72$$

由相应表格查得， $\alpha=0.05$ 时 χ^2 界值为 3.84，由此可得出 $P < 0.05$

结论： P_1 与 P_2 的差分按 0.05 的显著水平属于抽样误差，不拒绝 H_0 ，认为该公司在广州、北京的占有率相同。

3.6 方差分析

方差分析(Analysis of Variance, 简称 ANOVA)是一种对多个样本均数进行分析比较的统计方法，其基本原理是：指标的总变异又类别间的变异与随机变异组成，通过分析总变异的构成，推断各类别总体均数是否相同。

若需要检验两个或两个以上独立样本平均数的差异，方差分析(ANOVA)是一种合适的统计工具。虽然它可以用来检验两个平均数的差异，但是它更主要地是用于对 3 个或 3 个以上独立群体的平均数差异的假设检验。可以用方差分析确定样本之间平均数的差异是否由抽样误差而引致的。上面论及的 t 检验可用于只涉及两个样本平均数时的检验假设，当有 3 个或 3 个以上样本时，利用这两种方法缺乏效率。例如，如有 5 个样本及相应的平均，那么就需 10 次 t 检验来检测每一对平均数。值得注意的是，当有 3 个或更多个平均数出现时，使用 Z 检验或 t 检验，也增加了出现每类错误的可能性。每两个不同平均数的组合都必须被检验。检验的次数越多，夸大一对(或更多对)平均数真正由于抽样误差引起显著差异的机会就越大。若显著性水平为 0.05，平均在 20 次检验中就会出现一次。

单向方差分析经常被用来分析实验结果。例如，一家汽车配件连锁店的市场总经理正在考虑采用三种可能的服务来促销：车轮校正、换油、调试。她想知道这三种方法带来的潜在销售量是存在显著差异。

在 3 个城市中共随机出 60 家连锁店(每个城市选 20 家)，每个城市分别增加一项服务，实验期间的价格和广告等保持一致。实验共进行 30 天，其间的销售情况将逐一记录。

表 15-6 快步示了每家店的平均销售额。问题是，引起这些差异的主要原因是促销方式还是纯属偶然？

3.6.1 基本统计指标

(1) 总变异

统计量是总的离均差平方和，用 $SS_{\text{总}}$ 表示，假定共有 k 个类别，每类分别有 n_i ($i=1, 2, \dots, k$) 个个体，公式为：

$$SS_{\text{总}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{i,j}^2 - n\bar{x}^2$$

$$\text{其中，} \bar{x} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{i,j} / n, \text{ 自由度 } v = n-1$$

(2) 组间变异

$$\text{组间变异为：} SS_{\text{组间}} = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

$$\text{组间均方变异为：} MS_{\text{组间}} = SS_{\text{组间}} / v_{\text{组间}}$$



其中, $\bar{x}_i = \sum_{j=1}^{n_i} x_{i,j} / n_i$ ($i=1, 2, \dots, k$), 自由度 $\nu_{\text{组间}}=n-1$

(3) 组内变异

组内变异为: $SS_{\text{组内}} = SS_{\text{总}} - SS_{\text{组间}}$

组间均方变异为: $MS_{\text{组内}} = SS_{\text{组内}} / \nu_{\text{组内}}$

其中, 自由度 $\nu_{\text{组内}}=n-k$

方差分析的基本思想是: 当各类别总体均数相同时, $MS_{\text{组间}}$ 间的差异 $MS_{\text{组内}}$ 由随机误差造成, $MS_{\text{组间}}$ 与 $MS_{\text{组内}}$ 的比值应接近 1, 当其比值大于 1、并超出一定的范围时, 可以认为各类别间的总体均数不全相等、存在差异。

3.6.2 统计量

(1) 整体均数比较

对各类别均数整体方面的检验, 统计量为:

$$F = \frac{MS_{\text{组间}}}{MS_{\text{组内}}}$$

由分子自由度为 $\nu_1 = \nu_{\text{组间}}$ 、分母 $\nu_2 = \nu_{\text{组内}}$ 、显著水平 α , 查方差分析专用 F 界值表, 得界值 $F_{\alpha}(\nu_1, \nu_2)$, 统计量 F 与界值 $F_{\alpha}(\nu_1, \nu_2)$ 比较, 若 $F > F_{\alpha}(\nu_1, \nu_2)$, $P < 0.05$, 可以认为不同类别的总体均数不全相同, 需进一步对两两间的情况进行检验, 否则, $P > 0.05$, 认为各类总体均数相同。

(2) 两两比较

市场研究中, 多个样本均数间的两两比较 (称为多重比较, multiple comparison), 常采用 q 检验 (Newman-Keuls method, SPSS 中用 S-N-K 法), 统计量为:

$$q = (\bar{x}_A - \bar{x}_B) / \sqrt{\frac{MS_{\text{组内}}}{2} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}$$

q 界值除与显著水平 α 有关外, 还与分子、分母自由度有关, 分子自由度等于 A 类与 B 类间的类别数 a 与分母自由度为误差自由度 $\nu_{\text{组内}}$ 。分子自由度 a 的计算方法是, 将每类样本均数按从小到大的顺序排列, $a = A \text{ 类的序号} - B \text{ 类的序号} + 1$ 。

3.6.3 实例

为评估某行业职工收入情况, 调查三地该行业职工各 6 人, 结果见下表。

表 1 三地某行业 18 名职工年收入表 (万元)

序号	甲地	乙地	丙地
1	1.21	1.63	1.18
2	1.53	1.37	1.08
3	1.36	1.55	1.32
4	1.15	1.62	1.27
5	1.22	1.85	0.97
6	1.25	1.32	1.06

(1) 统计描述

sample	Number	Mean	Std.
1	6	1.287	0.138
2	6	1.557	0.193
3	6	1.147	0.134
Total	18	1.330	0.229



(2) F 检验

	SS	df	MS	F	P
Between Groups	0.521	2	0.261	10.542	0.0014
Within Groups	0.371	15	0.025		
Total	0.892	17			

界值 $F_{0.05(2,15)}=3.68$, $P<0.05$, 可以认为三地平均工资有差异。

(3) q 检验

$$q(1,2) = (1.287 - 1.557) / \sqrt{\frac{0.025}{2} \left(\frac{1}{6} + \frac{1}{6}\right)} = -4.2$$

$$q(1,3) = (1.287 - 1.147) / \sqrt{\frac{0.025}{2} \left(\frac{1}{6} + \frac{1}{6}\right)} = 2.17$$

$$q(2,3) = (1.557 - 1.147) / \sqrt{\frac{0.025}{2} \left(\frac{1}{6} + \frac{1}{6}\right)} = 6.36$$

按均数大小排列, (1) 类别 3, (2) 类别 1, (3) 类别 2, 所以, $a(1,2)=2$, $a(1,3)=2$, $a(2,3)=3$ 显著水平取 0.05, 查 q 界值表, $q_{0.05(2,15)}=3.01$, $q_{0.05(3,15)}=3.67$, 经比较, 可以认为甲、丙地区间无差别, 乙与甲、乙与丙间有差异。