

食品研究常用的试验设计与数据分析方法

刘魁英 湛江海洋大学农学院 524088

摘要 正确的试验方法可使食品研究得到正确的试验结果,而正确的统计分析可排除试验假象,增加试验的可靠性。但目前我国食品研究采用的试验设计与分析方法与国际同类研究差距较大,与现代飞速发展的计算机技术不相适应,因此介绍食品研究常用的试验方法非常必要,本文简要介绍了食品研究常用的试验设计和分析方法以及在分析过程中应注意的问题。

关键词 食品 研究 分析 方法

1 食品研究常用的试验方法

食品研究的重要表现形式是新产品、新工艺的研制与开发,把琳琅满目的食品提供给社会,以满足不同层次、不同需求、不同用途的人们对食品多样化、合理化的要求。从事食品生产的企业在激烈的市场竞争中要靠新产品、新工艺的研制与开发谋求企业的生存和发展,因而重视食品研究,用现代化的科学技术不断提高产品的科技含量,增加产品的附加值来谋求企业的发展。然而,食品研究与其他学科一样,实际上是一个数据的收集、整理、分析过程。翻阅国内外有关食品的期刊、杂志,可以明显的发现国内与国外食品研究在试验设计与数据分析方面存在的差距。因此将食品研究中常用的研究方法介绍给从事食品科研人员是非常必要的。

1.1 食品质量研究的线性研究与非线性研究

传统的观念认为食品只要符合技术标准和卫生标准就是质量好的产品,但从新的质量概念分析,食品质量不仅包括产品符合企业标准和卫生标准,而且要有不断采用新技术、新工艺来更新品种的能力,以不断提高产品质量。只有这样,才能生产出营养价值高、物美价廉、安全可靠的食品来。从食品的形成过程来看,食品质量不仅取决于生产质量,而且与设计质量及保存质量密切相关。从食品的特性分析,除了食品的一般性能外,还包括食品的货架寿命、营养价值、安全性和经济性等指标。质量高的食品必须具有良好的设计质量和制造质量。

食品质量研究包括线性质量研究和非线性质量研究,非线性质量研究是指新产品和新工艺设计的质量研究方法,非线性质量研究方法是在新产品开发过程中,紧密的把专业知识和统计分析结合起来,在保证食品质量的前提下,充分利用各种设计参数与食品特性的非线性关系,通过系统设计、参数设计和允许误差设计的三段优化设计方法,从设计上控制食品的输出特性和质量波动,或出于经济考虑,在不压缩原材料质量波动的情况下,仍然保证食品特性的一种稳定性优化设计方法。线性质量研究是指食品制造过程中的质量研究方法,是通过对生产工序的合理诊断、调节改善与检查,使生产工序的质量达到效果好、费用低的目的。

1.2 系统设计、参数设计和允许误差设计

系统设计即传统设计,是依靠专业技术进行设计。它是专业人员根据市场所需产品的性能、质量、价格情况决定采用何种原料、何种加工设备、怎样的加工工艺等。系统设计的质量完全取决于专业技术人员的技术高低。但结构复杂,特别是多参数、多特征值的食品,要全面考察各种参数组合的交互效

应,单凭专业技术进行定性的判断是很不够的,因为它无法定量地找出经济合理的最佳组合参数。系统设计是整个设计的基础,通过它我们可以选择需要考察的因素和水平。

参数设计是一种非线性设计,它是在系统设计的基础上,运用线性反应试验、面体反应试验、回归正交试验、均匀试验和混料试验以及相配套的统计分析方法来研究各种参数与食品质量特性的非线性关系,以便找出特征值波动最小的参数组合。所以参数设计又称为参数组合的中心之设计。

允许误差设计是系统设计和参数设计完成了产品或工艺最佳组合参数的选择,确定了参数组合的中心值以后,研究如何在保证产品质量的前提下,确定原材料公差(即允许误差)的设计方法。

1.3 对比试验设计

是食品研究中最简单的试验方法,它以差异对比的原则,设置简单的处理和对照试验,通过多次重复估计误差的方法比较和分析影响产品质量的因素,常常应用于新工艺、新技术的引进、工艺诊断等。按照传统的设计观念,对比试验设计属于顺序排列的一种,容易产生系统误差,特别是间比法和互比法排列,没有正确的误差估计,往往通过增加重复次数来提高试验的精确性。

对比试验包括共轭对比和交叉对比两种,共轭对比的特点是有一共轭轴,处理和对照可以横向比较也可以纵向比较,而交叉对比是在共轭对比试验的基础上,待数据调查完毕并试验效应失效以后,将处理与对照交叉对换进行第二次处理,处理安排对照,对照安排处理,分析是讲两次资料合并分析,以消除试验材料本身的差异。

1.4 随机试验设计

对比试验只研究两个处理的试验设计,当处理数目大于两个以上时,多采用随机试验。随机试验是严格按照设置重复、随机排列和局部控制的原则进行试验设计的,所谓局部控制是把要比较的一组处理控制在相对一致的时间或空间内以降低试验误差的方法,设置重复是为了降低试验误差,随机排列是为了无偏的估计试验误差,通过抽签、随机数字表和随机函数等有效的随机化方法,消除人为因素的干扰,尽最大努力克服人为地影响,使误差分布趋于正态,保证试验的精确性和准确性,以便通过方差分析进行统计推断。随机试验一般安排处理10个以下,最多不超过15个,处理过多会影响试验的精确性,重复次数4-7次。随机试验设计包括完全随机试验设计、随机区组试验设计、拉丁方试验设计和裂区试验设计。

1.5 线型反应试验设计

对比试验和随机试验是研究同群同类变数或不同群同类变数资料的分析方法, 而线型反应试验是研究试验效应随因素水平线性变化而呈线型反应的试验。即同群两类变数资料的分析方法, 因此, 研究的因素水平必须是连续可分的。线型反应试验设计包括等差设计、等比设计和0.618法的试验设计。

等差设计是指试验因素水平间隔是等差的, 一般适用于试验效应与因素水平呈直线相关的试验。试验因素水平间隔等差是每个水平间隔的差距是相等的, 如温度可采用30℃、40℃、50℃、60℃和70℃等水平, 它们的间隔都是10℃, 水平的数目一般不能小于3个, 最好包括对照采用5个水平点, 由于对照的试验效应等于0, 可作为坐标的原点, 其它水平可以看出试验效应随因素水平线性变化而呈现线型趋势。如果其中一点由于差错发生偏移, 可在坐标上清晰的表现出来, 以采取必要的措施。

等比设计是指试验因素水平的间隔是等比的, 一般适用于试验效应与因素水平呈对数或指数相关的试验。如果试验效应随因素水平呈对数效应, 时间因素的水平可选用5min、10min、20min和40min; 如果试验效应随因素水平的变化呈指数效应, 添加剂因素水平可选用1000mg/kg、1500mg/kg、1750mg/kg和1875mg/kg, 使试验效应变化率大的地方因素水平间隔排列的小一点, 而试验效应变化率大的地方因素水平间隔排列的大一点。其因素的水平数目应与等差间隔的试验设计一样, 因为试验效应的对数变化在对数坐标上也表现为直线关系, 而试验效应与因素水平呈指数效应在指数坐标上也表现为直线关系。一般包括对照选用5个水平, 以便反映试验效应的线性规律。

0.618法的试验设计适用于试验效应与因素水平呈二次曲线的线性反应试验设计。这种试验的目的在于从试验中寻求到一个最佳的极值试验点。例如食品中加入的糖精数量过多口味发苦, 加入的数量太少达不到甜度, 只有在合适的浓度时, 口味最佳; 这种试验的目的在于找到最佳的温度、最佳的浓度、最佳的时间和最佳的pH值等。试验因素水平间隔的0.618法排列, 一般要求在资料中或预备试验中预先知道或了解试验因素水平的大致变化范围, 即因素水平的上限与下限, 而且知道在这个范围内只有一个试验效应的极值点。它是以试验因素水平的上限与下限为两个端点, 以上限与下限之差与0.618的乘积为水平间隔从两端向中间展开的。例如山楂果冻中加入0.5%~4%的琼脂可达其硬度, 我们选用0.5%和4%为两个端点, 再以 $4 - 0.5 = 3.5$ 与0.618的乘积2.163为水平间隔从两端向中间扩展为 $0.5 + 2.16 = 2.7$ 和 $4 - 2.16 = 1.8$, 这样包括对照有0%、0.5%, 1.8%、2.7%和4%五个析因点。

1.6 面体反应试验设计

在食品研究中, 许多试验属于参数设计, 其目的在于通过试验找到最佳试验效应的因素水平点, 即参数。如最佳的温度、最佳的时期、最佳的使用浓度、最佳的pH等。但对于复因素来说, 由于因素之间的交互作用, 造成复杂的试验效应面, 因而试验设计不象线型反应那样简单, 例如, 食品中的含糖量与含酸量以什么样的组合配置才能够得到最佳的味觉指标? 为了解决类似的问题, 可以先研究其中一个因素, 得出结论后再

去研究第二个因素, 但由于交互作用的影响, 一个因素、一个因素的研究往往找不到最佳试验效应的处理。设计面体反应试验的目的, 多数在于寻求试验效应面的参数点对应下的因素水平。一般情况下, 很难准确的确定不同因素间水平的处理, 往往需要做一些预备试验或其他资料提供信息来确定试验的方向和因素水平的大致范围。面体反应的预备试验, 可根据专业知识采用线型反应设计的方法确定每个因素的水平, 实际上一个面体反应的试验是两个线型反应试验的组合。如果预测试验的效应面为平面反应, 因素的水平应采用等差间隔排列的试验设计。

1.7 回归正交试验设计

食品研究具有多因素的综合影响, 试验效应常常包括因素的主效应, 也包括因素间的交互作用, 因此, 试验设计者总希望安排足够多的研究因素以使试验效应有充分的试验论据。但因素和水平的增加造成试验规模庞大, 特别是对于多指标分析的试验往往由于分析困难而无法实施。线型反应试验一般是研究一个因素多水平的试验设计, 面体反应试验设计是研究两个因素多水平的试验设计, 当试验因素超过3个的多水平试验时, 由于采用组合处理, 处理数目等于因素水平间的乘积, 是随因素的增加呈几何级数增加。例如, 一个3因素5水平的试验, 共有 $5^3 = 125$ 个试验处理, 由于处理数目太大, 不仅增加了试验误差, 而且由于受试材和条件的限制, 这对食品研究来说有时是难以实施的。因此正交试验在食品工艺改革、新产品的试制中得到了广泛的应用, 虽然正交试验能够利用较少的处理安排较多的试验因素, 但难以表达试验效应与因素水平之间的线性规律, 试验效应与交互作用也常常出现混杂现象而难以对试验效应和交互作用做出正确的估计, 多元回归分析虽然可以对试验效应与交互作用做出正确的评价, 但计算比较复杂, 而回归正交试验设计把回归分析与正交试验结合起来, 这就是回归正交试验设计。

回归正交试验设计, 不仅把试验处理的数目大为简化, 而且把回归分析与试验设计有机的结合在一起。一次回归正交设计只能研究试验效应与因素水平的线性规律, 而在食品研究开发和质量管理过程中往往是研究试验效应与因素水平的非线性规律, 特别是研究试验效应与因素水平的二次规律, 利用试验效应与因素水平的二次曲线存在极值的特性, 确定生产工艺的最佳参数, 因此二次回归正交设计在参数优化设计中已得到广泛的应用。二次回归正交旋转组合设计不仅处理少、与回归分析结合、有正确的误差估计, 而且有一个共同的预测预报方差。

1.8 混料回归试验设计

所谓混料, 系指产品原料按一定成分百分比混合起来加工而成, 例如果冻、果茶、糕点、香肠、混配葡萄酒、饮料、巧克力等, 其产品质量指标仅与各因子成份的百分比有关, 而与混料总量无关。这种情况称为混料。在混料设计中, 各自变量的变化范围是受一定的约束的, 即各因子成分的百分比加起来必须等于1(即100%)。在这种情况下研究各因子成分比例与质量指标之间的关系, 建立回归方程, 进而通过探索效应面来估计整个多分量的系统性质, 这就是混料回归设计。因此, 混料设计是建立质量指标对于混料系统中各因子成分的回归方程, 利用回归方程探索效应面, 用来估计最佳混料比。

但是由于混料试验设计属于饱和设计,也就是说试验的处理数目与回归系数个数相等,那么剩余自由度为0,因而不能对回归方程进行显著性测验。

1.9 均匀试验设计

食品研究具有多因素的综合影响,新产品的研制开发总希望能够安排足够多的影响因素以使试验效应有足够充分的试验论据。但因素和水平的增加造成试验规模庞大,特别是对于多指标分析的试验往往由于分析困难而无法实施,应用二次回归正交旋转组合设计在食品工艺改革、新产品的研制、确定生产工艺的最佳参数,不仅可以使试验处理的数目大为简化,而且把回归分析与试验设计有机的结合在一起,还能够进行误差分析,是一种非常好的设计方法。但是,在许多实际问题中,不仅要考虑较多的因素,而且要求水平数目多一些,当每个因素的水平数目大于5个时,利用二次回归正交旋转组合设计就无能为力了,而均匀设计就是为这种试验而设计的。

2 食品研究常用的分析方法

2.1 对比试验的分析方法

对比试验依据调查指标的不同资料的性质也不同,不同性质的资料分析方法也不同,常见的对比试验资料包括计量资料、计数资料、百分数资料和非参数资料,在一般情况下,计量资料根据样本容量的大小确定分析方法,大洋本多采用 u 测验,而小样本多采用 t 测验;计数资料多采用 χ^2 测验,百分数资料要依据样本母体的性质,由计量资料得到的百分数资料仍然遵从正态分布,可采用 u 测验或 t 测验,而由计数资料得到的百分数资料遵从二项分布,应采用牛顿二项式展开计算误差概率,但利用牛顿二项展开式计算非常麻烦,多采用正态逼近的方法,而正态近似计算要求样本容量大一点,一般 $np > 5$ 或 $nq > 5$ (p 或 q 是低于50%的百分率);非参数资料是食品研究中出现频率比较高的资料,许多新产品都是采用数量化方法得到产品质量凭借标准的,因此,非参数分析方法在食品研究中显得非常重要,非参数资料多采用符号测验或秩和测验的方法。

2.2 随机试验的分析方法

随机试验的结果分析多采用方差分析的方法分析处理之间的显著性,与对比试验相同,随机试验也会遇到不同性质的资料,方差分析是利用统计代换的方法来解决不同性质资料的线性可加性的,一般计量资料可直接进行方差分析,而计数资料多采用平方根代换的方法或对数代换的方法,百分数资料多采用反正弦代换的方法,对于一些不知其分布的资料可利用中心极限定理,采用几个数的平均数进行方差分析,一般样本容量达到4个中心极限定理就成立了。

2.3 线性反应试验的分析方法

线形反应多研究线性关系,因此相关分析和回归分析是主要的分析方法,特别是食品研究多数是参数设计,其主要目的是利用回归分析得到二次回归方程,通过参数分析找到最佳设计参数,故曲线回归应用的比较普遍。

2.4 面体反应试验的分析方法

面体反应试验的结果分析实际上是线性反应试验从二维空间扩展到三维空间,一元线性回归或非线性回归扩展到多元

非线性回归,其原理基本相同。

2.5 多元分析方法

多元分析方法随着计算机的广泛应用日趋被广大研究者所接受,因为多元分析不仅由于研究因素的不断加多导致分析日趋复杂,更重要的事由于多元分析可以利用较少的处理得到足够多的试验信息,而且计算机到广泛应用使计算变得非常简单,分析方法也远远超出统计学的范畴。常用的多元分析方法主要包括通径分析、主成份分析、模糊综合评判和线性规划等。通径分析实际是把简单相关系数分解为各个因素相关系数的代数和,主要研究多个自变数与一个依变数的相关关系;主成份分析是利用变量间的线性关系压缩因子空间,利用尽量少的有效指标反应尽量全的指标信息特征,简称空间压缩;模糊综合评判是利用模糊数学原理把指标间不确定性模糊问题量化,以便根据多指标对产品质量做出综合评价。

3 食品研究与数据分析中应注意的几个问题

3.1 试验设计与分析方法相互吻合

食品研究在试验设计时就要考虑结果的分析方法,以便在试验过程中收集足够的信息,避免在结果分析过程中数据的缺乏。长期以来试验设计与统计分析的着眼点存在着本质的分歧,试验设计努力简化试验,而统计分析却在试验结果的可靠性大做文章,因此往往试验设计与统计分析不能相辅相成,甚至有的试验结果不能分析。原则上讲,什么样的试验设计就用什么样的分析方法,不同的分析方法往往造成结果不同。

3.2 不同性质的资料分析方法不同

如果两个处理差异比较大,不论采用什么样的分析方法,其结果应该是相同的,但如果两个处理间差异处于显著与不显著临界值上,分析方法不合适其分析结果可能不同。不同性质的资料其分布不同,因此假设检验的方法也不同。

3.3 明确试验是一尾测验还是两尾测验

试验是一尾测验还是两尾测验有时是很难确定的,显然一尾测验的否定域位于分布的一侧,也就是正态离差只有大于或小于检验水平临界值才能否定无效假设,而两尾测验的否定域位于分布的两侧,正态离差无论大于或小于检验水平临界值均可否定无效假设,为了便于区分一尾测验和两尾测验,我们提出显著性测验和判别性测验,一般情况下显著性测验为一尾测验,而判别性测验为两尾测验。

3.4 试验设计与抽样调查必须坚持随机的原则

误差分布是建立在随机的基础上的,如果试验排列和抽样调查不是随机的,其误差分布呈偏态分布,因此试验结果存在偏性,不能进行统计检验。

参 考 文 献

- 1 刘魁英等. 食品研究与数据分析. 中国轻工业出版社, 1998
- 2 Harald Martens. Food research and data analysis HELLMUT RUSSWURM JR. 1983.