

统计方法在药品工艺验证中的应用

单晓鹏

北京泽辉辰星生物科技有限公司

摘要：在制药行业，工艺验证是确保产品质量和疗效的关键环节。近年来，随着制药技术的发展和质量管理标准的提高，统计方法在工艺验证中的应用越来越受到重视。本文旨在探讨统计方法在药品工艺验证中的应用，包括设计实验（DoE）、过程能力指数（Cpk）、统计过程控制（SPC）等关键统计工具的使用。通过深入分析这些方法的原理、应用及其对药品质量的影响，本文提出了一系列基于统计分析的工艺验证策略，并通过案例研究展示了这些方法在实际中的应用。此外，本文还探讨了基于统计分析的工艺验证面临的挑战、未来趋势和技术进步，旨在为药品制造和质量控制专业人士提供一个全面的视角，理解统计方法如何加强药品工艺验证。

关键词：工艺验证；统计方法；质量控制

药品工艺验证是药品生产过程中不可或缺的一环，直接关系到最终产品的质量与安全。传统的工艺验证依赖于经验判断和定性分析，而现代工艺验证则更加注重于科学原理和定量分析。统计方法作为一种有效的工具，能够提供深入的数据分析和过程控制，帮助制药企业优化生产流程，提高产品质量。在验证中常用的统计学工具如表 1 所示。

1 统计方法在工艺验证中的应用

在工艺验证过程中，统计学的应用涵盖了从工艺设计、工艺确认到持续工艺验证的各个阶段。通过应用试验设计（DOE）、方差分析、回归分析等统计方法，可以系统地识别物料属性和工艺参数与产品质量之间的关系，从而优化配方和工艺，提高产品质量的一致性和可靠性。此外，稳定性研究、分析方法验证、中间控制和放行标准制定等环节同样离不开统计学的支持。统计工艺控制（SPC）和验收取样等方法，进一步确保了在整个生产过程中质量的持续

监控和改进。

特别是在第二阶段制剂工艺验证的过程中，容差区间、方差分析、控制图和过程能力指数等统计工具的应用，为评估和确保产品质量提供了科学、客观的依据。通过这些方法，可以准确地评估工艺的稳定性与可靠性，确保每一批次的产品都能满足既定的质量标准。

(1) 设计实验（DoE）：DoE 是一种系统的实验方法，通过计划实验的设计、执行和分析，来研究多个因素对一个或多个响应变量的影响。在药品工艺验证中，DoE 可以帮助确定关键过程参数（CPP）和关键质量属性（CQA），优化生产过程。

(2) 容差区间的概念：容差区间是一种统计工具，用来预测一个特定比例（如 90%、95% 或 99%）的未来样本将落在哪个区间内。与置信区间相比，它更关注于覆盖样本的范围，而不仅仅是估计的平均值或比例。

(3) 方差分析（ANOVA）的概念：方差分析（ANOVA）是一种用来比较三个或多个组数据平均值是否存在显著差异

的统计方法。简单来说，ANOVA 可以帮助判断不同地点生产的药品的有效成分含量是否有统计上的显著差异。

(4) 过程能力指数（Cpk）：Cpk 是衡量过程能力的一个重要指标，反映了过程在统计控制状态下，产出符合规格要求的能力。通过计算 Cpk 值，可以评估生产过程的稳定性和可靠性，为过程改进提供方向。

(5) 统计过程控制（SPC）：SPC 是一种通过监控生产过程，识别和消除异常的方法，用于图形化显示过程数据，以便我们可以看到数据随时间的变化情况，特别是识别出任何超出正常变异范围的情况。控制图通常包括中心线（CL，代表过程的平均水平），以及上控制限（UCL）和下控制限（LCL），这些控制限定义了过程变异的可接受范围。

2 案例研究

本文通过几个案例研究展示了统计方法在药品工艺验证中的实际应用。这些案例涵盖了不同阶段的工艺验证活动，从 DoE 的实施、Cpk 的计算到 SPC 的应用，展示了统计方法如何帮助

表 1 验证中常用的统计学工具^[1]

序号	统计参数 / 工具	公式	含义	用途	说明
1	平均值 (Average)	$\bar{x} = \frac{\sum x_i}{n}$	数据集合的中心趋势	在药品制造中评估一批产品的整体性能	对于对称分布的数据集最有效
2	标准偏差 (Standard Deviation)	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$	衡量数据点围绕平均值的分散程度	用于评估过程的一致性和控制水平，是衡量过程能力的关键参数	
3	方差 (Variance)	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	标准偏差的平方，反映数据的分散程度	提供对数据变异性的另一种度量方式	方差的单位是原始数据单位的平方
4	过程能力指数 (Cpk)	见备注	衡量过程在规定的规格限内生产合格产品的能力	用于确定生产过程的能力，确保产品质量	Cpk 值高于 1.33 通常意味着过程良好
5	控制图 (Control Chart)	见备注	用于监控过程稳定性的图形工具	实时监控生产过程，识别过程是否处于控制状态	控制图类型包括 X-bar、R 图、S 图等
6	相关系数 (Correlation Coefficient)	$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$	量化两个变量之间线性相关程度的统计度量	用于评估两个变量之间是否存在线性关系	r 值范围为 -1 到 1，接近 1 或 -1 表示强烈的线性关系
7	设计实验 (Design of Experiments, DoE)	见备注	一组统计原理，用于计划、执行和分析实验，以有效地评估多个输入变量对输出变量的影响	在药品开发和制造过程中，优化过程参数，减少实验次数和成本	DoE 可以是全因子设计、分数因子设计等
8	可靠度区间 (Confidence Interval)	见备注	以给定的置信水平，估计一个参数值将位于某个范围内的概率	用于估计参数值的不确定性范围，并对结果进行解释	
9	假设检验 (Hypothesis Testing)	见备注	一种统计方法，用于检验关于总体参数的假设是否成立	用于验证过程改进措施的效果或比较两个或多个组的性能	包括 t 检验、ANOVA 等
10	方差分析 (ANOVA)	见备注	一种用于比较三个或多个样本均值差异性的统计方法	用于识别不同水平的因子对结果变量的影响	适用于处理设计实验的数

备注：
Cpk 公式： $Cpk = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right)$

其中 USL 是上规格限，LSL 是下规格限，μ 是过程平均值，σ 是过程标准偏差。
控制图、设计实验 (DoE)、可靠度区间、假设检验和方差分析的具体计算方法和应用，因其复杂性和多样性，在实际应用中需要参照具体的统计书籍或软件工具进行。

制药企业提高生产效率和产品质量。

2.1 案例一：使用设计实验 (DoE) 优化混合工艺

背景：此案例涉及三个关键过程参数，分别是混合速度 (A)、混合时间 (B) 和填充物量 (C)，每个参数设定两个水平（低水平表示为 -1，高水平表示为 1）。表 2 列出的是一个基于 2³ 完全因子设计的模拟数据表，用于混合均匀性的分析。在这个模拟数据表中，假设混合均匀性得分是基于这三个关键过程参数的设置而变化的。混合时间 (B) 对混合均匀性的影响最大，其次是混合

表 2 模拟数据表

试验编号	混合速度 (A)	混合时间 (B)	填充物量 (C)	混合均匀性得分
1	-1	-1	-1	75
2	1	-1	-1	80
3	-1	1	-1	85
4	1	1	-1	90
5	-1	-1	1	78
6	1	-1	1	83
7	-1	1	1	88
8	1	1	1	93

速度 (A) 和填充物量 (C)。

分析步骤和结果讨论：

(1) 实验设计

采用 2³ 完全因子设计，考虑三

个参数各有两个水平。

(2) 执行与数据收集

进行 8 次试验，每种参数组合进行一次，记录混合均匀性得分。

(3) 数据分析

分析模拟数据,发现混合时间 (B) 显著影响混合均匀性得分,混合速度(A) 和填充物量 (C) 也有影响,但影响较小。

(4) 结果与讨论

根据数据分析结果,混合时间 (B) 是影响混合均匀性的关键因素,优化混合时间可以显著提高混合均匀性得分。

(5) 实施优化

调整混合工艺,将混合时间设置为优化后的水平 (高水平),重新进行生产试验,验证混合均匀性得分提高,从而证实了优化措施的有效性。

此案例分析展示了如何运用 DoE 方法系统地分析和解决药品生产中的混合均匀性问题,实现了通过优化混合工艺参数提高产品质量的目标。

2.2 案例二: 计算过程能力指数 (Cpk) 评估压片工艺

背景: 为了确保药效,每片药的有效成分含量必须控制在 95 mg ~ 105 mg 之间。实验需要确认生产过程是否能稳定地生产出符合规格的药片。

实施过程:

(1) 数据收集

随机抽取 30 批药片进行测试,发现有效成分的平均含量是 100 mg,标准偏差是 3 mg。

(2) 过程能力指数 (CPK 和 PPK) 的解释

CPK (过程能力指数): CPK 是衡量生产过程如何更好地符合规格要求的指标。简单来说,CPK 指出,如果生产过程继续保持当前的状态,则多大程度上能够确保未来生产的药片会符合规格要求。

PPK (过程性能指数): PPK 与 CPK 类似,但它以生产过程的长期表现而不是短期表现为基础。PPK 考虑了过程中可能发生的任何变化,如机器老化或原料批次间的差异。

(3) 实际计算

规格限: 上限 (USL) 是 105 mg,

下限 (LSL) 是 95 mg。

平均值 (μ): 100 mg

标准偏差 (σ): 3 mg

CPK 计算公式:

$$CPK = \min \left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma} \right)$$

将数据代入公式:

$$CPK = \min \left(\frac{105 - 100}{3 \times 3}, \frac{100 - 95}{3 \times 3} \right)$$

$$CPK = \min \left(\frac{5}{9}, \frac{5}{9} \right)$$

$$CPK = 0.56$$

解释: CPK 值为 0.56,这在制药行业通常是不可接受的,因为它表明过程中有相当比例的产品可能不会符合规格要求。一般来说,CPK 值如果低于 1.33,意味着过程需要改进。

2.3 案例三: 应用统计过程控制 (SPC) 监测生产过程

背景: 大多数的 SPC 参考文献都建议收集 20-30 组数据来计算控制限以评估工艺是否受控。可以从 3 批或更多批中取 30 组样本。对于 3 批来说,可以将每批分为 10 组合理子群并取样。然后计算 SPC 图的限度并对工艺进行统计学控制评估。使用批次的数量取决于 SPC 图检测到非期望的批间变化的能力。

实施过程和结果讨论:

(1) 数据收集

在连续的生产批次中,对每个批次进行片重的测量。

(2) 控制图的应用

选用 I-MR 控制图来监控 CQA (片重) 的平均值和范围,通过这些控制图实时跟踪过程的变异,如图 1 和图 2 所示。

(3) 过程调整

一旦控制图显示过程偏离控制限值,即时进行过程调整和根本原因分析。

(4) 结果与讨论

通过持续的 SPC 监控,表明片重的控制过程稳定,

2.4 案例四: 方差分析考察不同生产地点的产品质量 (简化计算,不考虑每个地点产品的批间差异)

背景: 假设制药公司正在测试一种新药的有效成分含量是否受到制造地点的影响。公司在三个不同的地点 (A、B、C) 生产了这种药物,并从每个地点随机抽取了五批药品进行测试。

实施过程和结果讨论:

(1) 数据收集

表 3 显示的是三个地点生产的产品有效成分含量测量值。

(2) 过方差分析 (ANOVA)

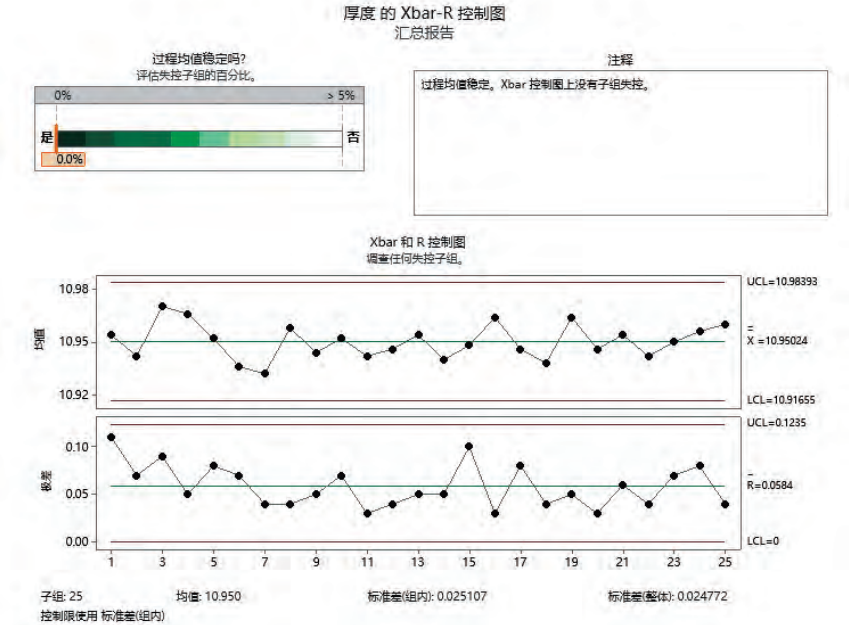


图 1 厚度的 Xbar-R 控制图汇总报告

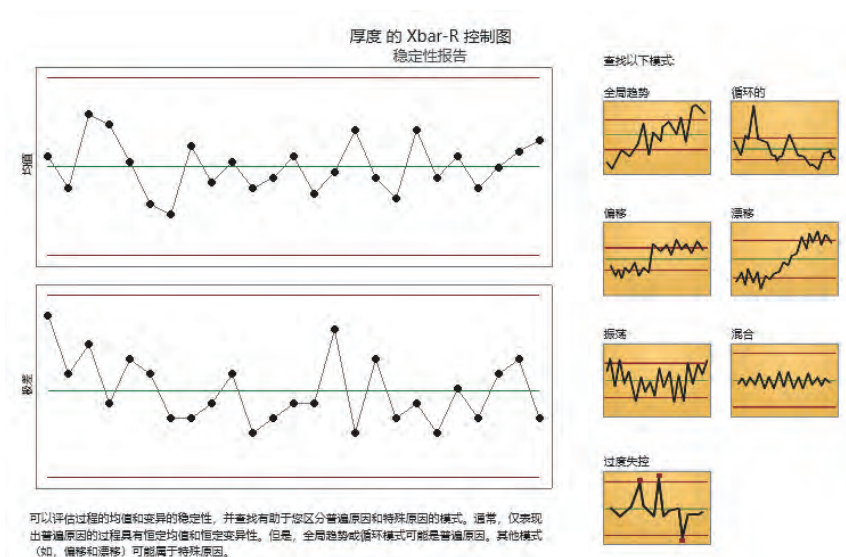


图 2 厚度的 Xbar-R 控制图稳定性报告

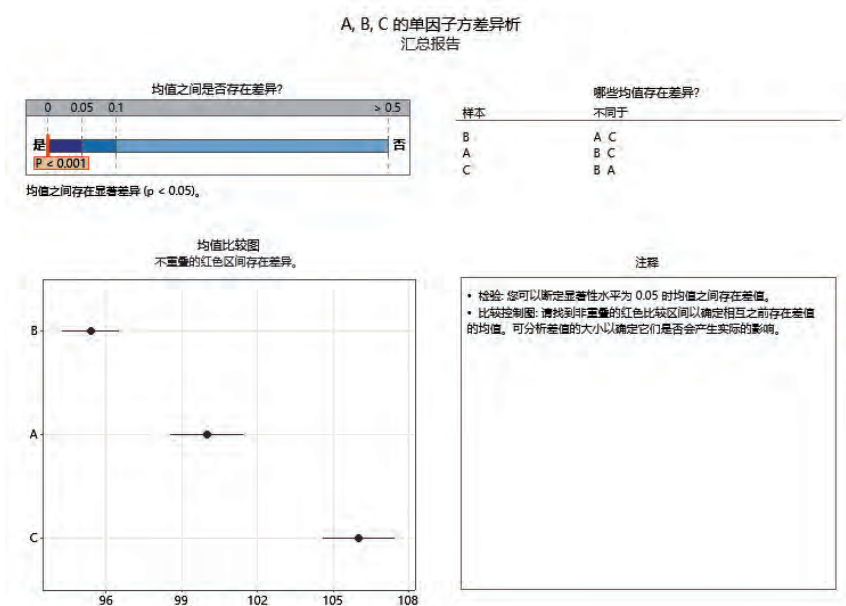


图 3 A、B、C 的单因子方差分析汇总报告

表 3 三个地点生产的药品的有效成分含量测量值 (单位: mg):

地点	批次 1	批次 2	批次 3	批次 4	批次 5
A	99	101	100	102	98
B	95	96	94	97	95
C	105	107	106	104	108

制药公司可以系统地评估和确保不同生产地点的药品在有效成分含量上的一致性。这种方法不仅适用于制药行业，也广泛应用于各种科学研究和工业生产中，以保证产品和服务的质量。

(3) 实际计算

为了简化计算，本文只介绍

ANOVA 的基本概念而不进行详细的数学运算。通常，ANOVA 的计算需要通过统计软件来完成。

总平均计算: 计算所有样本的总平均含量。

组内变异: 查看每个地点内部的测量值与该地点平均值的差异。

组间变异: 查看每个地点的平均含量与总平均含量的差异。

(4) 结果与讨论

如果组间变异显著大于组内变异，这意味着不同地点间的差异不仅仅是随机变异，而是有统计学上的显著差异，如图 3 所示。

3 面临的挑战与未来趋势

尽管统计方法在药品工艺验证中展现出显著的优势，但在实践中仍面临诸多挑战，包括数据质量的保证、统计技能的培养以及方法的正确应用等。未来，随着人工智能和大数据技术的发展，统计方法将更加智能化和自动化，为药品工艺验证提供更强大的支持。

4 结论

统计方法为药品工艺验证提供了一种科学、系统的分析工具，有助于提高生产过程的控制水平和产品质量。制药和生物制药工业越来越认识到统计方法的重要性，以持续生产符合预定质量标准的产品。统计方法为符合这一目的提供了客观证据，它也是理解工艺的基础，使持续改进和开发成为可能。

通过实施 DoE、计算 Cpk 和应用 SPC 等统计工具，制药企业可以优化生产流程，确保产品符合质量标准。面向未来，统计方法的应用将更加广泛和深入，为药品制造业的持续改进和创新提供强有力的支持。

参考文献:

- [1] PDA TB Validation by Design The Statistical Handbook for Pharmaceutical Process Validation 2015 official version