

假设检验的三个决策方法

郭 莹

北京科技大学天津学院基础部, 天津

收稿日期: 2025年10月27日; 录用日期: 2025年11月18日; 发布日期: 2025年12月1日

摘 要

文章介绍了假设检验决策的三种方法: 临界值法、p值法和区间估计法。临界值法适合检验统计量手工计算时使用, 临界值法通过显著性水平 α 确定临界值与检验统计量的值比较大小做出决策; 若检验统计量值的计算依赖于软件, 适合使用p值做决策, p值小于显著性水平 α , 拒绝原假设, 否则接受原假设; 区间估计法帮助深层次理解区间估计与假设检验间的关系, 对于同一总体参数, 其区间估计和假设检验都依赖于同一样本统计量的抽样分布, 参数的置信区间为参数假设检验的接受域。

关键词

假设检验, 临界值法, p值法, 区间估计法

Three Decision-Making Methods in Hypothesis Testing

Xuan Guo

Department of Basic Courses, Tianjin College of University of Science and Technology Beijing, Tianjin

Received: October 27, 2025; accepted: November 18, 2025; published: December 1, 2025

Abstract

The article introduces three methods for making decisions in hypothesis testing: critical value method, p-value method and interval estimation method. The critical value method is suitable for use when calculating statistical quantities manually. The critical value method determines the critical value based on the significance level α , and then compares the value of the test statistic with the critical value to make a decision. If the calculation of the test statistic value relies on software, it is appropriate to make decisions based on the p-value. If the p-value is less than the significance level α , the null hypothesis is rejected; otherwise, the null hypothesis is accepted. The interval estimation method helps to deeply understand the relationship between interval estimation and hypothesis

testing, For the same overall parameter, both interval estimation and hypothesis testing rely on the sampling distribution of the same sample statistics. The confidence interval of the parameter serves as the acceptance domain for the parameter hypothesis test.

Keywords

Hypothesis Testing, Critical Value Method, p-Value Method, Interval Estimation Method

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

统计学的分析方法分为描述统计和推断统计两大类，描述统计法是对数据的初步分析，通过整理排序发现数据基本规律、绘制图形、计算一组数据的平均指标和离散程度测度指标等。推断统计法是在描述统计的基础上结合概率论的知识由样本统计量推断总体参数的方法。推断统计包括参数估计和假设检验两种方法，参数估计是利用样本统计量，估计得到总体参数的一个估计值或者将总体参数估计在某个区间中。假设检验的思想最早是由统计学家费歇尔提出的，后经奈曼和皮尔逊等统计学家的改进完善，形成了一种程式化的统计推断模式[1]。假设检验是对总体参数事先有一个认识，然后利用样本构造一个检验统计量，通过检验统计量的值判断总体参数是否等于某个设定值。假设检验具有标准的计算步骤，第一步对总体参数提出原假设和备择假设，第二步通过样本构造检验统计量并计算其取值，第三步在给定的显著性水平 α 下查找得到临界值，通过检验统计量的值与临界值大小比较做出决策。文章在讨论假设检验决策的临界值法和 p 值法的基础上，使用区间估计法对总体参数进行假设检验，探讨参数的区间估计和假设检验之间的联系。

2. 临界值法

大部分教材中将临界值法作为假设检验决策的主要方法，该方法会事先给定一个显著性水平 α ， α 在概率统计中是研究者选取的一个非常小的概率值，表示犯第一类错误的概率。第一类错误是指假设检验中原假设命题为真，却将其拒绝所犯的错误，通常取 0.01, 0.05, 0.1 三个值，其中最常用的是 $\alpha = 0.05$ ，在社科、医学、心理学等领域常用。 $\alpha = 0.01$ 表明控制犯第一类错误的概率为 0.01，这个显著性水平的选取表示只有样本提供的证据足够强，才能拒绝原假设，适合高风险的研究，比如药物的临床研究、精密制造等领域。在具体假设检验决策中，研究者会根据可容忍的犯第一类错误的概率选取合适的 α 进行决策。

2.1. 选择不同的 α 得到的结论一致

α 是假设检验前研究者选取的，通过 α 可以查表得到检验的临界值，在计算得到检验统计量的值后，将临界值和检验统计量的值对比做出决策。假设检验中首先会提出假设，一般将原假设记为 H_0 ，备择假设记为 H_1 ，在提出假设时，将研究者希望证明的命题放在备择假设，再提出原假设。

例 1：一家制药公司生产一种药片，其有效成分的理论标准含量为 50.0 毫克，如果含量过低影响患者治疗，药效不足，如果含量过高，可能会增加副作用的风险。药片有效成分严格控制在 50.0 毫克至关重要。已知总体标准差 $\sigma = 1.2$ 毫克，有效成分服从正态分布。某一天质检员从生产线随机抽取了 20 批药

片, 进行精密检测, 测得样本平均有效成分含量为 49.3 毫克, 这次抽样是随机波动还是生产流程出现了系统性偏差, 导致平均含量不再是 50.0 毫克?

该研究背景关注的是有效成分的平均含量是否为 50.0 毫克, 低于或者高于 50 毫克都属于不达标, 故为双侧假设检验, 总体服从正态分布, 总体标准差已知, 使用 z 检验统计量进行假设检验。

1) 提出假设 $H_0: \mu = 50$, $H_1: \mu \neq 50$ 。

2) 计算检验统计量:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{49.3 - 50}{1.2/\sqrt{20}} \approx -2.61$$

3) 做出决策, 当 $\alpha = 0.01$ 时, 查表得 $z_{0.005} = 2.58$, $|-2.61| > 2.58$, 拒绝原假设, 有足够证据说明平均含量不等于 50.0 毫克, 这次抽样的随机波动使生产流程出现了系统性偏差。当 $\alpha = 0.05$ 时, 查表得 $z_{0.025} = 1.96$, $|-2.61| > 1.96$, 同样拒绝原假设, 有足够证据证明生产流程出现了系统性偏差。

例 1 中选择了不同的显著性水平 α , 得到的结论是一致的, 都是拒绝原假设, 说明样本数据所包含的信息具有足够的代表性, 样本统计量的取值偏离 H_0 较远, 不管是相对宽松的犯第一类错误的概率 0.05, 还是相对严格的犯第一类错误的概率 0.01, 统计结果都是显著的。此时制药公司应果断停工对生产线进行检测, 检测结束后需再次抽样进行假设检验。

2.2. 选择不同的 α 得到的结论不一致

例 2: 根据历史资料某食品公司产品中的农药残留物含量服从正态分布, $X \sim N(1020, 100^2)$, 经过技术改造, 采用了新的生产方法, 现在从最近生产的一批产品中随机抽取 16 袋, 测得样本平均值为 980 微克, 这批产品农药残留物的含量是否有显著降低?

该研究背景关注采用新生产方法后, 农药残留物是否显著降低, 是否显著降低是研究者想要证明的命题放在备择假设, 该假设检验为左侧假设检验。已知总体服从正态分布, 总体方差已知, 采用 z 检验统计量。

1) 提出假设 $H_0: \mu \geq 1020$, $H_1: \mu < 1020$ 。

2) 计算检验统计量的值:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{980 - 1020}{100/\sqrt{16}} = -1.6$$

3) 做出决策, 当 $\alpha = 0.05$ 时, 查表得 $z_{0.05} = 1.645$, 左侧假设检验的临界值为 -1.645 , $-1.6 > -1.645$, 落在接受域, 没有足够的证据说明这批农药残留物的含量有显著降低。当 $\alpha = 0.1$ 时, 查表得 $z_{0.1} = 1.28$, 临界值为 -1.28 , $-1.6 < -1.28$, 落入拒绝域, 有足够证据说明这批产品农药残留物含量显著降低。

例 2 中选择了不同的显著性水平 α , 得到不一样的结论, 在具体问题中该如何决策。显著性水平从 0.05 取到 0.1, 增大了犯第一类错误的概率, 条件变得宽松时落入拒绝域认为农药残留物降低, 此时对应的决策是采用新的生产方法, 其缺点是犯第一类错误的概率较高, 若可以容忍这个犯错的概率, 就采用新的生产方式进行生产。若农药残留物带来的危害大, 控制犯第一类错误的概率为 0.05, 落在接受域没有足够的证据证明新的方法确实可以降低农药残留物, 此时对应的决策是新的生产方法不投入使用, 继续使用旧生产方法且不断改良新生产方法, 改良后继续抽样进行假设检验。是否采用新生产方法依据食品公司实际情况决定。

假设检验中临界值法人为控制了犯第一类错误的概率, 真实的犯第一类错误的概率可能比给定的 α 高或者低, 假设检验针对具体研究目的展开, 检验的目的是做出某种决策, 从上述分析看, 假设检验决策相对主观, 依赖于研究者的经验同时依赖于研究背景所能容忍的犯第一类错误的概率。

3. p 值法

3.1. p 值做决策

假设检验中的 p 值是能够做出拒绝原假设的最小显著性水平[2]，用 p 值做决策相对比较客观。统计学中使用 p 值做决策的准则为：若 $p \leq \alpha$ ，在显著性水平 α 下拒绝原假设；若 $p > \alpha$ ，在显著性水平 α 下接受原假设。下面分别计算例 1 和例 2 中的 p 值。假设检验的 p 值可以通过 EXCEL 中的函数计算得到。

例 1 中，对于药片有效成分平均含量为 50.0 毫克的检验为双侧假设检验，双侧假设检验的 p 值为： $p = P(|z| \geq z_c)$ ， z_c 表示检验统计量值的绝对值，例 1 中检验统计量的值为 -2.61，故 $z_c = 2.61$ ， $P(z \geq 2.61) = 1 - P(z \leq 2.61) = 1 - \Phi(2.61) = 1 - 0.995473 = 0.004527$ ，故 p 值 $= 2 \times 0.004527 = 0.009054$ 。对于给定的显著性水平 $\alpha = 0.01$ ， $p < \alpha$ ，拒绝原假设，有足够证据证明药片有效成分平均含量不是 50 毫克。对于给定的显著性水平 $\alpha = 0.05$ ， $p < \alpha$ ，拒绝原假设，同样有足够证据证明药片有效成分平均含量不是 50 毫克。p 值取 0.009054，只要 α 取值大于 p 值，都可以做出拒绝原假设的决策。

例 2 中对于农药残留物的含量是否有显著降低的假设检验为左侧假设检验，左侧假设检验的 p 值为 $p = P(z \leq z_c)$ ，例 2 中 $z_c = -1.6$ ， $p = P(z \leq -1.6) = \Phi(-1.6) = 1 - \Phi(1.6) = 1 - 0.945221 = 0.054779$ ，对于给定的显著性水平 $\alpha = 0.05$ ， $p > \alpha$ ，不拒绝原假设，认为农药残留物没有显著降低。只有当 α 取值大于 0.054779 时，才能得到拒绝原假设的结论。

3.2. 临界值法与 p 值法等价

假设检验中的临界值法和 p 值法是等价的，临界值法是通过比较检验统计量的值与一个临界值的大小做出决策，无论是检验统计量还是临界值，本质上都是检验统计量分布的分位点值；p 值法是通过与显著性水平 α 的大小比较做出决策，无论是 p 值还是 α 值本质都是一个小的概率值，在统计量分布的概率密度函数图形中表现为曲线、横轴和分位点围成的面积。临界值法和 p 值法只是角度不同，效果是一样的。只不过临界值法需要通过查表寻找临界值，同时还需要计算检验统计量的值，p 值法只需要 p 值与 α 值比较即可，如果是软件输出的假设检验结果，可以直观看到 p 值的大小，省去了检验统计量计算的过程。图 1 给出了例 1 中的临界值和检验统计量的值，-1.96 左侧横轴与曲线围成的面积为显著性水平 $\alpha = 0.05$ ，-2.61 左侧横轴与曲线围成的面积为假设检验的 p 值， $p = 0.009054$ 。故临界值法是通过分位点的值做决策，p 值法是通过比较面积大小做决策。

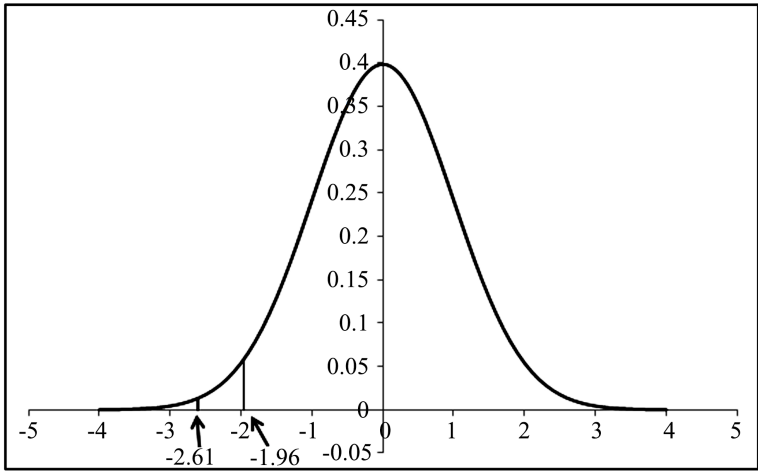


Figure 1. Standard normal distribution quantile
图 1. 标准正态分布分位点

4. 区间估计法

临界值法和 p 值法是常用的假设检验决策方法, 假设检验的区间估计法做决策是用参数的区间估计得到假设检验的结论。在参数的区间估计中研究者关心的是总体参数在给定的置信度 $1-\alpha$ 下, 通过样本计算得到的参数的估计区间, 这个估计区间对应的是参数假设检验的接受域。假设检验中研究者关注的是备择假设命题发生的概率 α , 是小概率事件。

4.1. 双侧假设检验的区间估计法

对于总体均值 μ 双侧假设检验, 说明如何通过 z 检验统计量推出假设检验接受域。显著性水平取 α , 提出假设为 $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$, μ_0 为总体均值的一个已知的设定值, 检验统计量 z 落在 $(-z_{\alpha/2}, z_{\alpha/2})$ 则接受原假设, 否则拒绝原假设。 $-z_{\alpha/2} \leq z \leq z_{\alpha/2} \Rightarrow -z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \Rightarrow -z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \Rightarrow \bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$, z 落在 $(-z_{\alpha/2}, z_{\alpha/2})$ 等价于 μ_0 落在 $\left(\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right)$, 即 μ_0 落在 $\left(\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right)$ 则接受原假设, 否则拒绝原假设。该方法对其他总体参数同样适用。这个推导过程就是总体均值在置信度为 $1-\alpha$ 下总体均值 μ 的区间估计的推导, 区间估计中更关注 μ 以一较大概率落入的区间, 抽样允许误差是多大, $\bar{x}$ 为 μ 的点估计, $z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ 为样本均值的抽样允许误差, 区间估计本质上是在对总体参数信息未知时, 通过总体参数的点估计加减一个抽样允许误差得到的, 保证在多次抽样中, 总体参数以概率 $1-\alpha$ 落入置信区间。假设检验事先对总体参数有一个认识, 更关注在一次具体以抽样中, 通过样本计算得到的置信区间是否能包含总体参数的设定值 μ_0 。由此可见, 参数估计和假设检验站在不同的角度对总体参数做出统计推断, 但其依据都是样本统计量的抽样分布。

例 1 中对药片有效成分的平均含量为 50.0 毫克的双侧假设检验 $H_0: \mu = 50$, $H_1: \mu \neq 50$, $\bar{x} = 49.3$, $\alpha = 0.05$, $z_{0.025} = 1.96$, $\sigma = 1.2$, $n = 20$, 双侧假设检验的接受域为: $\left(\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \left(49.3 - 1.96 \times \frac{1.2}{\sqrt{20}}, 49.3 + 1.96 \times \frac{1.2}{\sqrt{20}}\right) \Rightarrow (48.77, 49.83)$, $\mu_0 = 50$ 不在这个区间中, 故拒绝原假设, 有足够证据证明 $\mu \neq 50$ 。说明生产流程出现了系统性偏差。

4.2. 单侧假设检验的区间估计法

对于总体均值 μ 的单侧假设检验, 分为左单侧假设检验和右单侧假设检验。对于左侧假设检验 $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$, 用 z 检验统计量推出接受域。显著性水平取 α , μ_0 为总体均值的一个已知的设定值, 检验统计量 z 落在 $(-z_{\alpha}, +\infty)$ 则接受原假设, 否则拒绝原假设。 $z \geq -z_{\alpha} \Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq -z_{\alpha} \Rightarrow \bar{x} - \mu \geq -z_{\alpha} \times \frac{\sigma}{\sqrt{n}} \Rightarrow \mu \leq \bar{x} + z_{\alpha} \times \frac{\sigma}{\sqrt{n}}$ 。 z 落在 $(-z_{\alpha}, +\infty)$ 等价于 μ_0 落在 $\left(-\infty, \bar{x} + z_{\alpha} \times \frac{\sigma}{\sqrt{n}}\right)$, 即 μ_0 落在 $\left(-\infty, \bar{x} + z_{\alpha} \times \frac{\sigma}{\sqrt{n}}\right)$ 则接受原假设, 否则拒绝原假设。

对于总体均值 μ 的右单侧假设检验, $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$, 用 z 检验统计量推出接受域。显著性水平取 α , μ_0 为总体均值的一个已知的设定值, 检验统计量 z 落在 $(-\infty, z_{\alpha})$ 则接受原假设, 否则拒绝原假设。 $z \leq z_{\alpha} \Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha} \Rightarrow \bar{x} - \mu \leq z_{\alpha} \times \frac{\sigma}{\sqrt{n}} \Rightarrow \mu \geq \bar{x} - z_{\alpha} \times \frac{\sigma}{\sqrt{n}}$ 。 z 落在 $(-\infty, z_{\alpha})$ 等价于 μ_0 落在

$\left(\bar{x} - z_{\alpha} \times \frac{\sigma}{\sqrt{n}}, +\infty\right)$, 即 μ_0 落在 $\left(\bar{x} - z_{\alpha} \times \frac{\sigma}{\sqrt{n}}, +\infty\right)$ 则接受原假设, 否则拒绝原假设。

例 2 中对农药残留物是否显著降低的假设检验为左侧假设检验, $H_0: \mu \geq 1020$, $H_1: \mu < 1020$, $\bar{x} = 980$, $\alpha = 0.05$ 时, 查表得 $z_{0.05} = 1.645$, $\sigma = 100$, $n = 16$, 左侧假设检验的接受域为 $\left(-\infty, \bar{x} + z_{\alpha} \times \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \left(-\infty, 980 + 1.645 \times \frac{100}{\sqrt{16}}\right) \Rightarrow (-\infty, 1021.125)$, $\mu_0 = 1020$, 落入接受域, 故没有足够的证据说明这批农药残留物的含量有显著降低。

5. 结论

本文介绍了假设检验决策的三种方法: 临界值法、p 值法和区间估计法, 三种方法在对总体参数进行假设检验时等价, 同一个假设检验分别用三种方法得出的结论是一致的。文章是以单个总体均值的单侧和双侧假设检验为例说明三种决策方法是等价的, 在实际应用中假设检验决策方法不会局限于单个总体, 对于两个总体均值之差、两个总体比例之差、两个总体方差之比、方差分析的假设检验中也可以使用三种决策方法做出决策。文章得出的结论主要有三点。

一是临界值检验法适合手工计算检验统计量的值时使用, 通过给定显著性水平 α 查表得到临界值, 检验统计量值与临界值大小比较做出决策。显著性水平 α 是研究者事先给定的, 在某些假设检验中, 选取不同的显著性水平 α , 得出的结论是不一致的, 此时需要结合研究背景考虑可容忍的犯第一类错误的概率, 再选择接受还是拒绝原假设。

二是 p 值可以通过统计软件输出, 当假设检验的计算依赖于统计软件时使用 p 值较为方便。p 值是观测到的显著性水平[3], 即通过样本计算出来的显著性水平, 当观测的显著性水平小于研究者选定的显著性水平时, 拒绝原假设, 否则接受原假设。

三是区间估计检验法可以深层次理解区间估计和假设检验之间的关系, 区间估计在估计之前对参数没有认识, 给定置信度 $1-\alpha$ 时, 由样本统计量加减误差得到参数的区间估计, 区间估计时更关注置信度 $1-\alpha$ 这个大概率; 假设检验在检验之前对总体参数有一个认识, 提出假设时更关注备择假设发生概率 α , α 为小概率。参数的区间估计与参数假设检验的关系主要体现在参数的置信区间为参数假设检验的接受域, 这是由于无论是区间估计还是假设检验, 对于同一个参数, 都依赖于同一个样本统计量的抽样分布。

基金项目

北京科技大学天津学院第四批校级一流课程, 项目名称: 《应用统计学》线下一流课程(项目编号: YLKC202430)。

参考文献

- [1] 王军虎. 统计检验中假设的设置方法[J]. 统计与决策, 2022(21): 57-59.
- [2] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程(第三版) [M]. 北京: 高等教育出版社, 2019.
- [3] Mendenhall, W.M. and Sincich, T.L. (2016) Statistics for Engineering and the Sciences. 6th Edition, CRC Press.