

假设检验

基本思路

正太总体-均值

单个正太总体-均值

方差已知- Z 检验

方差未知- t 检验

两个正太总体-均值差- t 检验

基于成对数据的检验- t 检验

正太总体-方差

单个总体- 卡方检验

两个总体- F 检验

置信区间与假设检验之间的关系

样本容量的选取

分布拟合检验

卡方拟合检验法

假设前提

拟合优度检验的基本原理和步骤

偏度、峰度检验

秩和检验

假设检验问题的 P 值法

基本思路

正太总体-均值

单个正太总体-均值

方差已知- Z 检验

方差未知- t 检验

两个正态总体-均值差- t 检验

基于成对数据的检验- t 检验

正态总体-方差

单个总体- 卡方检验

两个总体- F 检验

置信区间与假设检验之间的关系

样本容量的选取

分布拟合检验

上面介绍的各种检验方法都是在总体分布形式为已知的前提下进行的讨论，但在实际问题中，有时不能知道总体服从什么类型的分布，这时就要根据样本来检验关于分布的假设。

这里通过 χ^2 拟合检验法，来检验总体是否具有某一个指定的分布或属于某一个分布族（即分布中有未知参数）。

还介绍专用于检验分布是否为正态的偏度和峰度检验法。

卡方拟合检验法

假设前提

记： $F(x)$ 为总体 X 的位置的分布函数。

假设： $F_0(x)$ 是形式已知，但可能含有若干个未知参数的分布函数。

检验假设： $H_0 : F(x) = F_0(x) \quad \forall x \in$

注意：

1. 一般比如检验是否符合孟德尔遗传定律这种没有参数的，但大多是检验有参数的分布，比如检验是否符合泊松分布，泊松分布是有参数 λ 的，这类有参数的分布拟合假设也叫做分布族的 χ^2 拟合检验。
2. 若总体 X 为离散型，则原假设 H_0 ：总体 X 的分布律为 $P\{X = t_i\} = p_i, i = 1, 2, \dots$
3. 若总体 X 为连续型，则原假设 H_0 ：总体 X 的概率密度为 $f(x)$
4. 备择假设是除了这个分布之外所有的分布，故不用写出。

拟合优度检验的基本原理和步骤

1. 在 H_0 下，总体 X 取值的全体分成 k 组，即 k 个两两不想交的子集 A_1, \dots, A_k .
2. 以 $n_i (i = 1, \dots, k)$ 记样本观察值 x_1, \dots, x_n 中落 A_i 内的个数（实际频数），且 $\sum_{i=1}^k n_i = n$.
3. 当 H_0 为真且 $F_0(x)$ 完全已知时，计算事件 A_i 发生的概率 $p_i = P_{F_0}(A_i), i = 1, \dots, k$ ；
当 $F_0(x)$ 含有 r 个未知参数时，先利用极大似然法估计 r 个未知参数，然后求得 p_i 的估计 \hat{p}_i .
此时称 np_i （或 $n\hat{p}_i$ ）为理论频数.

4. 直观来看, 如果 H_0 成立, 实际频数 n_i 与理论频数 np_i 相差不会太大, 基于这种想法, 我们会选择:

$$\text{检验统计量形式: } \sum_{i=1}^k h_i (n_i - np_i)^2, h_i = ?$$

$$\text{检验的拒绝域形式: } \sum_{i=1}^k h_i (n_i - np_i)^2 \geq c$$

统计量分布: 若 n 充分大, 则当 H_0 为真时, 统计量

$$\chi^2 = \sum_{i=1}^k \frac{h_i (n_i - np_i)^2}{np_i} \stackrel{\text{近似}}{\sim} \chi^2 (k-1)$$

$$\chi^2 = \sum_{i=1}^k \frac{h_i (n_i - n\hat{p}_i)^2}{n\hat{p}_i} \stackrel{\text{近似}}{\sim} \chi^2 (k-r-1)$$

其中 k 为分类数, r 为 $F_0(x)$ 中被估未知参数的个数。

检验统计量:

$$\chi^2 = \sum_{i=1}^k \frac{h_i (n_i - np_i)^2}{np_i} = \sum_{i=1}^k \frac{n_i^2}{np_i} - n \quad \text{或}$$

$$\chi^2 = \sum_{i=1}^k \frac{h_i (n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \sum_{i=1}^k \frac{n_i^2}{n\hat{p}_i} - n$$

显著水平 α 下拒绝域:

$$\chi^2 = \sum_{i=1}^k \frac{n_i^2}{np_i} - n \geq \chi_{\alpha}^2 (k-1), \quad (\text{没有参数需要估计})$$

$$\chi^2 = \sum_{i=1}^k \frac{n_i^2}{n\hat{p}_i} - n \geq \chi_{\alpha}^2 (k-r-1), \quad (\text{有 } r \text{ 个参数需要估计})$$

5. 卡方拟合检验使用时必须注意:

n 要足够大, $n \geq 50$;

np_i (或 $n\hat{p}_i$) ≥ 5 ;

否则应适当合并相邻的类 (组), 以满足要求。

偏度、峰度检验

秩和检验

假设检验问题的 P 值法

p-value 是指在一个概率模型中, 统计摘要 (如两组样本均值差) 与实际观测数据相同, 或甚至更大这一事件发生的概率。换言之, 是检验假设零假设成立或表现更严重的可能性。**p-value** 若与选定显著性水平

(0.05 或 0.01) 相比更小, 则零假设会被否定而不可接受。然而这并不直接表明原假设正确。通常在零假设下, `p-value` 是一个服从 $[0, 1]$ 区间均匀分布的随机变量, 在实际使用中因样本等各种因素存在不确定性。产生的结果可能会带来争议。

简单来说, `p-value` 就是在假设原假设 (H_0) 正确时, 出现现状或更差的情况的概率。

从研究总体中抽取一个随机样本计算检验统计量的值计算概率 `p-value` 或者说观测的显著水平, 即在假设为真时的前提下, 检验统计量大于或等于实际观测值的概率, 当然要看假设的情况选择单侧 `p-value` 和双侧 `p-value`:

1. 如果 `p-value < 0.01`, 说明是较强的判定结果, 拒绝假定的参数取值。
2. 如果 `0.01 < p-value < 0.05`, 说明较弱的判定结果, 拒绝假定的参数取值。
3. 如果 `p-value > 0.05`, 说明结果更倾向于接受假定的参数取值。

可是, 那个年代, 由于硬件的问题, 计算 `p-value` 并非易事, 人们就采用了统计量检验方法, 也就是我们最初学的 t 值和 t 临界值比较的方法。统计检验法是在检验之前确定显著性水平 α , 也就是说事先确定了拒绝域。但是, 如果选中相同的 α , 所有检验结论的可靠性都一样, 无法给出观测数据与原假设之间不一致程度的精确度量。只要统计量落在拒绝域, 假设的结果都是一样, 即结果显著。但实际上, 统计量落在拒绝域不同的地方, 实际上的显著性有较大的差异。因此, 随着计算机的发展, P 值的计算不再是个难题, 使得 P 值变成最常用的统计指标之一。

以上关于 `p-value` 内容分别摘录自 [Wikipedia](#), [知乎](#) 和 [百度百科](#)。