卡方检验——挽救参数检验的一颗稻草

作者：张倩 来源：爱科学

本文原地址：https://www.iikx.com/news/statistics/5962.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

到底什么是卡方检验?

卡方检验就是卡方分布为基础的一种检验方法，主要用于分类变量，根据样本数据推断总体的分布与期望分布是否有显著差异，或推断两个分类变量是否相关或相互独立。其原假设为：观察频数与期望频数没有差别。凡是可以应用比率进行检验的资料，都可以用卡方检验。

要注意的是，卡方检验受样本量的影响很大，同样两个变量，不同的样本量，可能得出不同的结论。解决这个问题的办法是对卡方值进行修正，最常用的是列联系数。对较大样本，当卡方检验的的结果显著，并且列联系数也显著时(列联系数至少超过
0.16)，才可拒绝原假设;当卡方检验的结果显著，列联系数不显著时，不能轻易下结论。

卡方检验的用途?

一般情况下分类变量用卡方检验，连续性变量用 t 检验或者 u 检验。小样本 n<50 时，通常是 t 检验，大样本用 u 检验。

卡方检验最常见的用途就是考察某无序分类变量各水平在两组或多组间的分布是否一致实际上，除了这个用途之外，卡方检验还有更广泛的应用。具体而言，其用途主要包括以下几个方面：

1、检验某个连续变量的分布是否与某种理论分布相一致。如是否符合正态分布、是否服从均匀分布、是否服从 Poisson 分布等。

2、检验某个分类变量各类的出现概率是否等于指定概率。如在 36 选 7
的彩票抽奖中，每个数字出现的概率是否各为 1/36;掷硬币时，正反两面出现的概率是否均为 0.5。

3、检验某两个分类变量是否相互独立。如吸烟 (二分类变量：是、否) 是否与呼吸道疾病 (二分类变量：是、否) 有关;产品原料种类 (多分类变量) 是否与产品合格 (二分类变量) 有关。

4、检验控制某种或某几种分类因素的作用以后，另两个分类变量是否相互独立。如在上例中，控制性别、年龄因素影响以后，吸烟是否和呼吸道疾病有关;控制产品加工工艺的影响后，产品原料类别是否与产品合格有关。

5、检验某两种方法的结果是否一致。如采用两种诊断方法对同一批人进行诊断，其诊断结果是

否一致;采用两种方法对客户进行价值类别预测，预测结果是否一致。

什么情况下使用卡方检验?

参数检验对观测值的 普遍要求是总体呈正态分布，但实际研究中，不是所有观测值都呈正态分布，或者无法确定其是否正态分布，由于缺乏足够信息，总体的分布未知，这些情况下，参数检验技术就未必适用了，因此我们还需要掌握一些非参数检验技术，其中最为常用的就是卡方检验，它最适合于次数分布检验。

卡方检验作为非参数检验的一种，其稳健性不及参赛检验，因此，从使用的角度来看，应首选参数检验，如果在无法满足参数检验基础条件的前提下，再考虑使用非参数检验，例如，样本来自的总体数据不符合正态分布，此时可采用卡方检验完成。另外，研究定类变量和定序变量之间的关系时，由于定类或者定序变量都不具有完备的运算性能，因此无法对总体某种参赛的计算，可采用非参数检验如卡方检验来完成。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发