
两个变量都是多分类变量时，应该选用什么统计方法

作者：王晓晓，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1918.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！

两个分类变量都是多分类变量时，常称为 $R \times C$ 列表资料，也就是两个多分类变量。这种数据常分为以下几种情况：两个变量都是无序的、两个变量都是有序的、只有一个变量是有秩序的，下面我们分别进行讨论。这里仅限于数据是独立的情况，不涉及配对设计的问题。

一、两个变量，都是无序多分类的

研究者针对这种资料常常会提出这样的问题：

Q1

：比较多个样本的构成比。比如，研究3个病区，胃溃疡和胃癌的构成比有无差异。这种情况下，直接采用R × C列表的卡方检验即可。

Q2：研究两个变量之间的关联。比如，研究血型与胃溃疡、胃癌的关系。

这种情况下，可采用卡方检验来推断两个变量是否有关系，在有关系的前提下，再计算Pearson列联系数。

二、两个变量，都是有序多分类的

研究者针对这种资料常常会提出这样的问题：

Q1

：比较不同组间某指标的变化。比如，研究某药对不同病情患者的疗效，这种情况下可以采用卡方检验，但卡方检验只能告诉我们不同组间在疗效按等级的构成上有没有不同，而不能说明哪一组疗效较好，哪一组疗效较差。如果想知道哪组疗效较好，可以采用非参数秩和检验。

Q2：研究两个变量之间的关联。这时候需要用到Spearman秩相关或Kendall相关，在SPSS中通过分析-相关-双变量相关实现。

Q3

：研究两个变量之间是否存在线性变化趋势。比如，研究某疾病的严重程度是否随年龄增加而加重。这种情况，需要借助趋势卡方，英文为Linear by Linear Association，中文SPSS常写作线性关联。

三、两个变量，只有一个变量是有序多分类的

研究者针对这种资料常常会提出这样的问题：

Q1

：比较多个样本的构成比。比如，不同年份某医院体检人员甲状腺疾病的检出情况，其中年份是有序多分类变量，甲状腺疾病的检出情况(异常、可疑、正常)是无序多分类变量。这种情况下可以采用卡方检验。

Q2

：比较不同组间某指标的变化。比如，研究不同用药组患者的疗效，其中用药组是无序多分类变量，疗效为有序多分类变量。采用卡方检验和非参数秩和检验都是可以的，但一般推荐非参数检验，因为非参可以告诉我们哪一组疗效较好，哪一组疗效较差。

四、小结

说这么多，两个多分类变量，应该选用什么统计方法，这里总结一下，供大家参考。

两个变量都是无序多分类时，选用卡方检验。

两个变量都是有序多分类时，可选用卡方检验或非参数检验、趋势卡方检验或者秩相关、Kendall 相关。

只有一个变量是有序多分类时，选用卡方检验或非参数秩和检验。

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](http://www.iikx.com)转发