



四格表卡方检验不同组频数的简便计算

简便计算

曹守波
哈尔滨医科大学附属医院

四格表卡方检验是我们研究分析中常用的方法之一，比如说，分析某个蛋白或 RNA 的表达与临床病理特征的关系。而实际最困扰我们的可能是分析过程中不同组频数的计算，我一开始都是用最原始的方法来计数不同组的频数，相对比较繁琐，接下来我为大家介绍一下如何利用公式快速计算出各个组的频数。

比如，我们想要比较蛋白 A 的表达与疾病分期是否相关，其中，蛋白 A 阴性表达记为 0，阳性表达记为 1。疾病分期 IIIA 期记为“0”，IIIB 期记为“1”，而此时蛋白 A 阳性和阴性表达、IIIA 和 IIIB 期的频数是已知的。如果蛋白 A 在前，疾病分期在后，两两组合就是“0 0”、“0 1”、“1 0”、“1 1”这 4 种情况，只要记录好这 4 组数据的频数就可以分析蛋白 A 的表达是否与疾病分期相关。

首先，在 C 列记录一行二者关系，然后利用公式进行计算（图 1），在第二行第三列 (C2) 输入“=”，此时也就是选择公式进行计算，然后输入“A2+B2”，或者你输入“=”后，左侧的空白区域会出现函数这一栏，你选择“sum”求和，然后数值 1 输入 A2，数值 2 输入 B2，具体操作请参考图 1a-b，也就是说 C2=A2+B2，点 Enter 键会有相应结果出现，然后点击 C2 这一栏，待其右下角变为 +，点击下拉菜单至最后一行数据，然后就会出现如图 2 所示的数据，

也就是整个 C 竖列对应的结果均为同一行 A 和 B 数值之和，此时的 C 竖列结果均为公式结果，这种分析方法在数据比较大的时候会显现出其独特的优势，会节省很多时间。

分析数据时，我一般会选择把 C 竖列结果复制，然后点击选择性粘贴，粘贴为数值，或直接点击粘贴为数值，这样就会产生 D 竖列的结果，此时的结果只是纯粹的数字结果，不带有公式，这样 D 竖列的结果就会有 0,1,2 共 3 组，其中 2 代表的是蛋白 A 阳性、IIIB 期，0 代表的是蛋白 A 阴性、IIIA 期，这时候选中 D 竖列，点击查找，输入 0，然后就会出现结果为 0 的频数，也就是蛋白 A 阴性、IIIA 期的频数。相应的，输入 2，你就会得到蛋白 A 阳性、IIIB 期的频数，如图 3 所示。

由于蛋白 A 阴性和阳性以及 IIIA 期和 IIIB 期的频数均为已知，如果你知道其中任意一组的频数，剩下 3 组的频数可以迅速得到。但是有一点需要注意，由于我在 C、D 竖列输入的都是汉字，所以不存在这个问题，如果你输入的内容包含数字，且数字和其中两组有交叉，比如说同样一组数据，只是第一行的标识改一下，A 竖列记为蛋白 2，D 竖列记为蛋白 2 与疾病分期关系，这时候你点击查找 2，查到的频数是包含 D 竖列中你输入的蛋白 2 中的数字 2 的，如图 4 所示，记得要把这部分出现的频数去掉才是

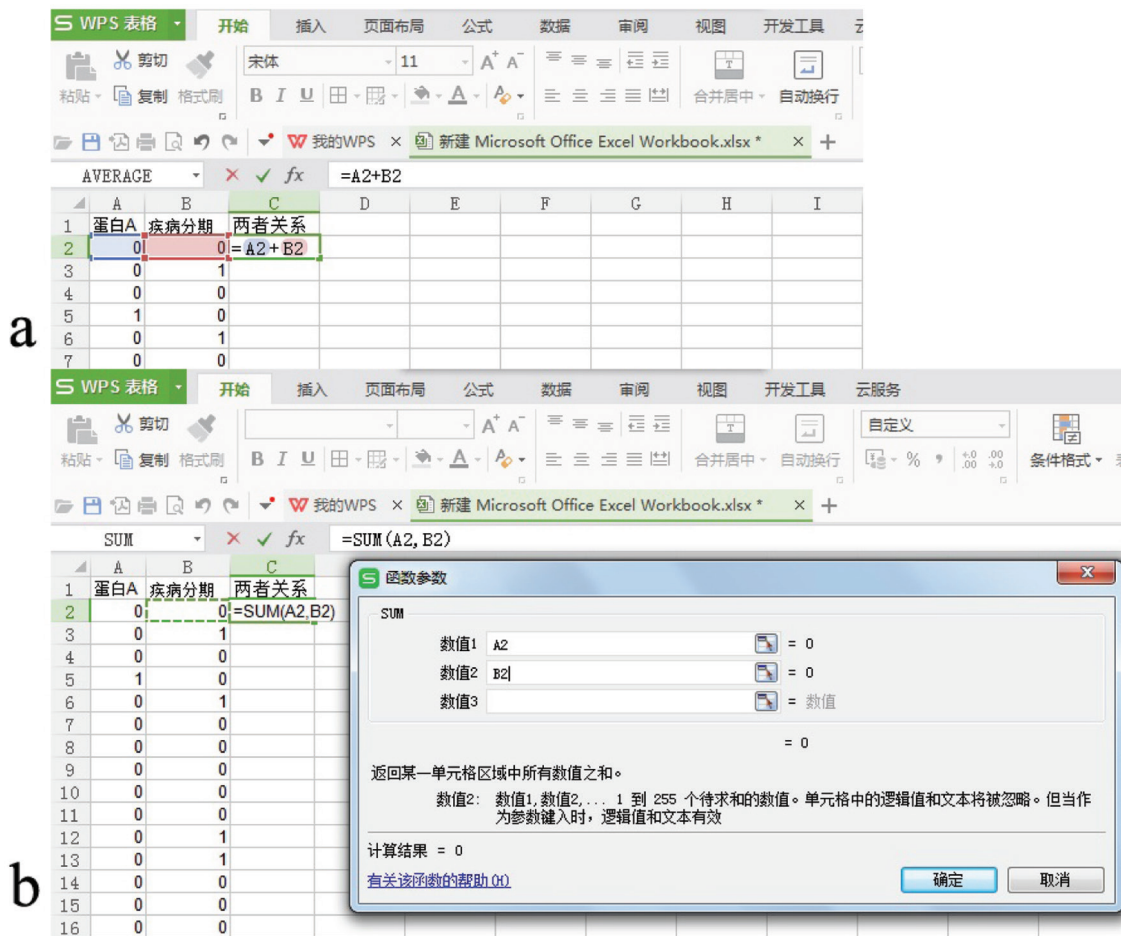


图1

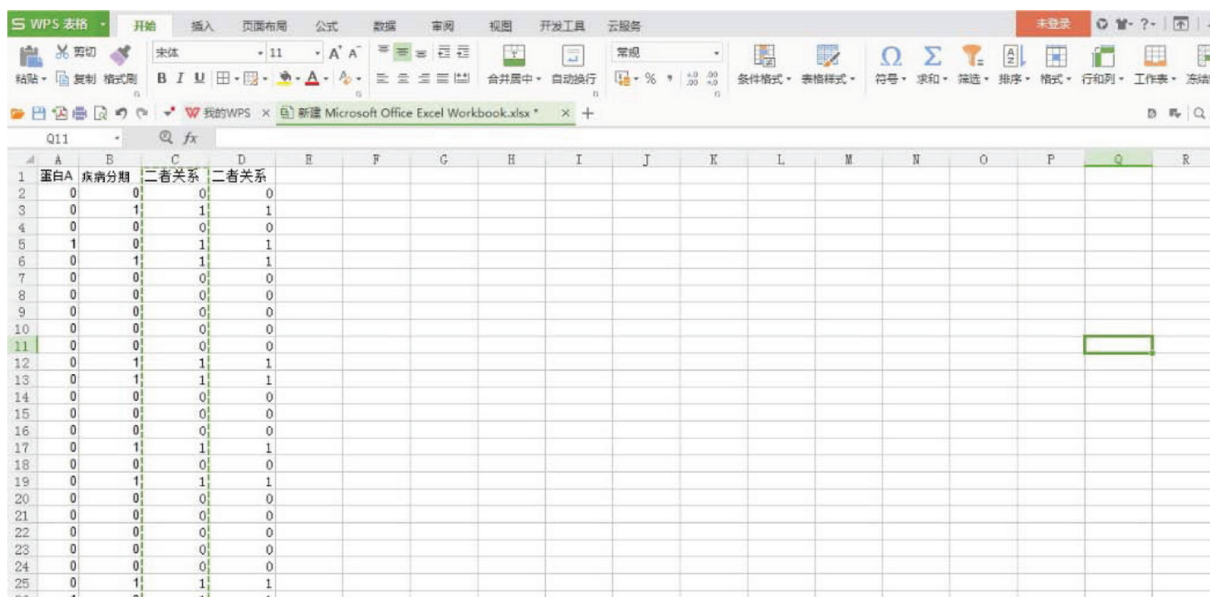


图2

你想要分析的结果，所以尤其注意。但是如果你只选择在数字这一竖列查找就不会出现这个问题了。

以上就是我个人利用公式快速计算不同组频数的方法，欢迎大家批评指正。

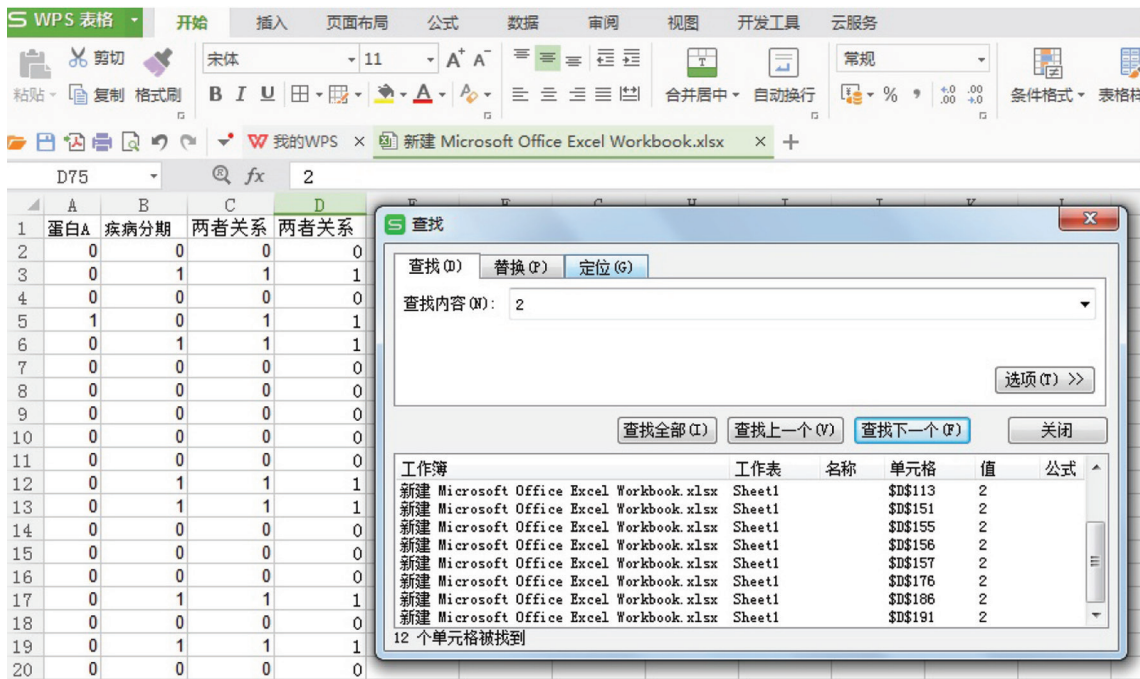


图3

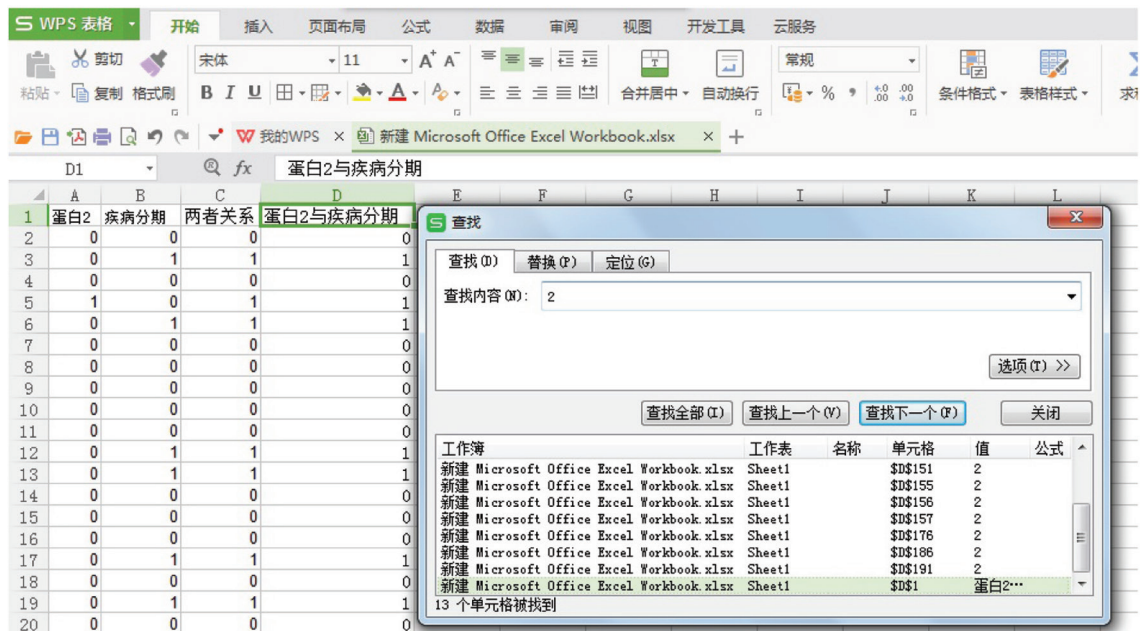


图4