**典型相关分析**

**1202110887 徐鸿飞**

**题目**（来源网络）：某个研究人员收集了600名大学新生的三个心理变量，四个学术变量（标准化考试成绩） 。他希望研究者3个心理变量与4个学术变量间的相关关系。

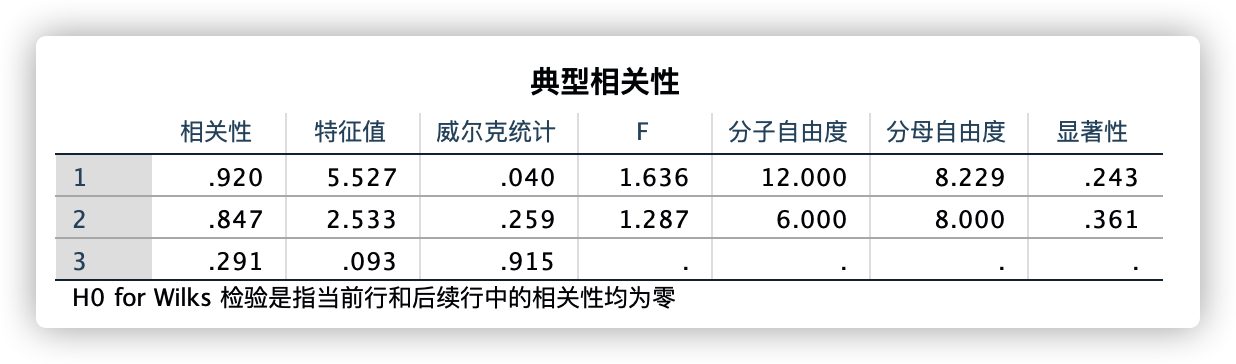


解答：



此图反映了各变量间的相关系数，从中可以看出不同变量间的相关程度。

如果组内变量间的相关系数高，说明两者包含的信息有重叠部分；如果组间变量相关系数高，则说明两者有一定相关性。



此图给出了典型相关系数及其检验，结果表明前两个典型相关系数是显著的，因此我们选择前两个典型相关变量进行解释。

具体来说，第一对典型相关变量的相关系数是0.446，p< .001；第二对典型相关变量的相关系数是0.153，p= .025。



上图分别是两组变量的标准化相关系数和未标准化的相关系数。

根据此图，可以写出各典型变量的表达式，如对于第一对典型变量u1和v1：

其标准化的表达式为（Z外向倾向表示将该变量标准化后的值）：

u1 = 1.963\*Z外向倾向-1.474\*Z自我概念-0.082\*Z动机水平

v1 = 0.857\*Z阅读成绩-0.847\*Z写作成绩+0.321\*Z数学成绩+0.342\*Z理科成绩

非标准化的表达式为：

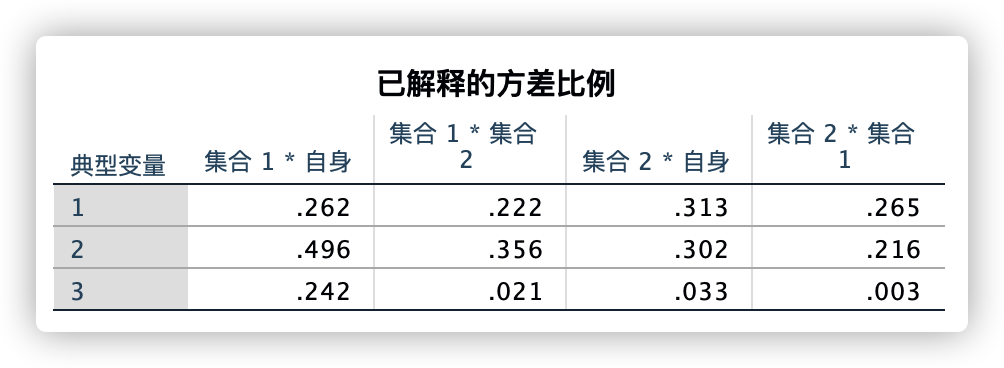
u1 = -2.687\*外向倾向-3.267\*自我概念-0.388\*动机水平

v1 = 0.091\*阅读成绩-0.083\*写作成绩+0.031\*数学成绩+0.041\*理科成绩

标准化的系数由于经过标准化，因此系数相互之间是可比的，用处是用于比较不同自变量对应变量的影响程度。比如在集合1的标准化系数中，外向倾向的系数是1.963，自我概念的系数是-1.474，因此我们可以认为外向倾向对成绩的影响比自我概念影响更大。

而未标准化的系数因为每个变量没有标准化，量纲不一样，因此不能直接用系数大小比较自变量贡献程度，它的用处是可以用于计算CCA得分,（直接用系数乘以原始数据）





上图是冗余分析的结果，它说明各典型变量对各变量组方差解释的比例。