**信息测度**

——《信息论基础》第三次上机作业

科研人员都终有一天会将自己生活的重心从学术工作中移走，或因为正常退休，或因为转向行政工作，或因为家庭需要更多时间投入，或单纯因为改换了职业……简单起见，我们称这种重心的转移为其从学术界“退出”。

学术论文是基础理论研究人员最重要的成果，那么停止发表任何学术论文必然是上文提到的这种退出的一个典型表现，但反过来则不然：退出者在退出之后未必会停止一切论文发表。这里的情况较为复杂，如果其仍留在学界工作，即使本人已经无心科研，大概率也会偶尔参与学生或合作者的研究并署名发表论文；即使已经完全变更职业，最后的几项研究也可能在其进入新岗位之后才陆续发表（发表周期一般需要1-5年）……

典型退出者的成果数量走向图：自第13年之后发生了上文中提到的“退出

因此，在研究大批量科研人员的职业生涯轨迹时，为了避免潜在的偏差，我们不能将“退出”完全等同于“停止一切论文发表”，而是应该设计更为合理的测度规则。

**实验目标：请设计一个测度方案，以量化测度科研人员“退出”学术界的时间。**

附件包含100位科研人员生涯论文发表记录，每一行存储一位科研人员的信息，“sy”字段指示其发表第一篇论文的年份，“pub\_list”字段包含其逐年发表论文数量的列表。请借助这些数据来对你定义的规则进行验证和调整，最终请提交两项内容：1）判断“退出”的规则的说明2）根据这一规则，判断这100位科研人员的退出情况（是否退出？如果退出，在生涯第几年？）。

请于当天**17:00**前将**PDF文档**发放到邮箱： **XXX**