

法拟合回归模型来预测事件发展变化的因果关系，要么采用 Logistic 回归来预测结局事件是否发生。这些传统方法都是基于截面数据的静态分析，无法反映教师离职或留任状况随时间而改变的规律，也不能对观测期内不同时间截面上的变量取值变化给予正确的描述和解释，从而造成信息的损失，甚至引起系统估计的偏差。生存分析方法通过为预测变量设置相应的示性函数，对分布复杂的时间数据进行生存函数估计，不仅解决了传统方法的限制，揭示出传统方法无法得到的生存特征，而且也精确地反映了预测事件发展变化的动态过程。

本文将采用生存分析的寿命表法（Life Tables）计算不同知识结构的高职教师在各时点上生存函数的估计值（留任率），从时间维度上反映高职教师总体的生存状况；采用 Kaplan - Meier 法比较不同知识结构教师的留任时间差别；采用 Cox Regression 模型分析在控制其他因素的情况下知识结构对高职教师生存状况的影响。

（二）数据处理

定义高职院校教师留任时间为教师从进入该校到辞职离开（中间没有间隔）所经历的时间。为方便后续的数据分析和处理，本研究以“年”为单位，对高职教师的留任时间进行记录。教师离职称之为“失败事件”（failure event）。关于数据处理有两点需要说明：（1）数据删失（censor）问题。数据删失是生存分析中非常普遍但须加以控制的现象。主要有两种类型：左删失（left censoring）和右删失（right censoring）。前者是指事件在观测之前即已发生并持续至观测期内的样本，后者是指到观测期末仍未终止的样本。右删失问题对于研究分析没有影响，生存分析方法能够有效进行处理，但对于左删失问题目前仍缺乏有效的解决手段，大多数研究采取舍弃左删失观测值的做法。本研究的样本为 2007 ~ 2012 年高职院校离（在）职教师，进入观测期内的样本都记录了明确的起始时间（即入校时间），因此可以恰当地处理左删失问题，避免了教师的留任时间被低估。（2）多个持续时间段（multiple spells）问题。这是指高职教师如果在学校持续工作一段时间，离开学校后（至少一年），有可能再次返回该校工作，因此同一名教师可能存在多个留任时间。我们采取的处理方法是将同一名教师的多个留任时间视为相互独立的留任时间段。

本文的样本数据来源于广东省某高职院校的人

事档案，主要分为两部分：一是该校离职教师数据，共 221 例，留任时间为教师入校时间与离职时间的差值；二是该校在职教师数据，共 373 例，留任时间为教师入校时间与研究截止时间的差值。由此，我们得到了 594 个观测样本。接着，我们对高职教师的留任时间进行统计，为每一个时间段定义了结局变量（out variable），并对多个持续时间段进行了标记，最终的统计分析结果如表 1 所示。

表 1 高职教师连续留任类型

连续留任类型	观测值 个数	百分比 (%)	累计百分比 (%)
连续在职 1 年以内	142	23.9	23.9
连续在职 1 ~ 3 年	238	40.1	64.0
连续在职 4 ~ 6 年	129	21.7	85.7
连续在职 7 ~ 9 年	66	11.1	96.8
连续在职 10 年以上	19	3.2	100
总 计	594	100	—

三、高职教师留任时间的分布估计

（一）高职教师留任时间的总体情况

表 2 是描述性统计分析结果。第 1 行针对全部样本，给出了高职教师留任时间的均值和中值，分别是 3.05 年和 2.00 年；第 2 行将分析样本局限于离职教师，此时的均值和中值都有所下降，分别降至 2.03 年和 1.00 年；第 3 行选择了在职教师样本，相对于离职样本和全样本而言，留任时间均有所提升，均值和中值分别上升至 3.65 年和 3.00 年。整体而言，离职教师的留任时间普遍较短，并拉低了全体教师留任时间的均值和中值。

表 2 高职教师留任时间（年）的描述性统计

变量名称	均值	中值	标准差	频数	百分比 (%)
全样本	3.05	2.00	2.870	594	100
离职教师	2.03	1.00	2.301	221	37.21
在职教师	3.65	3.00	3.003	373	62.79

（二）高职院校不同知识结构教师的留任情况

表 3 的上半部分反映了高职院校不同教育背景教师的留任情况。这里，我们将高职教师留任率定义为高职院校在职教师人数与总人数之比。总体而言，高职教师留任率普遍偏低，平均留任率为 62.79%，其中最低为高中及以下学历教师，为 39.29%；最高为大专学历教师，为 68.92%。数据结果与高职院校近年来教师队伍建设的总体发展情