《信息论基础》第四次上机作业

——信息分析：相关关系与因果关系

文件ReturnSeaR.csv包含2400余条发生过跨国流动的科学家的信息。其中各字段的含义分别如下：

* ORCID：科学家的ORCID编号，类似于“身份证号”
* last\_phd：流动前是否刚博士毕业
* tota\_pub：流动前发表过的论文数量
* to\_China：是否流动至中国大陆
* age：流动时的学术年龄
* conn：该科学家与其流入的机构的关联强度
* is\_Chinese：该科学家是不是中国人
* work\_diff：流动前后的业绩变化

需要解决的问题是：

相比于流动至其他国家（地区）的科学家，流动至中国大陆的科学家的业绩是否变得更好？如果更好，好了多少？

请以work\_diff为被解释变量，以to\_China为解释变量，以其他可以使用的变量为控制变量，建立多元线性回归模型回答以上问题。

请于当天**18:00**前将**PDF文档**发放到邮箱： informatics\_course@163.com。请将文件命名为:学号\_姓名\_HW4。