

《数据分析与处理技术》
南京审计大学2015级物流管理专业

作业2

2017.9.18

1-练习使用包中的命令

- 线性规划一直在管理学当中发挥着极为重要的作用，R语言当中有一个著名的程序包名叫Rglpk，专门用来计算线性规划问题。
 - 要求：从运筹学课本上找一道线性规划问题，加载Rglpk包并使用其中的命令解决问题
 - 参考资料：`help(package="Rglpk")` #注意，需安装该包后才能够看到说明文档
 - 关联包：该包依赖slam包才能够运行，即需先加载slam包
 - 另：不得使用附录中的指导案例作为作业

2数据集的操作

经理人	晋升日期	国籍	性别	年龄	Q1	Q2
William	10/24/08	US	M	32	5	4
Sarah	10/28/08	US	F	45	3	5
Debby	10/01/08	UK	F	25	3	5
Johnson	10/12/08	UK	M	39	3	3
Rosa	05/01/09	UK	F	99	2	2

上表是一张管理人员状况的记录表，请使用R语言创建该表，数据集名称自由设定，变量名称可依次转化为任意缩写形式.

- 请将“晋升日期”设为日期型变量，“国籍”“性别”设置为因子型变量
- 将“经理人”属性设置为数据集的对象名称，即row.names
- 删除“年龄”属性为99的对象
- 在剩下的记录数据当中，计算晋升日期最大差距为多少天
- 令Q1和Q2相加求出每个经理人的总分，并将总分变量插入数据集中
- 提取出其中的男性管理人员的数据，存入新变量MaleManager，求男性管理人员总分平均为多少

3作业-数据导入

中国国家调查数据库中开放了大量关于中国经济、社会方面的基础情况调查数据，其中有一个关于大学生成长的项目《大学生成长追踪调查（2009）基线调查》，请自行下载它的数据文件和调查问卷，参照调查问卷理清数据中的变量含义，并对数据进行导入和简单分析。

摘要： 大学生成长追踪调查开始于2009年，当年成功访问了通过科学抽样程序从北京地区的公立高校中抽取的15所高校5100名在校生中的4771人。随后在2010、2011、2013年又对2009年成功访问的学生进行了追踪访。研究通过反映与中国高等教育及中国大学生相关的问题与规律。从大学生的来源与大学的进入、大学生的学业、大学生的日常生活、大学生的课外活动、大学生的社会观念与行为、大学生的心理健康情况、大学生的爱情观与恋爱行为、大学生的职业与人生规划、大学生的毕业出路等方面进行描述和探索

3作业-操作要求

- 翻阅PPT，根据课堂讲解练习导入数据
- 将原始数据集赋予变量mydata
- 后续操作请不要直接改变mydata,计算过程中产生的数据可以设置新的变量进行保存
- 请对照问卷调查表弄清楚mydata中变量的含义
- 作业请写出操作的代码，以及你的分析结果
- 不接受电子稿作业

3作业-问题：

1. mydata中一共多少属性（变量），多少调查对象？
2. 该调查数据集中男女比例为多少？
3. 数据集的样本中有多少人在大学期间存在抽烟习惯？
4. 第一次谈恋爱比例最高的时间段是大学第几年？没有谈恋爱的人占多少比例？

提示：先抽取出需要的对象，该过程中数据集索引、取子集等方法都可以实现，方法不唯一；再对抽取出的对象进行处理和分析；注意缺失值的存在。