

定量研究方法课程作业（R语言操作1）

1. 在中国综合社会调查网站（<http://cgss.ruc.edu.cn/>）上下载cgss2015的数据（spss版本）。
2. 在Rstudio中将数据导入，选择样本类型（s1）、居住省份（s41）、个人全年总收入（A8a）、出生年份（a301）、性别（A2）、教育程度（A7a）、政治面貌（A10）、工作年限（A59c）、工作经历及状态（A58）、父亲教育程度（A89b）、母亲教育程度（A90b）、父亲的工作单位类型（A89g）、母亲的工作单位类型（A90g）等变量构成新的工作数据集cgssincome，再根据样本类型（s1）变量选取城市的被调查对象（s1值不为2）并保存（注意：如果采用菜单方式载入，haven包会默认给变量添加值标签，应采用zap_labels函数去除值标签，否则有些函数无法使用，如cgssincome <- zap_labels(cgssincome)）。
3. 新建一个项目，载入cgssincome数据集，将出生年份转化为年龄（2015年），教育程度（本人、父母）重新进行编码转化为教育年限，政治面貌重新进行编码转化为是否党员。并且设定所有的缺失值为NA（CGSS设定为-8是无法回答；-3拒绝回答；-2是不知道；-1是不适用），最后成行删除掉具有缺失值的观测（建议使用cgssincome <- na.omit(cgssincome)）。
4. 将性别、是否党员转化为类别变量（因子factor函数），计算男女党员的比例列联表；制作收入的直方图(注意用参数breaks多分割一些区间)，收入是正态分布吗？如何让它变得更像正态分布？制作收入的箱线图，选取收入小于15万的观测再试试看；制作男女收入对比箱线图，收入存在性别差异吗？
5. 计算在置信水平为95%的条件下，居民收入的置信区间。
6. 假设有人提出2015年全国居民年平均收入为5万块，那么在0.05的显著性水平下，cgss的样本结果和这一提法一致吗？
7. 计算男性与女性的平均收入，在95%的置信水平下，收入是否存在性别差异？
8. 检验被调查者的父亲与母亲教育程度是否有差异？