homework2

Chiffon

7 October 2016

### 阅读内容

* R语言编程艺术
* 6-8,10章
* or数据科学中的R语言
* 3,5,6章

### 1. 薪水数据分析

利用ISLR包的Wage数据进行计算:

* 不同职业的人群收入有没有显著差别？
* 人们的健康和其职业是否有关系？(chisq.test)

### 2. 抓取豆瓣所有的250部最佳电影

网址:

[<http://movie.douban.com/top250?start=50&filter>=&format=text](http://movie.douban.com/top250?start=50&filter=&format=text)

### 3. 描述性数据分析

1. 从网页中下载数据并命名为snee <http://www.biostatisticien.eu/springeR/snee74en.txt>
2. 一共有多少个观测值？有多少个变量？变量的类型是什么？
3. 对每个变量制作一个单变量描述研究，数值的结果.
4. 我们准备研究eyes和 hair两个变量的相关性，创建一个二变量的频数表。
5. 计算各个不同水平下的眼睛与头发颜色占总体的比例。
6. 假定两个变量是独立的，计算每个水平下频数的理论值。
7. 比较观测的频数表和理论的频数表的区别（对每种情况计算误差的平方
8. 用卡方检验来检验独立性

### 4. 解析该网页JSON

解析该网页JSON,制作如下的表格,保存到csv文件

[<https://secure.bmw.com.cn/cn/_common/_js/dealer_locator/dealer_locator.json>](https://secure.bmw.com.cn/cn/_common/_js/dealer_locator/dealer_locator.json)

str(output)

'data.frame': 573 obs. of 9 variables:  
 $ city : chr "南京" "南京" "南京" "台州" ...  
 $ pvid : chr "江苏省" "江苏省" "江苏省" "浙江省" ...  
 $ shortName: chr "南京宁宝" "南京星之宝" "南京宝利丰" "台州好德宝" ...  
 $ fullName : chr "南京宁宝汽车服务有限公司" "南京星之宝汽车销售服务有限公司" "南京宝利丰汽车销售服务有限公司" "台州好德宝汽车服务有限公司 / 温州市好达机电有限公司台州分公司" ...  
 $ address : chr "江苏省南京市秦淮区大明路278号" "江苏省南京市江宁区东麒路" "江苏省南京市栖霞区尧化街道马家园106号" "浙江省台州市路桥区西迎宾大道11号" ...  
 $ url : chr "http://www.ningbao.bmw.com.cn" "http://www.njxingzhibao.bmw.com.cn" "http://www.njpolyfine.bmw.com.cn" "http://www.tzhaodebao.bmw.com.cn" ...  
 $ lon : num 119 119 119 121 121 ...  
 $ lat : num 32 32 32.1 28.6 28.6 ...  
 $ service : chr "5S 金融服务 预约快修通道服务 BMW 尊选二手车 M展区 大客户中心 大客户服务 插电式混合动力 BMW明悦晚间服务 BMW i" "4S 金融服务 预约快修通道服务 BMW 尊选二手车 大客户服务 插电式混合动力" "4S 金融服务 预约快修通道服务 大客户服务 插电式混合动力" "4S 金融服务 预约快修通道服务 BMW 尊选二手车 大客户中心 大客户服务 插电式混合动力" ...