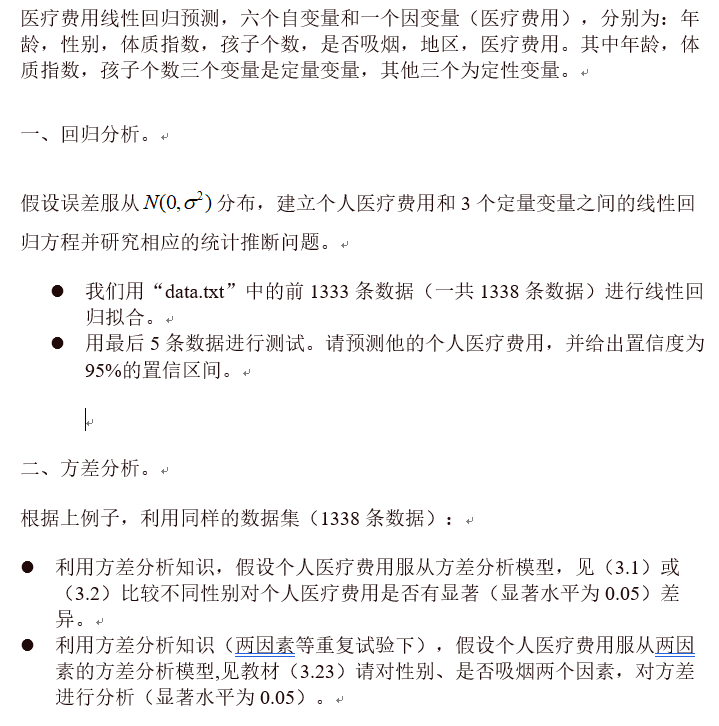
**统计分析方法第二次作业**

16337063 宫凯凤

**实验内容**



**实验结果**

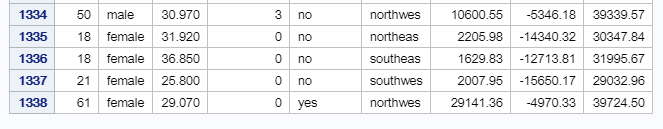
Hw1：



CHARGES = 237.74407 \* AGE + 333.74999 \* BMI + 546.27972 \* CHILDREN – 6872.96706

最后五组的个人医疗费用估计以及95%的置信区间如下图的最后三列所示。





Hw2：



由于（Pr > F）为0.0361，小于0.05，所以不同性别对个人医疗费用有显著性差异。

Hw3：



由于SEX的一个（Pr > F）大于0.05，其余两个因素的（Pr > F）小于0.05，所以是否吸烟对个人医疗费用有显著性影响，而性别以及性别和是否吸烟的交互效应对个人医疗费用没有显著性影响。

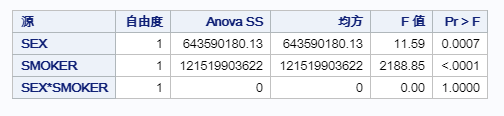
**实验心得**

我这次实验改用了SAS，首先在安装上就遇到了问题，我是照着老师的PPT做的，从那个百度网盘下载的安装程序，并且特意开了D盘（我的电脑原来只有C盘和E盘），但是安装完成后还是无法正常使用。打开中文版就会弹出一个log然后报错，打开英文版就是OLE未注册什么的。最后在舍友的帮助下在网上下载了大学版并用虚拟机打开了。

遇到的第二个问题就是大学版因为是在虚拟机上运行的，所以我给出的data.txt的原地址是在E盘的，然后就无法读取，后来通过查阅资料我把data那个数据拷贝到了SAS的myfolders文件夹下，并改了读取路径，才能正常读取（同时把“\”改为了“/”，因为根据log给出的错误信息，可以看出SAS是在Linux下运行的）。

然后就是读取txt文件时我发现读出来的结果和预期的不符，经过思考我认为是由于数据之间的逗号影响了SAS的读取，所以我决定将txt转化为csv后再进行尝试。

在第三题遇到了一个很诡异的问题，我的SEX\*SMOKER的结果全是1或0



后来经过仔细阅读课本第八章的内容发现，是由于函数调用错误，不应该继续用ANOVA，而应该使用GLM。因为给定的数据集中相同因素组合的情况下，数据的数量不相同。所以不能直接将其当成等重复实验来做。SAS中的GLM可以处理这种情况，而ANOVA则不能。不过如果使用ANOVA的话也可以做，只不过需要将部分数据进行截断，使其数量相同，不过这样处理起来比较麻烦，所以就直接改用GLM函数。

**参考文献**

<https://blog.csdn.net/qq_17534301/article/details/80882253>

<https://jingyan.baidu.com/article/fd8044fa2e0ff35031137a05.html>