

# 统计分析方法作业 2

姓名：孟衍璋 学号：16337183

## 题目 1：回归分析

假设误差服从  $N(0, \sigma^2)$  分布，建立个人医疗费用和 3 个定量变量之间的线性回归方程并研究相应的统计推断问题。

- 我们用“data.txt”中的前 1333 条数据（一共 1338 条数据）进行线性回归拟合。
- 用最后 5 条数据进行测试。请预测他的个人医疗费用，并给出置信度为 95% 的置信区间。

### 线性回归拟合：

首先使用 textscan 将 txt 文件中的内容读入矩阵中，再将三个定量变量存入数组之中，使用 regress 函数求出其线性回归的形式。

```
X = [ones(length(charges),1), age, bmi, children];  
[b,bint,r,rint,stats] = regress(charges,X,0.05);
```

其中变量 b 表示回归系数；

变量 bint 表示回归系数的区间估计；

变量 r 表示残差向量；

变量 rint 表示置信区间；

变量 stats 表示用于检验回归模型的统计量，有三个数值，相关系数  $r^2$ 、F 值、与 F 对应的概率 p。

其中回归系数如下：

|             |
|-------------|
| -6.8730e+03 |
| 237.7441    |
| 333.7500    |
| 546.2797    |

因此拟合出的曲线为：

$$charges = -6873 + 237.74 * age + 333.75 * bmi + 546.28 * children$$

预测：

用上述公式验证后 5 个数据得到 charges\_estimate：

|            |
|------------|
| 1.6989e+04 |
| 8.0597e+03 |
| 9.7051e+03 |
| 6.7304e+03 |
| 1.7332e+04 |

置信度为 95%的置信区间为（左为下界，右为上界）：

|            |            |
|------------|------------|
| 1.3601e+04 | 2.0378e+04 |
| 7.5052e+03 | 8.6143e+03 |
| 8.6534e+03 | 1.0757e+04 |
| 6.6612e+03 | 6.7996e+03 |
| 1.5176e+04 | 1.9487e+04 |

参考资料：

<https://wenku.baidu.com/view/139ec687b9d528ea81c779dd.html>

<https://www.mathworks.com/help/stats/regress.html>

## 题目 2：方差分析

根据上例，利用同样的数据集（1338 条数据）：

- 利用方差分析知识，假设个人医疗费用服从方差分析模型，见（3.1）或（3.2）比较不同性别对个人医疗费用是否有显著（显著水平为 0.05）差异。
- 利用方差分析知识（两因素等重复试验下），假设个人医疗费用服从两因素的方差分析模型,见教材（3.23）请对性别、是否吸烟两个因素，对方差进行分析（显著水平为 0.05）。

首先使用 textscan 将 txt 文件中的内容读入矩阵中，再将性别和是否吸烟两项分别提取出来，将他们转化为字符串数组。

### 单因素方差分析：

进行单因素方差分析使用 anova1 函数：

```
p1 = anova1(charges, sex_str, 'off');
```

得到检验 p 值：

```
p1          0.0361
```

小于给定显著性水平 0.05，所以可以认为性别对个人医疗费用有显著影响。

### 双因素方差分析：

进行双因素方差分析使用 anovan 函数：

```
p2 = anovan(charges, {sex_str smoker_str}, 'model', 'interaction', 'varnames', {'sex_str', 'smoker_str'});
```

得到检验 p 值：

|             |
|-------------|
| 0.0983      |
| 4.3266e-277 |
| 0.0030      |

性别因素的检验 p 值大于显著性水平 0.05，可认为没有显著影响；是否吸烟的检验 p 值远小于显著性水平 0.05，所以有显著影响；两者的交叉影响的显著性水平小于 0.05，所以可认为有显著影响。

参考资料：

<https://www.mathworks.com/help/stats/anova1.html>

<https://www.mathworks.com/help/stats/anova2.html>