2018春季《回归分析》研究课题

**注：**如果响应变量的个数比较少，可以逐个拟合模型。如果响应变量的个数较多，可以：

       1. 重新整合数据集：将全部响应变量拉直到一个变量Y中，在数据集中加入一个名义变量用于指示响应变量。如加入名义变量index, index="1"时，对应的Y是第一个响应变量的值； index="2"时，对应的Y是第二个响应变量的值，等等。这样只需要考虑Y和原先所有自变量以及index的建模就可以了。

       2. 将响应变量进行压缩：利用主成分或其他方法将响应变量压缩成少数几个综合响应变量，然后用这些综合响应变量对所有自变量建模。

**背景及统计的作用：**

洗衣粉是通过其中的化学成分溶于水后改变水溶液的物理化学性质来实现去污的作用的，因此通过测量洗衣产品溶于水后的溶液的一些属性就可以了解产品去污的功效。如果能建立溶液属性和产品功效之间的模型，就可以找出能够最大化产品功效的溶液的属性，根据这些属性和化工技术知识我们就可以找出最优的配方。

**试验：**

为了研究洗衣粉溶液的物理属性对去污功效的影响，我们分别测量了86个不同产品溶液的物理属性和它们的去污效果的数据

**已有的数据：**

1. 现有86个产品的物理属性及功效数据，请用这86组数据来建立模型
2. 每一个产品的21个属性作为输入变量 (PP1—PP21)
3. 产品在18种污渍上的功效作为输出变量 (O1—O18)



**要求：**

1. 请根据现有数据拟合出一个统计模型，模型能够基于产品的属性数据对产品的功效做出比较可靠的预测
2. 考虑所有输入变量的线性项，根据模型的需要选择它们的平方项及交互作用项；
3. 对此数据用多种不同的方法进行分析；
4. 选择合适的能够反映模型预测能力的评价准则（可以根据需要提出新的准则），并根据准则选出最优的建模方法和最优模型；
5. 再利用前10组数据说明拟合出的模型的预测能力。

**注意：**某些产品的某些属性有缺失数据。