**预测工作错误：基于累计工作量和当天工作模式**

算法目的：本算法的主要目的是帮助公司预测员工在使用进销存系统时创建的错误的销售单数量。将前期28天（包含当天）累计的工作量（所有任务的）和当天工作模式结合起来，共同预测错误的销售单数量。当天工作模式分为两个方面，一是员工在8:00到18:00里每小时创建的销售单数量的分布情况，二是员工在8:00到18:00里完成各类任务数量的分布情况。

数据说明：contactid：店铺名 userid：员工ID moduleid ：各任务模块，创建销售单对应数值为2 operatetime：操作时间

数据预处理：

1. 计算每一天工作量（workload）

2. 计算28天累计的工作量（Cumulative workload），比如2018年1月1日，则计算包含1月1日在内的和前面的27天的工作量之和，公式为，是工作者每天完成的任务数量

3. 计算变量workday,该天是否是工作日，是工作日则为1

4. 计算number of working days，即计算从最开始到当前工作日的天数

5. 计算time allocation，公式为，为员工每小时创建的销售单数量除以员工当天创建的销售单数量，时间是8:00到18:00,共10小时，n为10。假如总共完成10单，某小时完成了1单，则p为1/10。

6. 计算effort allocation,公式为，为员工完成某类任务占完成的所有任务数量的比值，moduleid里面不同的数字代表不同的任务，比如创建销售单3单，所有任务数量加起来为100，则q为0.03。

7. 计算number of task categories，计算员工当天完成的任务数量，当天完成10类任务，则该值为10

8. 处理year变量，若为2016年，则year1=1，为2017年，则year2=1，为208年，则year3=1，为2019年，则year4=1，不在这些年，则输入时year1、year2、year3和year4均为0。

8,。 处理month变量，为2月，则month2=1，为3月，则month3=1，为4月，则month4=1，为5月，则month5=1，为6月，则month6=1，为7月，则month7=1，为8月，则month8=1，为9月，则month9=1，为10月，则month10=1，为11月，则month11=1，为12月，则month12=1，若为1月，这month2-12均为0。

10. 对Cumulative workload，number of working days，number of task categories分别+1之后，取ln值得到：ln(Cumulative\_workload)，ln(number\_of\_working\_days)，ln(number\_of\_task\_categories)

输入1：ln(Cumulative\_workload)，ln(number\_of\_task\_categories)，workday，ln(number\_of\_working\_days)，time allocation，year，month

处理1：



输出1：得到，再求的值，即为预测的被创建的错误的销售单数量

输入2：ln(Cumulative\_workload)，ln(number\_of\_task\_categories)，workday，ln(number\_of\_working\_days)，effort allocation，year，month

处理2：



输出2：得到，再求的值，即为预测的被创建的错误的销售单数量

结果：根据输出1和输出2的结果，取两者均值，作为最终预测的被创建的错误的销售单数量