

说明文档-数据预处理

320180941980 薛睿卓

去除空文件

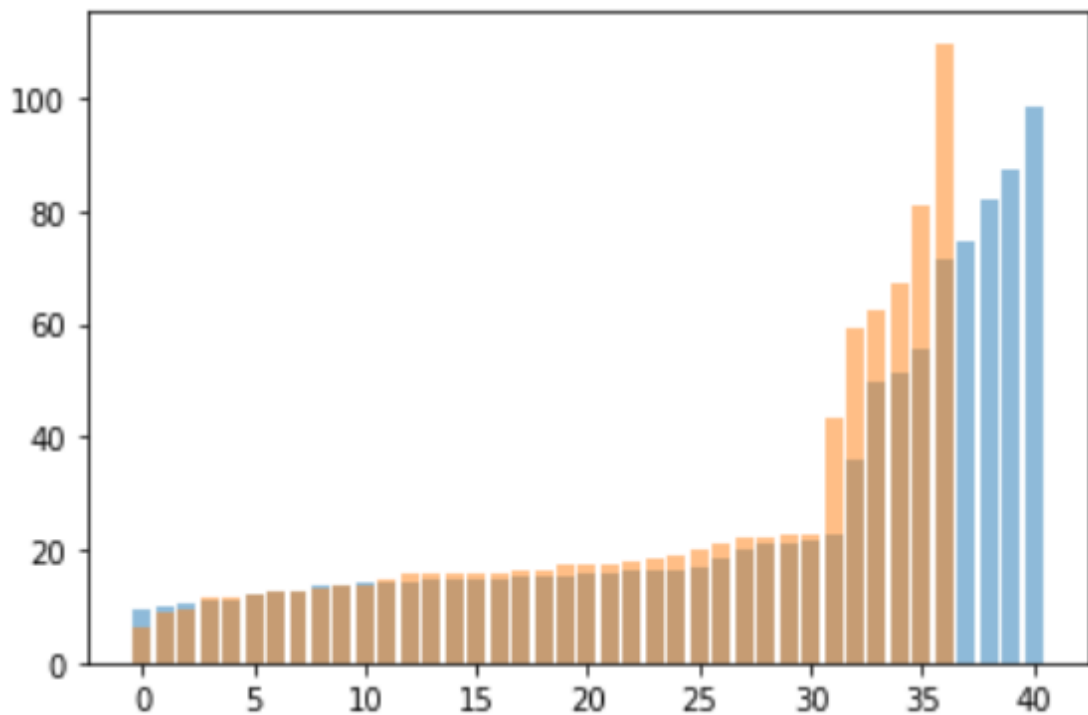
用read_json()方法读取下载后的文件，将文件转化为DataFrame对象，同时用empty方法判断该对象是否为空DataFrame，从而将空文件剔除。

画出时间分布对比图

按照 5 Hz的采样率计算出每份测试题对应的时间，以分钟为单位。然后将得到的数据排序并画出分布图。

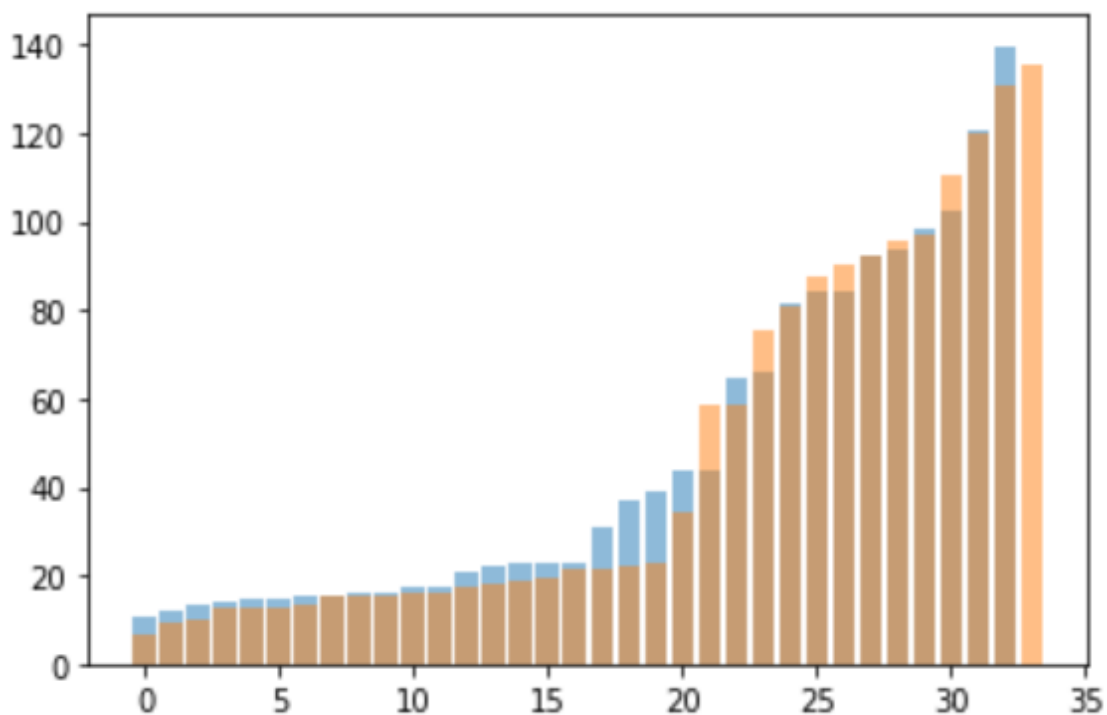
蓝色为health，橙色为anxiety

- accelerometer组



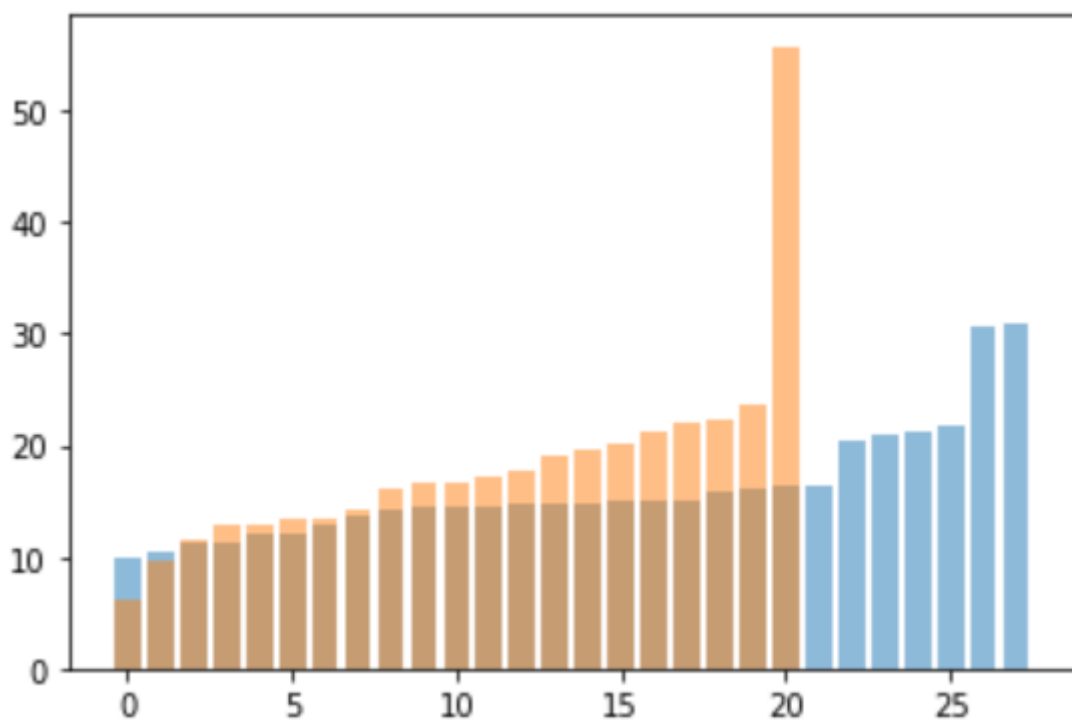
从对比中可以看出，健康者的三轴加速度要略低于焦虑症患者，且部分焦虑症患者做题时间格外长。

- device motion组



根据device motion的采样时间来看，焦虑症患者做题时间总体略低于健康者。

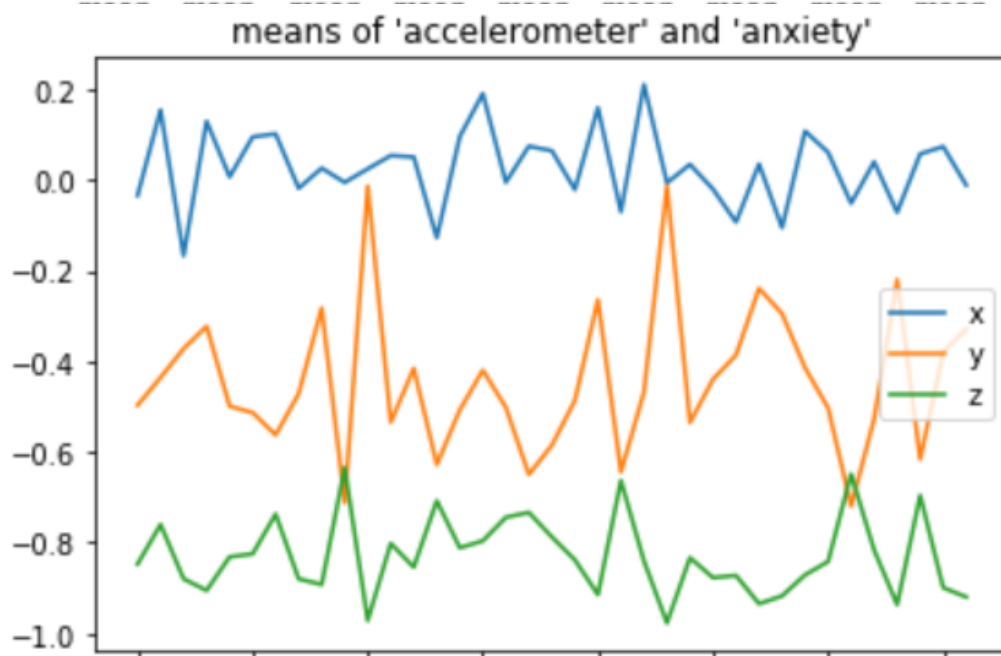
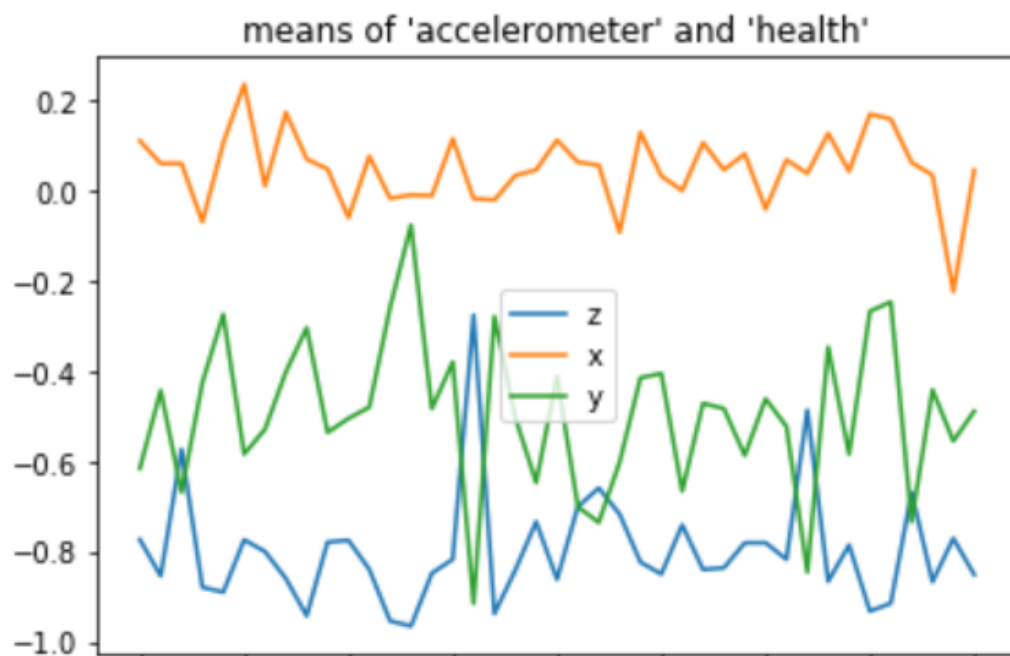
- gyroscope组



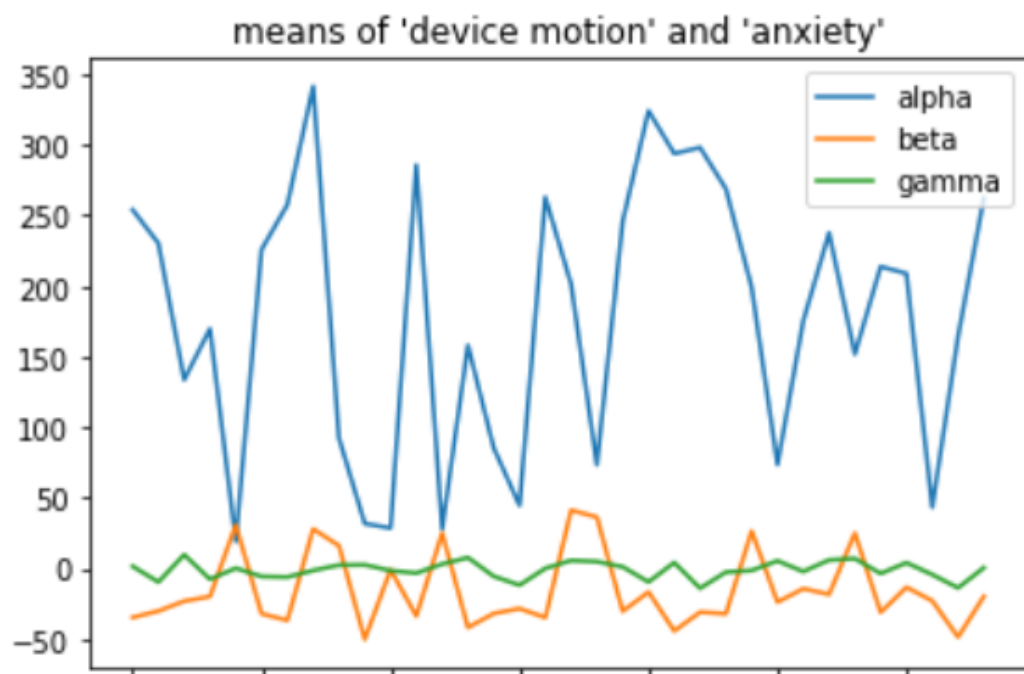
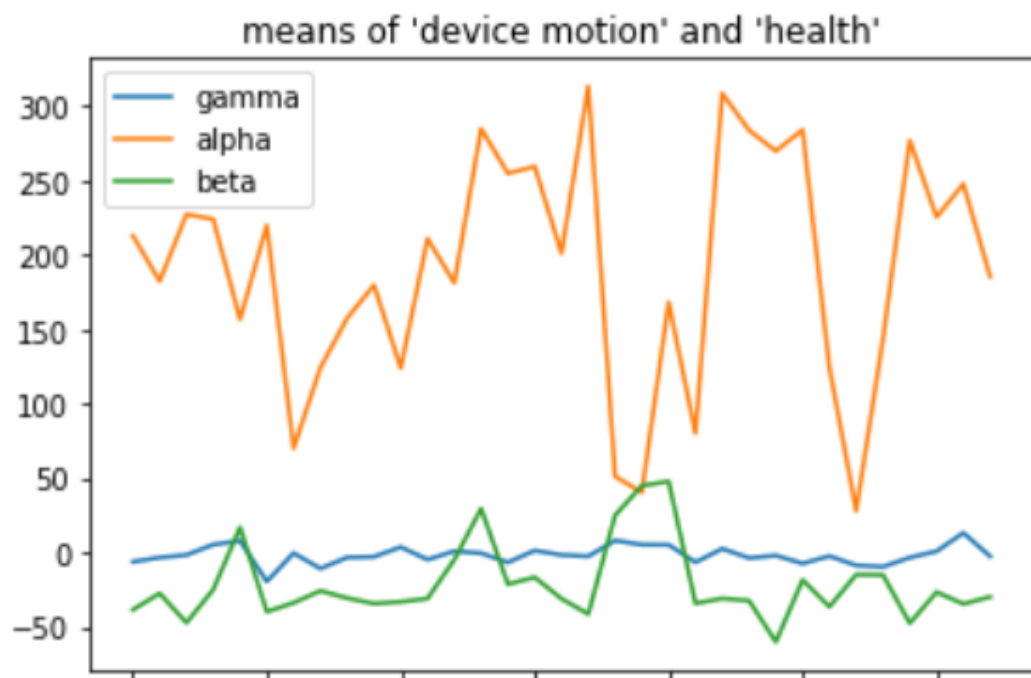
根据gyroscope组的采样时间看，焦虑症患者做题时间普遍略高于健康者。

画出各轴加速度的均值和标准差分布

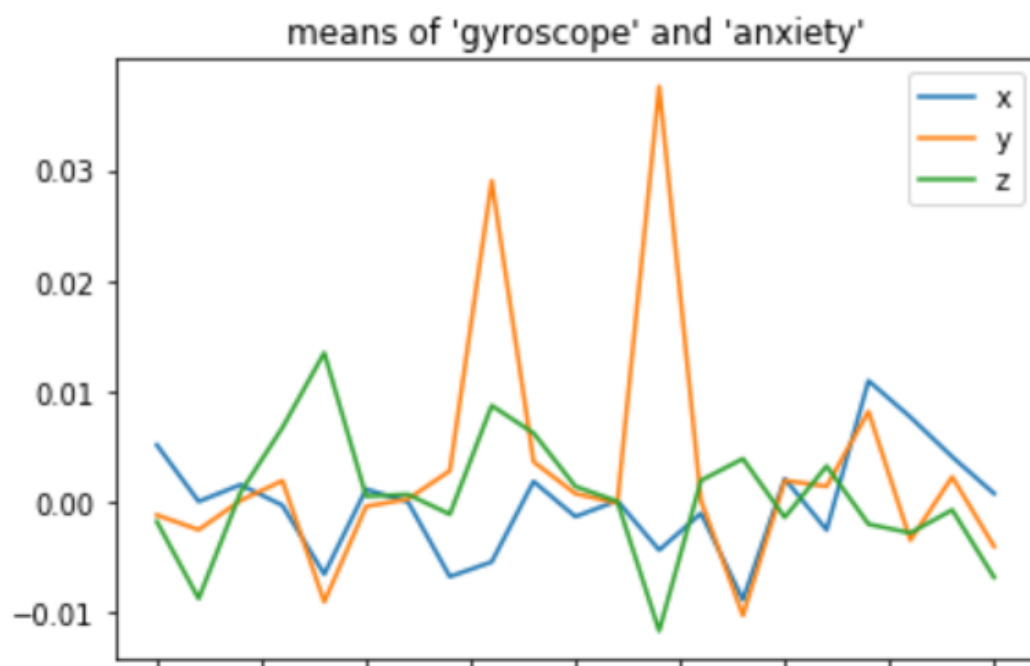
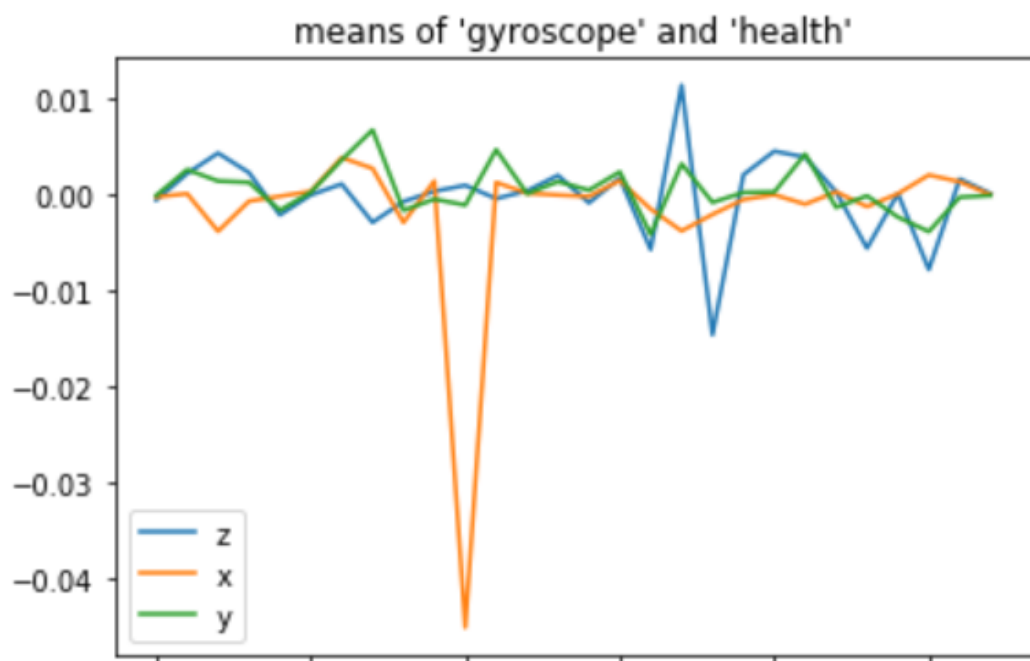
- 平均值
 - accelerometer组



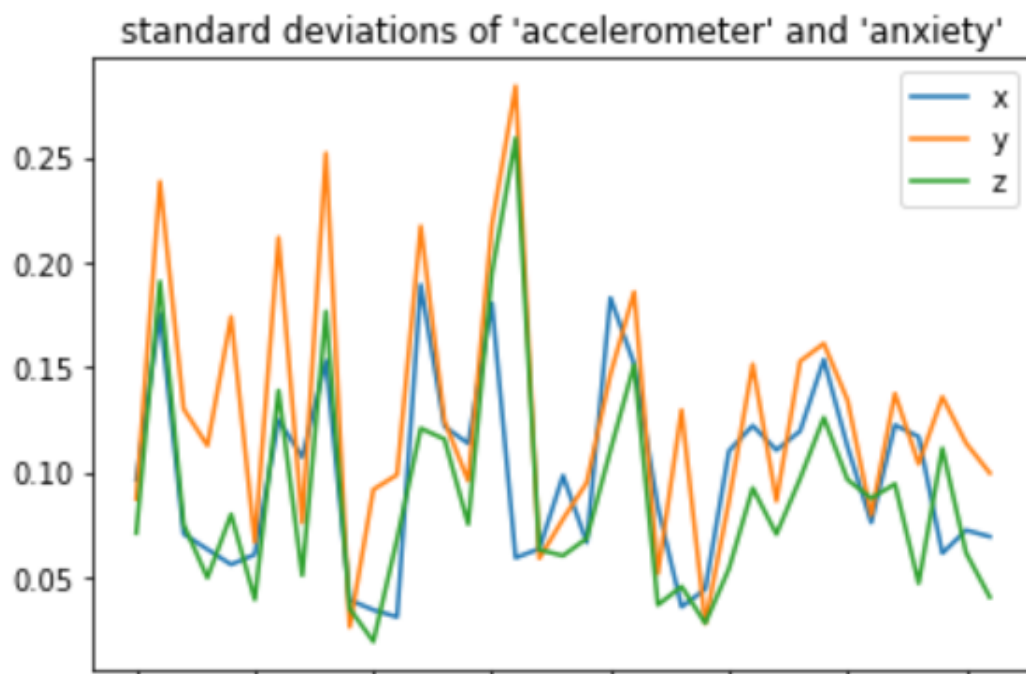
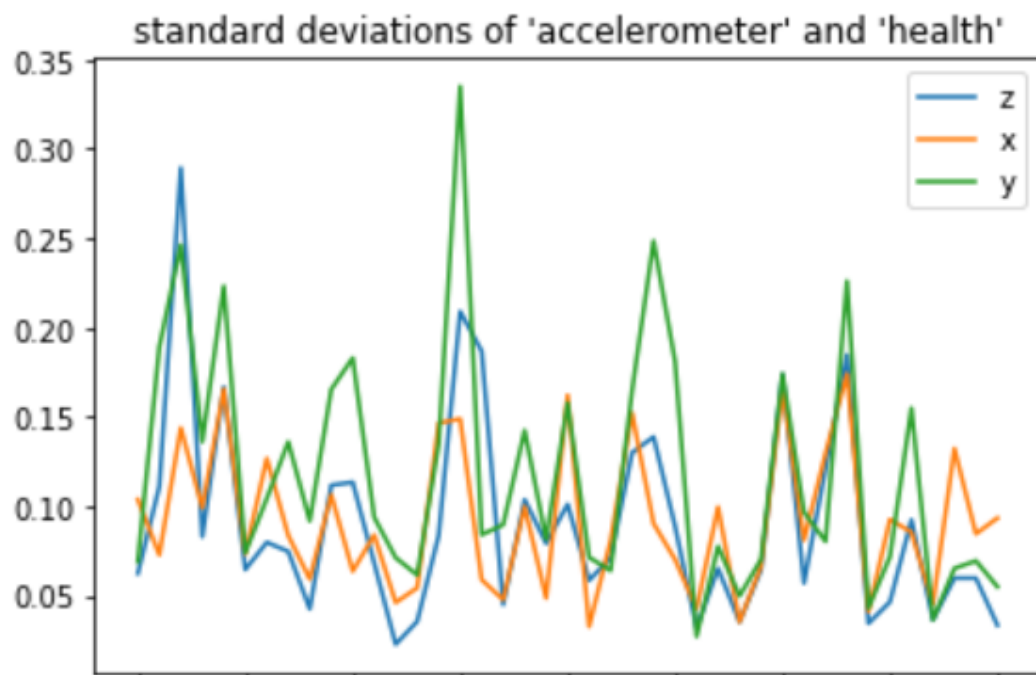
○ device motion组



○ gyroscope组

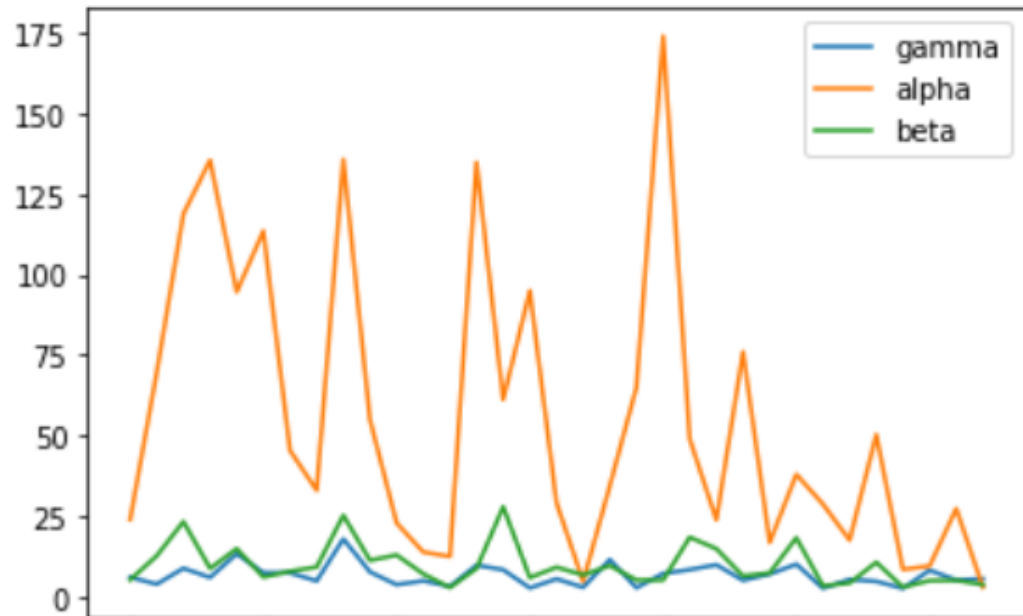


- 标准差
 - accelerometer组

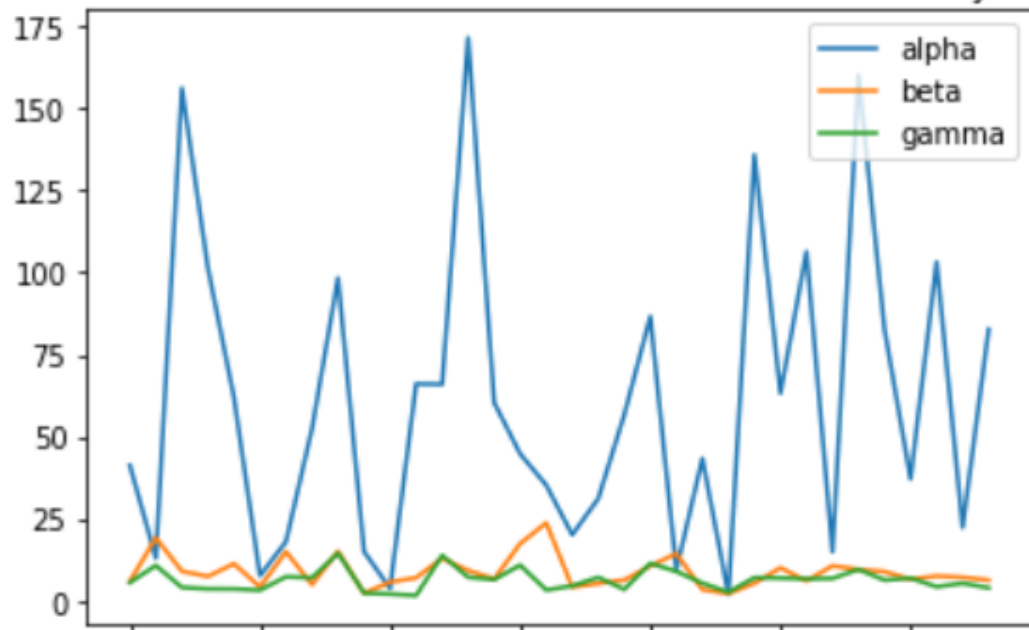


○ device motion组

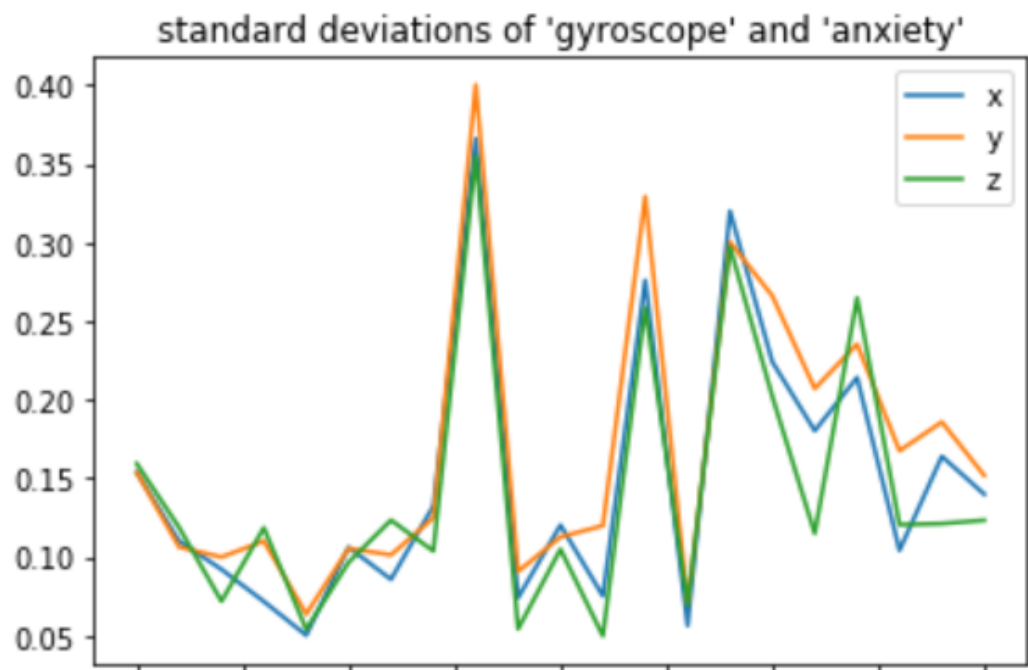
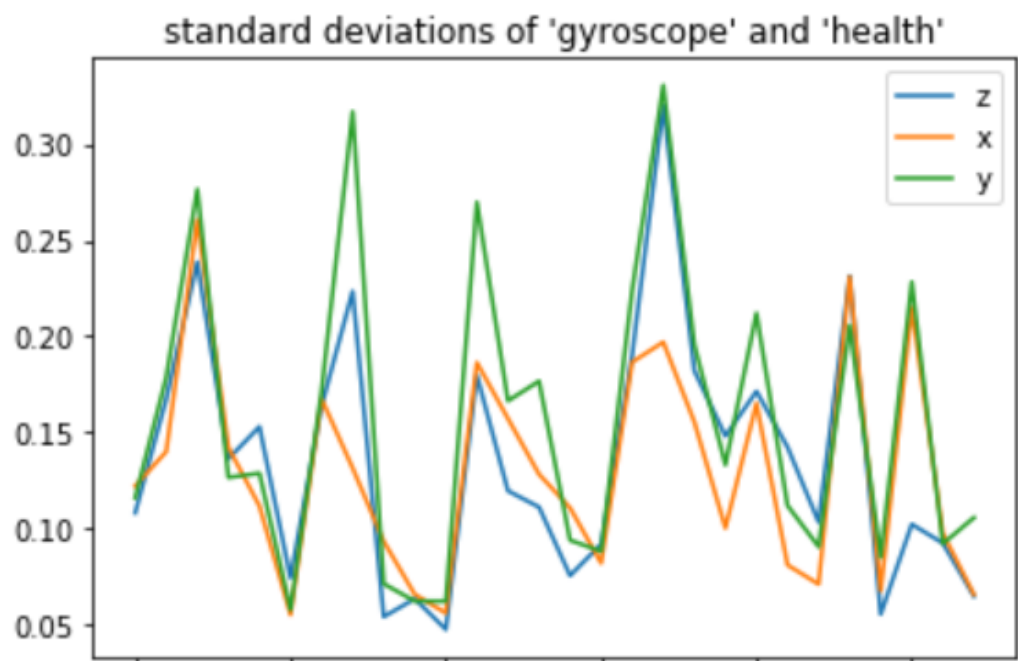
standard deviations of 'device motion' and 'health'



standard deviations of 'device motion' and 'anxiety'

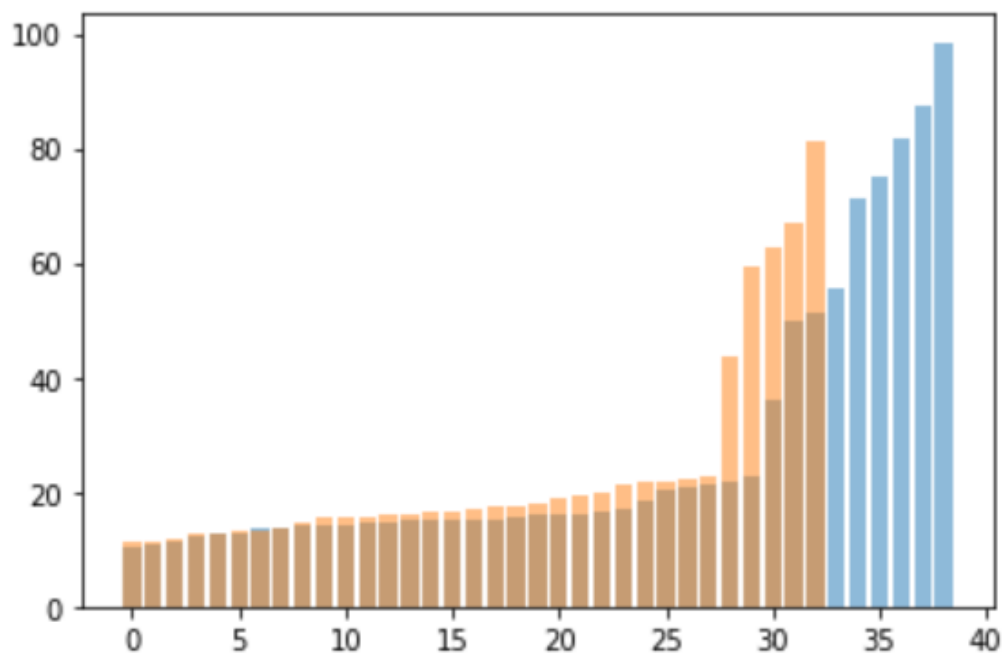


◦ gyroscope组

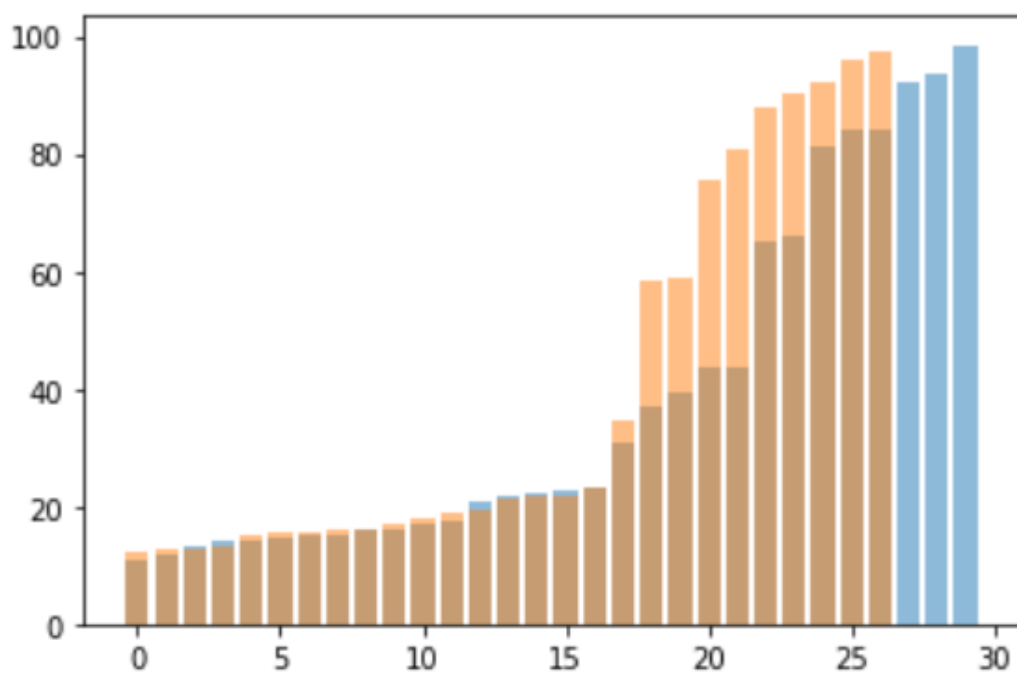


去除噪声

- 将做一份问卷的时间控制在10~100分钟内，得到如下结果：
 - accelerometer组



◦ device motion组



◦ gyroscope组

◦

