**规则一：针对于离散变量。**针对这类数据，在定义区间范围时有两种方式。第一类为已知存在于数据库内的信息。对于此类数据，建议将已知信息存放在不同的枚举类中，主控方传回的数值先于相对应的枚举类中的数据进行比较。若主控方传回数据与枚举类中数据不匹配时，系统记录并与后台数据库中已有数据进行比对且会自动抛出异常信息。此时应与相关部门人员或现场人员二次确认异常数据再进行修改。另一类为未知的离散数据。当接收到的数据总是为固定的整数值时，猜测为人为规定的数值，系统应先记录所有的信息值并存储于数据库内，再与现场人员或相关部门负责人进行确认。数据被确认后，定义区间与校验方法与第一类相同。

创建枚举类，Eg.

public enum WtWrongInfoEnum {

XXX("0","正常"),

XXX("1","待机"),

XXX("2","维护"),

XXX("3","中断"),

XXX("4","故障"),

;

……

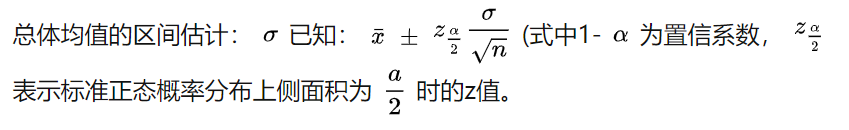
}

对于连续变量的定义区间范围的取值，应采用随机抽样调查法。即对固定时间段内(例如: 一小时内取10分钟的标签点值)的所有标签点值进行运算确定正常值的取值范围。例如: 风速，基于国家气象部门为每隔六小时进行一次观测(每天02时、08时、14时、20时),即要求主控方定时传回数据并对指定时间段内数据进行数据分析。此类数据也分为两种且同种数据可同时进行两种规则的校验，具体规则如下:

**规则二：变化范围比较大的连续变量。**应根据历史数据求出正常值的取值范围，即大部分值落在的区间为正常取值范围，少部分为异常值。当然整体中所有的数据都在这个范围也不现实，我们只需要绝大多数出现在置信区间就可以了，这里的绝大多数就是置信水平的概念，通常情况我们的置信水平是95%。置信区间[a，b]的计算方法为：（z分数：由置信水平决定，网上查表可得。）在校验过程中，可以选取同时间段的数据进行数值分析，若该区间段数值与定义区间差异极大，系统应自动返回异常信息且应与工作人员或现场人员联系并进行再次确认。

a = 样本均值 – z\*标准误差，b = 样本均值 + z\*标准误差

标准误差SE= s/ （s样本标准差， 样本大小）



**规则三：变化范围比较小的连续变量。**应根据历史数据进行运算求出离散程度(数据中最大值和最小值的差值)。之后可以选取相同时间段的数据进行数值分析并求出离散程度进行比较，如果离散程度相似，则认为主控方传回来的数据无异常值，反之系统返回异常信息并及时与工作人员或现场人员进行二次确认，确保对数据进行及时优化处理。

当要决定测量值是否符合预测值，测量值的标准差占有决定性重要角色：如果测量平均值与预测值相差太远（同时与标准差数值做比较），则认为测量值与预测值互相矛盾。因为值都落在一定数值范围之外，可以合理推论预测值是否正确。

方差(方差描述随机变量对于数学期望的偏离程度。)计算公式：

