在原始数据预处理中：

根据附件四提供的数据变量操作范围，使用最大最小限幅筛选方法，筛选出了68个超出范围的操作变量，并依据拉依达准则（3准则）进行异常值检测，共计删除了31条样本原始数据，保留了285号样本原始数据40 条，313号样本原始数据9条。

在保留的49条原始数据中，有5处数据项存在缺失，使用前后两小时数据平均值进行了插值补全。

为了克服测定数据较少、辛烷值测定存在一定滞后的局限性，我们最终计算了2小时内操作变量的平均值作为辛烷值操作变量数据插入附件一的样本数据中。

在样本数据预处理中：

使用插值补全了122项缺失数据，剩余的18个位点仅仅含有部分时间点数据，无法使用插值补充，直接将此位点删除，保留了336项操作变量。

根据附件四提供的操作变量范围，样本号为301、302、303、304的操作变量超出范围被删除，保留了321条样本数据。

*一左一右两幅饼状图，左边为样本数据，右边为原始数据*

*图上标出所占比例，不用标出具体数值*

*表名为“原始数据处理情况”，“样本数据处理情况”*

|  |  |  |
| --- | --- | --- |
| 原始数据 | 数据总计 | 28320 |
| 删除超出操作范围数 | 10974 |
| 替代数据缺失 | 5 |
| 删除空白点位 | 760 |
| 删除拉依达准则异常值 | 0 |
| 剩余有效数据 | 16586 |
| 样本数据 | 数据总计 | 119275 |
| 删除超出操作范围数 | 1950 |
| 替代数据缺失 | 176 |
| 删除空白点位 | 5525 |
| 剩余有效数据 | 111624 |