关于泰坦尼克号数据分析报告

1. 生还率分析

自变量：

1. pclass

选取pclass作为自变量的原因是，pclass分为三个等级，依据常识猜测，是否因为选择高等级船票的人较之选择低等船票的人生还几率更大。

1. sex

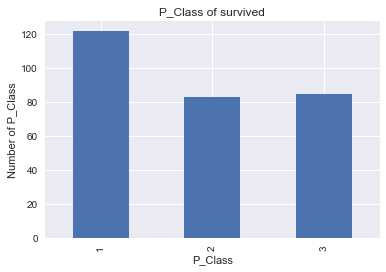
选取sex作为自变量的原因是，由于男女双方性别而导致身体质素的差异，是否会影响生存几率

1. age

判断是否由于年龄而影响生存几率

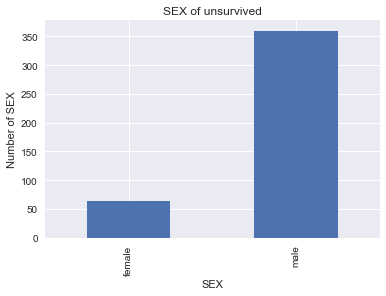
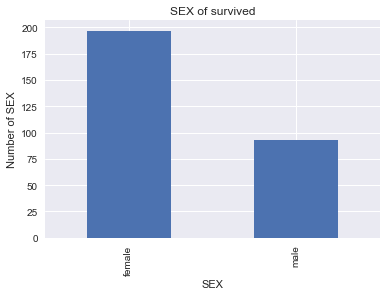
因变量：是否生还

关于Pclass的分析：



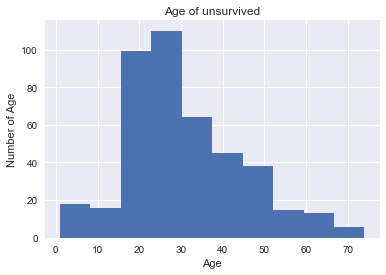
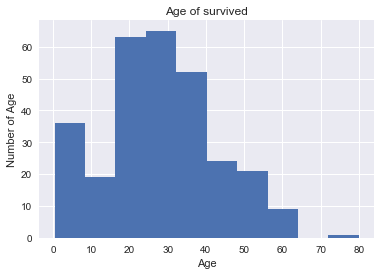
结论：1等船票生还几率最高，遇难几率最低

关于性别的分析：

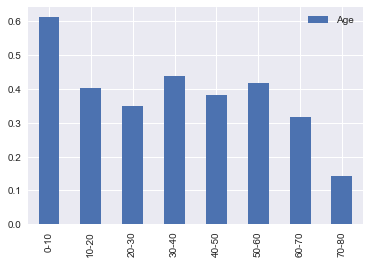


结论：女性生存人数最高，且遇难人数最低

关于age的分析：



结论1：由于age生存几率和遇难几率趋势相同，所以需要进一步进行分组操作，计算每个年龄段生还概率



结论2：由图表可知0-10岁儿童生存几率最高。