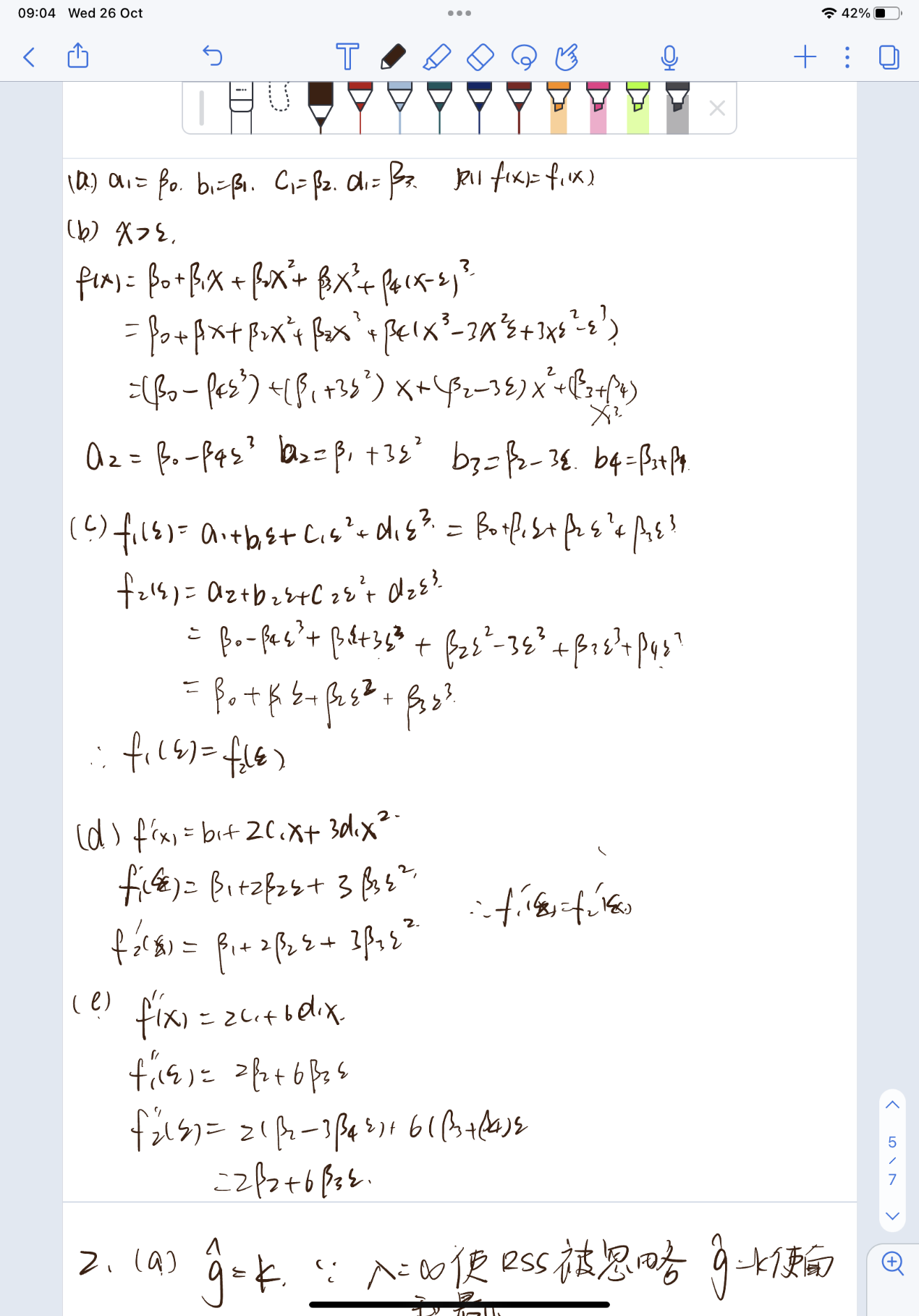
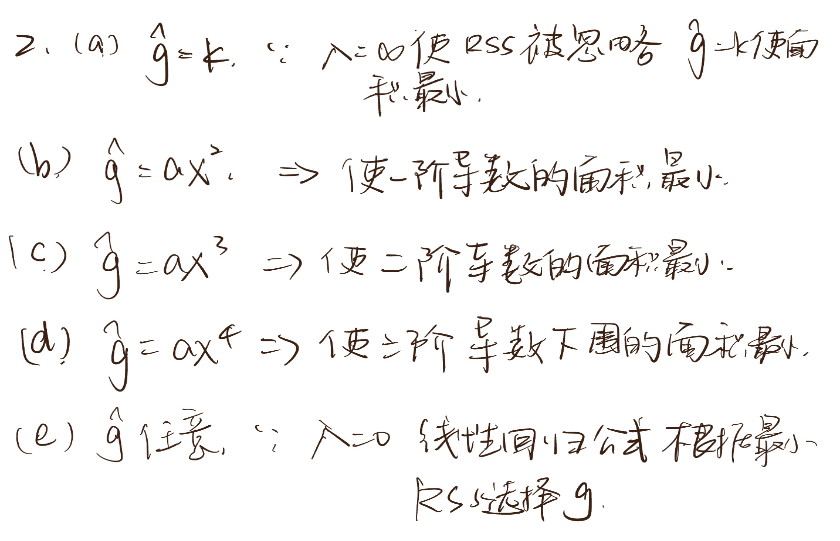
1



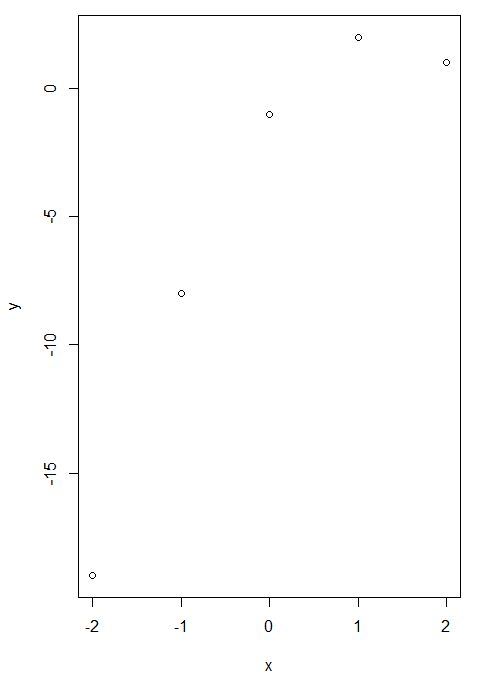
2



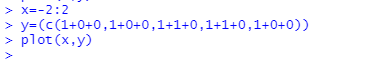
3

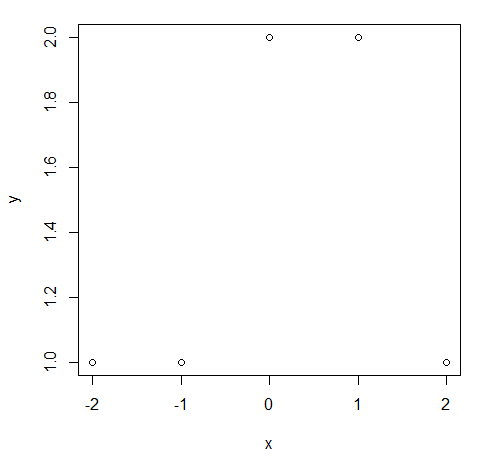




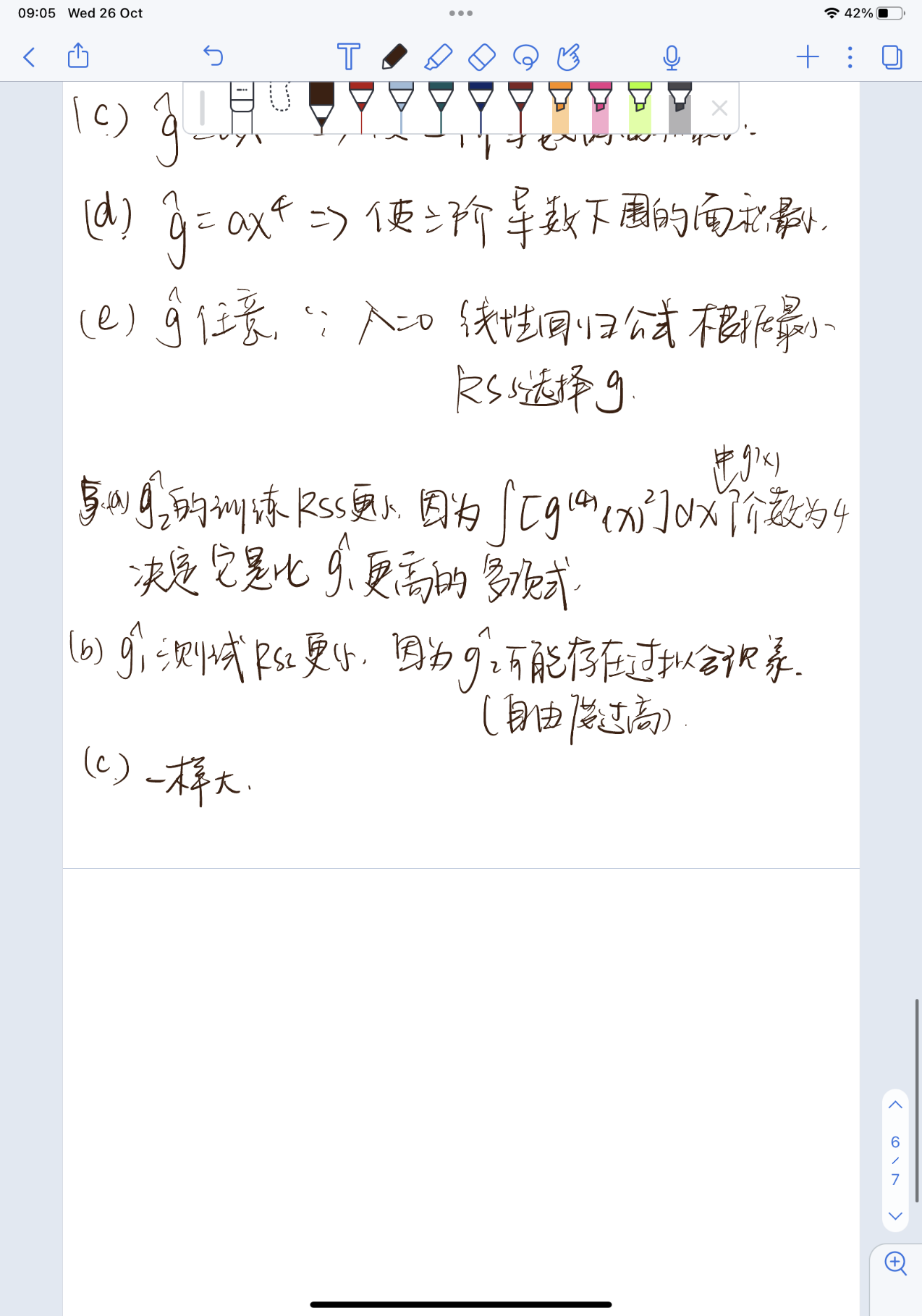


4.

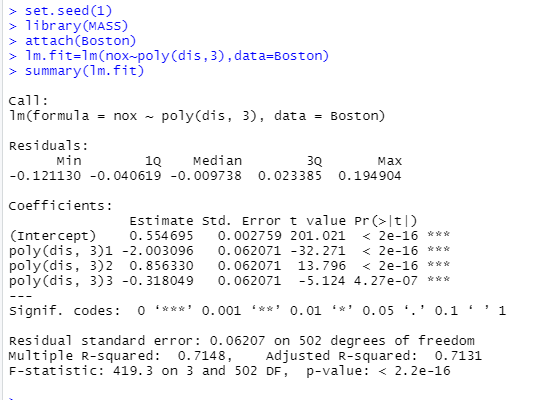


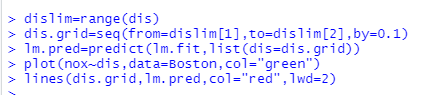


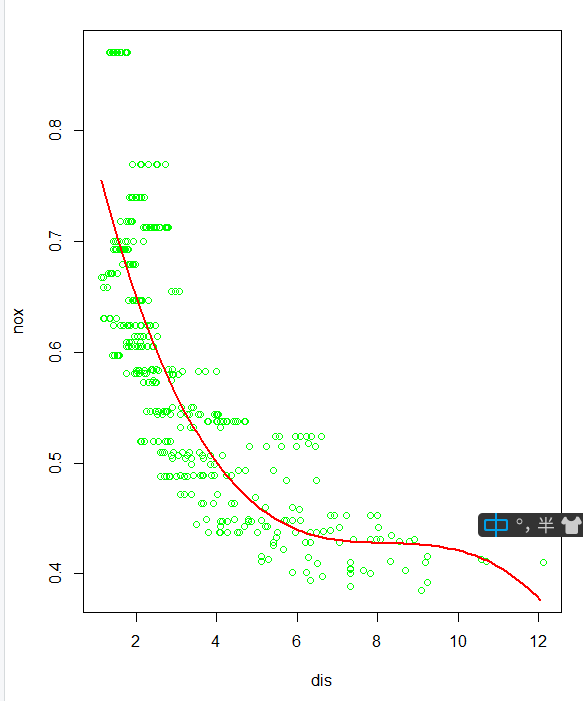
5



9（a）

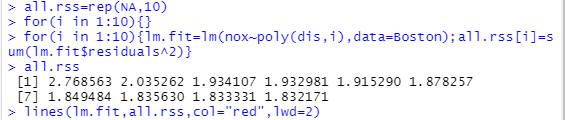


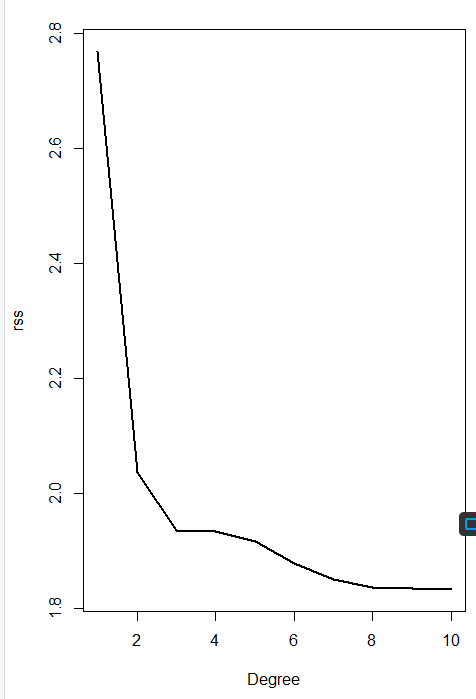




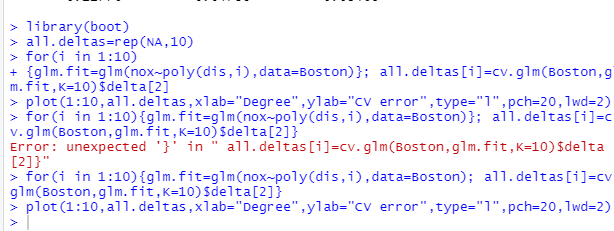
Summary函数显示的pr值表示在用dis预测cox的时候，所有多项式项都是有用并且显著的（因为pr值很小），并且拟合的曲线也很光滑

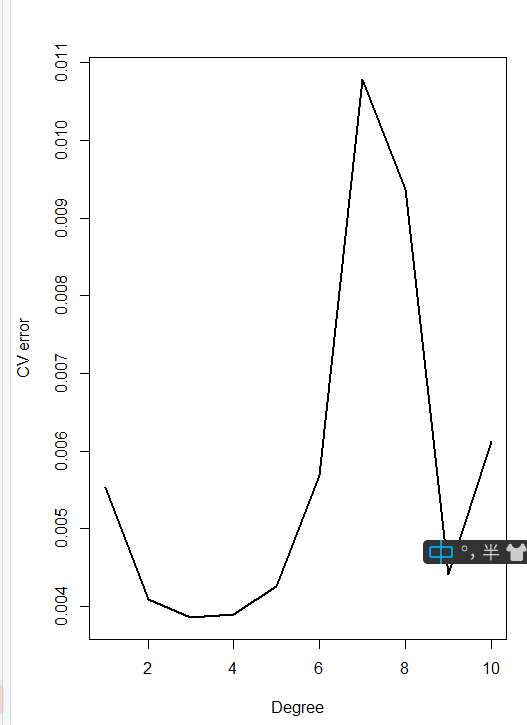
9（b）





9（c）

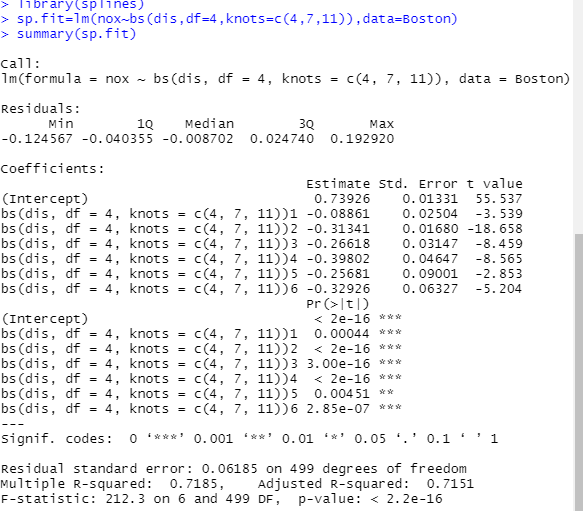


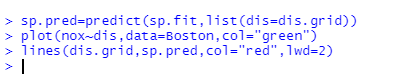


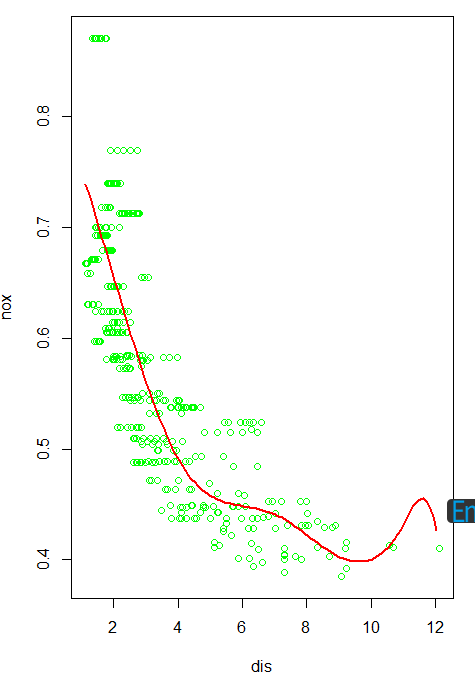
10倍的CV表明，当我们将度数从1增加到3时，CV误差会减少，当度数达到4时，CV误差达到最低。而度数越高，CV误差开始增加。我们选择4作为最佳多项式次。

9（d）

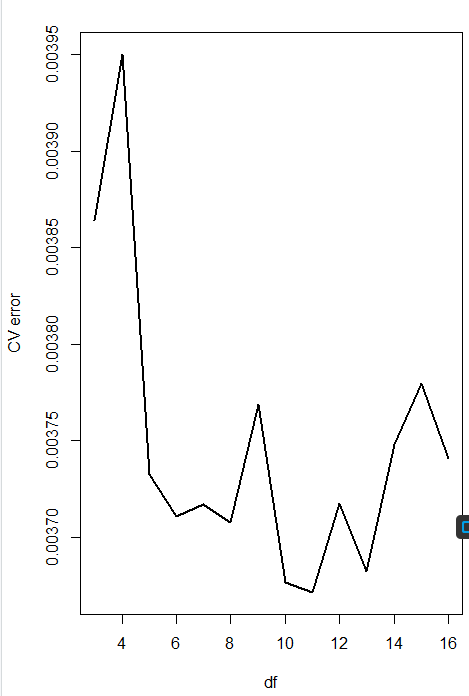
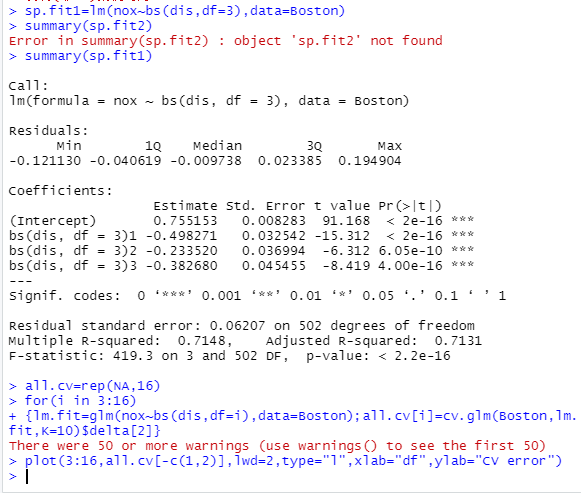
我们看到dis的极限分别是1和13。我们将这个范围分成大致相等的4个间隔，并在[4,7,11]处建立结点







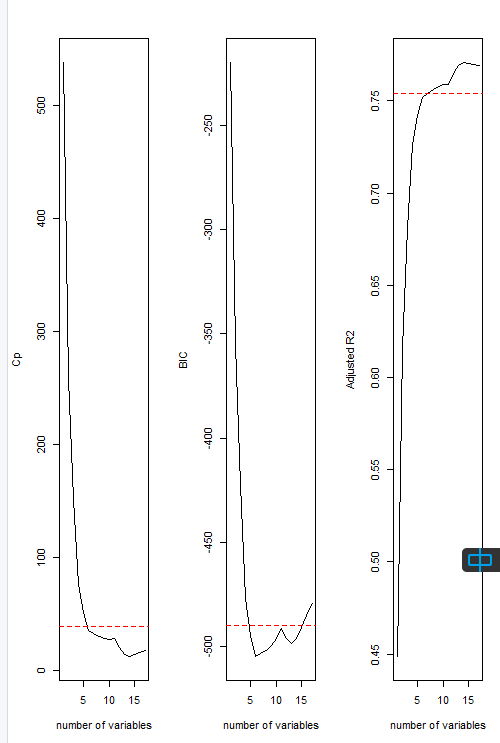
9(e.f)



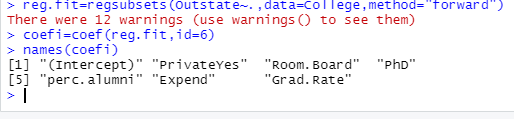
在这种情况下CV误差波动较大，但在df=10时达到最小值。我们选择10作为最优自由度。

10（a）

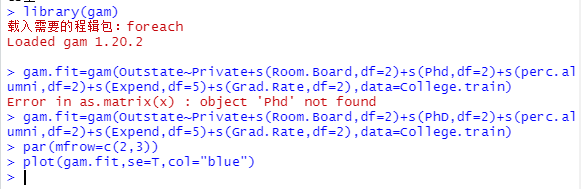


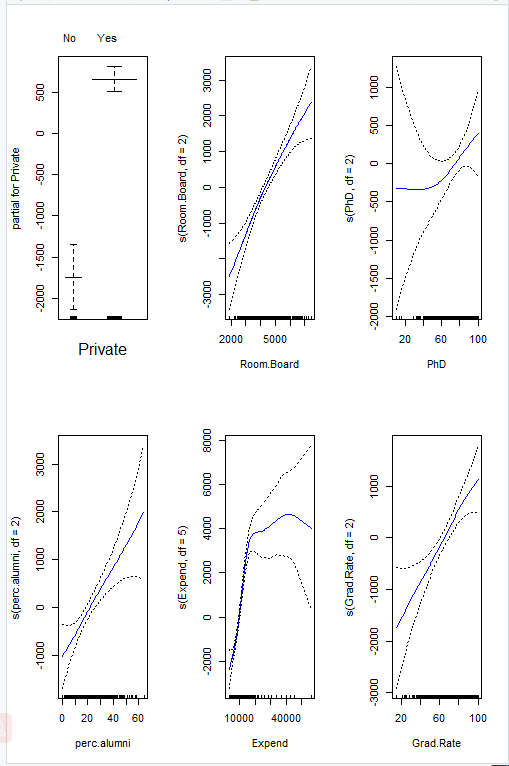


所有的cp、BIC和adjr2分数都表明，在分数与最优值的0.2标准差范围内，size 6是子集的最小大小。我们选择6作为最佳的子集大小，并使用整个数据找到最佳的6个变量

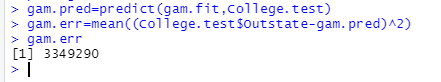


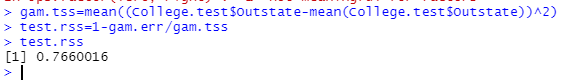
10（b）





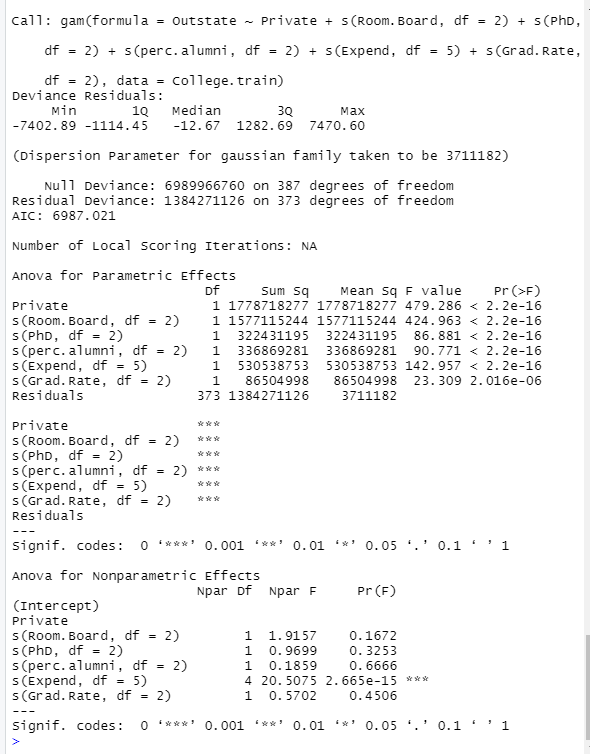
10（c）





我们使用带有6个预测因子的GAM得到测试总体r平方为0.77。

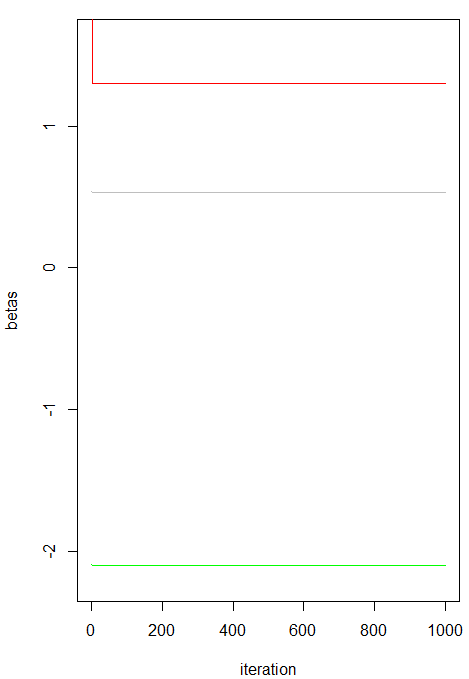
10（d）



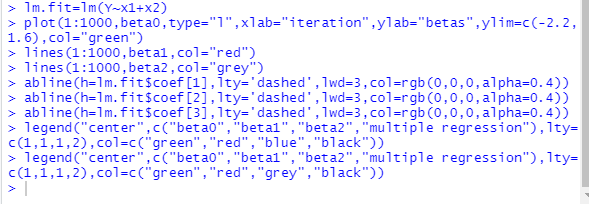
由数据可以看到expend和响应变量有很强的非线性关系。

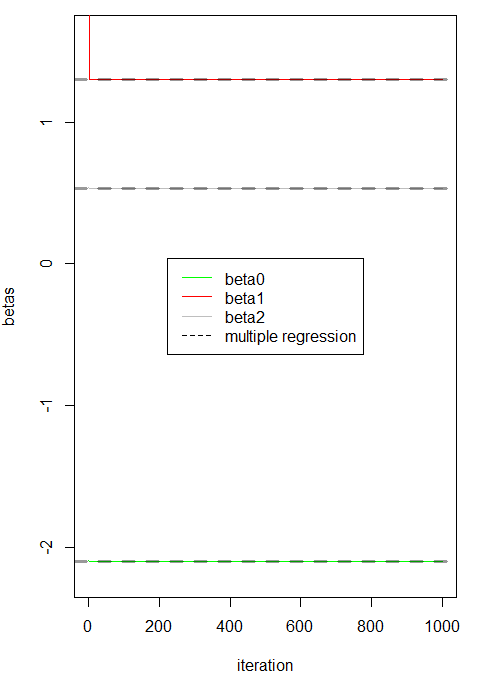
11（a-e）





11（f）





虚线表示估计的多元回归系数与用回拟法得到的系数完全匹配。

11（g）

当Y和X之间的关系是线性的，一次迭代就足以获得真实回归系数的良好近似。