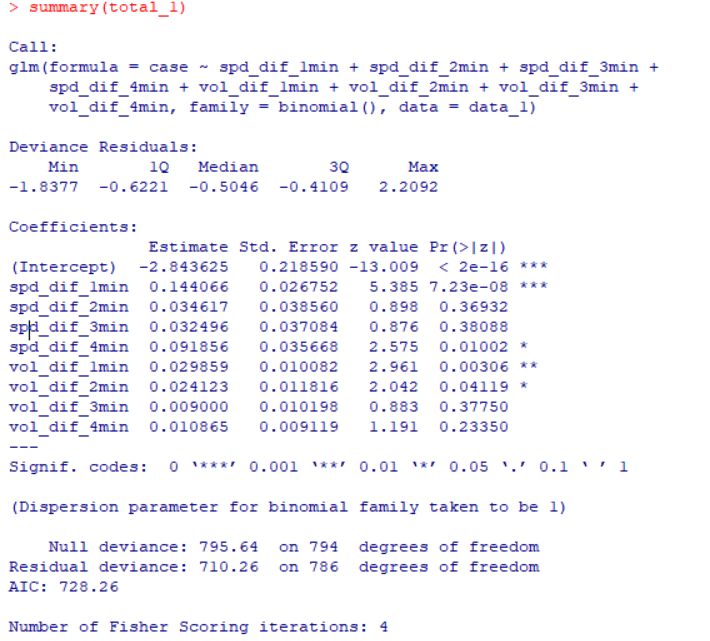
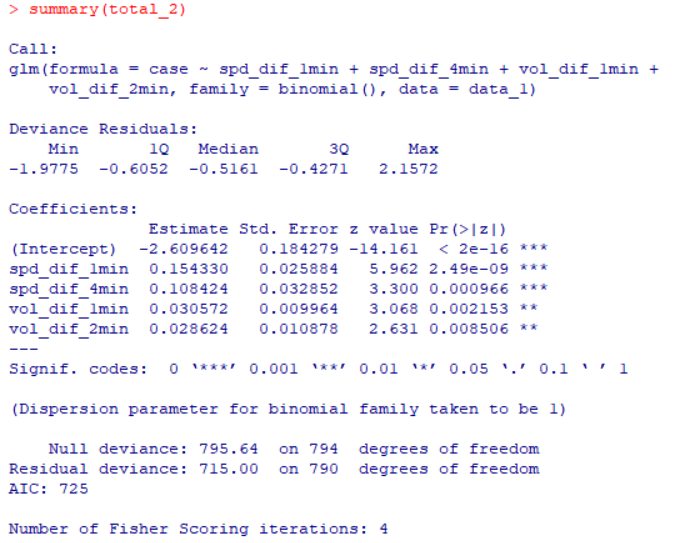
第一步建立事故是否发生与速度和流量之间的模型，由于事故是否发生为二值变量，因此采用逻辑回归模型，得到下图所示结果

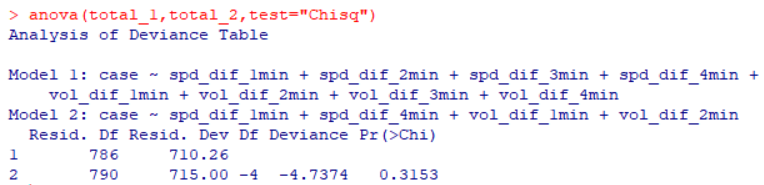


由上可以看出，回归系数比较显著（p<0.05）的是事故发生前1分钟内，4分钟内的速度变化值和事故发生前1分钟内，2分钟内的流量变化值。

因此去除不明显的，用显著变量建立新的逻辑回归模型如下



从上面的结果可以看出，4个参数的回归系数都非常显著，我们再用anova()函数对两个模型进行比较，用卡方检验，可得



可以看出p=0.3153,并不显著，因此用后面的四个变量模型和前面的拟合的一样好。所以就可以得到最佳模型如下

logit(p) = -2.6096+0.1543spd\_dif\_1min+0.1084spd\_dif\_4min+0.0306vol\_dif\_1min

+0.0286vol\_dif\_2min; y~dbern(p).